




Stability and folding pathways of tetra-nucleosome from six-dimensional free energy surface

Xinqiang Ding^{1,2}, Xingcheng Lin^{1,2}  & Bin Zhang¹  

The three-dimensional organization of chromatin is expected to play critical roles in regulating genome functions. High-resolution characterization of its structure and dynamics could improve our understanding of gene regulation mechanisms but has remained challenging. Using a near-atomistic model that preserves the chemical specificity of protein-DNA interactions at residue and base-pair resolution, we studied the stability and folding pathways of a tetra-nucleosome. Dynamical simulations performed with an advanced sampling technique uncovered multiple pathways that connect open chromatin configurations with the zigzag crystal structure. Intermediate states along the simulated folding pathways resemble chromatin configurations reported from in situ experiments. We further determined a six-dimensional free energy surface as a function of the inter-nucleosome distances via a deep learning approach. The zigzag structure can indeed be seen as the global minimum of the surface. However, it is not favored by a significant amount relative to the partially unfolded, in situ configurations. Chemical perturbations such as histone H4 tail acetylation and thermal fluctuations can further tilt the energetic balance to stabilize intermediate states. Our study provides insight into the connection between various reported chromatin configurations and has implications on the in situ relevance of the 30 nm fiber.

¹Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. ²These authors contributed equally: Xinqiang Ding and Xingcheng Lin. ✉email: binz@mit.edu

The organization for a string of nucleosomes is of fundamental interest and may dictate the transcriptional fate of enclosed genes¹. Significant research effort has been devoted to characterize chromatin structure *in vitro*^{2–4}, and has led to the discovery of the 30 nm fiber^{5–8}. The precise arrangement of nucleosomes inside the fiber remains controversial, though high-resolution structures for an array of nucleosomes with well-defined positions are becoming available^{9,10}. Even less known is the dynamical process along which chromatin folds from extended configurations to these regular fibril structures¹¹. Methodologies that enable efficient conformational exploration and accurate stability evaluation are needed for a detailed characterization of chromatin structure and dynamics. They may also provide insight into the *in vivo* relevance of the 30 nm fiber, which has been under debate after numerous failed attempts at finding similar motifs inside the nucleus with a variety of techniques^{12–16}.

Experimentally probing chromatin organization at atomistic resolution remains challenging due to difficulties in precise chromatin engineering and assembly. Computational modeling offers an alternative and promising approach, and has played crucial roles in uncovering the possible topology of nucleosomes in fibril structures^{17,18}. For example, the two angle model proposed by Woodcock and coworkers⁷ was instrumental for interpreting electron microscopy images of chromatin and for providing a consistent description of various chromatin fiber models^{19–21}. Schlick and coworkers developed a mesoscale model that explicitly describes the nucleosome core particle and its charge distribution, the histone tails, and the linker DNA^{2,22,23}. Applications of these models have resulted in numerous insights into the sensitivity of chromatin organization with respect to nucleosome repeat length, linker histones, salt concentrations, etc.^{24–28}. To better characterize the impact of DNA sequences, the Wilson group^{29,30}, Zhurkin group^{31–34}, and van Noort group³⁵ have developed chromatin models in which the DNA molecule is represented at a single base-pair resolution. The de Pablo group further introduced a systematic parameterization and coarse-graining procedure to derive mesoscopic models that are potentially of improved accuracy³⁶.

Here, we applied a near atomistic model to study the basic unit of chromatin fibers, the tetra-nucleosome. Protein and DNA molecules were represented with residue and base-pair resolution to provide a particle-based description of electrostatic interactions and specific protein–protein interactions that are crucial at stabilizing inter-nucleosome contacts. Dynamical simulations of this model uncovered two folding pathways that differ in the ordering of nucleosome contact formation. While the sequential pathway connects open chromatin configurations to the compact zigzag structure¹⁰ with tri-nucleosome structures, the concerted pathway goes through conformations that resemble a β -rhombus. These folding intermediates bear striking similarity to chromatin configurations observed *in situ*^{16,37}. We then computed the free energy surface as a function of the six inter-nucleosomal distances with a neural network approach and confirmed the statistical significance of these pathways. The free energy surface further suggests that while the stacked, zigzag configuration resides in the global minimum, the folding intermediates bear comparable stability. Significantly, these intermediate configurations can be further stabilized by configurational entropy, histone modifications, and variation in the secondary structure of histone tails, highlighting the sensitivity of chromatin organization to thermal and chemical perturbations. Our study suggests that chromatin configurations observed *in situ* might form as a result of local excitations or unfolding from the global minimum of the *in vitro* fibril structures.

Results

Direct simulations of tetra-nucleosome folding. We applied a near-atomistic model to characterize the folding process during which chromatin transitions from extended configurations with minimal contacts between nucleosomes to the compact zigzag crystal structure (PDB ID: 1ZBB)¹⁰. The near-atomistic model has been used extensively in prior studies to investigate nucleosome unwinding and nucleosome–nucleosome interactions (see Methods). A tetra-nucleosome with 20-bp-long linker DNA representing a minimal stable structural unit of chromatin fibers^{9,11} was employed in our simulations.

To facilitate observations of large-scale conformational rearrangements, we combined meta-dynamics³⁸ with temperature-accelerated molecular dynamics (TAMD)^{39,40} to bias simulation trajectories along two collective variables, Q and R_g . The fraction of native contacts Q is defined as

$$Q = \frac{1}{6} \sum_{i=1}^3 \sum_{j=i+1}^4 \exp \left[-\frac{(r_{ij} - r_{ij}^o)^2}{2\sigma^2} \right], \quad (1)$$

where i and j are indices of nucleosomes and $\sigma = 2$ nm. r_{ij} measures the distance between the center of nucleosome i and j , and r_{ij}^o is the corresponding value in the crystal structure¹⁰. Q measures the similarity between a given tetra-nucleosome configuration and the crystal structure, while R_g , the radius of gyration, quantifies the overall compaction and size of the structure. As shown in an example trajectory (Fig. 1a), the tetra-nucleosome undergoes multiple collapse/expansion transitions, as indicated by Q varying from 0.1 to 0.8, and R_g from 15 to 5 nm. On the other hand, without the advanced sampling techniques, tetra-nucleosome unfolding occurs on a much slower timescale

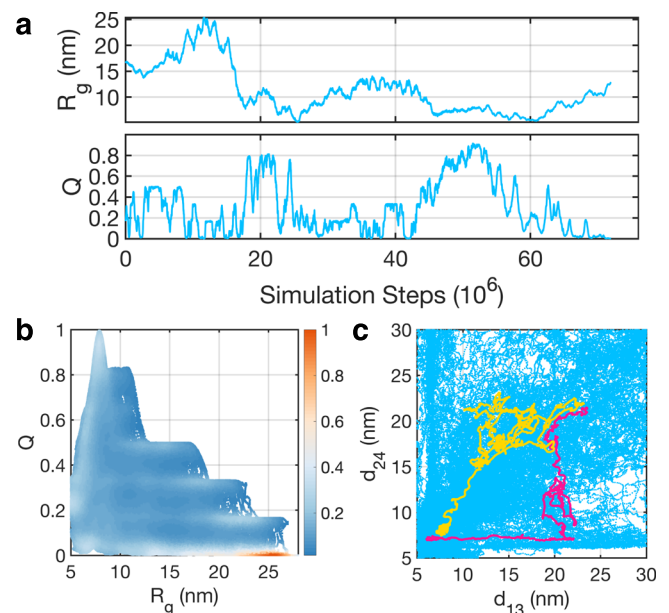


Fig. 1 Direct simulations of the tetra-nucleosome folding process enabled via an advanced sampling technique. **a** Time evolution of the two collective variables along an example trajectory during which chromatin undergoes multiple folding and unfolding transitions. **b** Scatter plot of all the sampled tetra-nucleosome configurations represented by R_g and Q . Color scale represents relative frequency estimated from direct counting that increases from blue to white and to red. **c** Scatter plot of all the tetra-nucleosome configurations represented by 1–3 and 2–4 nucleosome distances. Two example trajectories that correspond to the concerted and sequential folding pathway are highlighted in yellow and pink, respectively.

and was not observed during the simulation time (Supplementary Fig. 1).

We carried out a total of 20 independent simulations starting from different open tetra-nucleosome conformations for a comprehensive exploration of the configurational space. Time evolutions of Q and R_g along these trajectories are provided in Supplementary Fig. 2. Similar to the example shown in Fig. 1a, large scale changes in tetra-nucleosome conformations can be observed. Collectively, these trajectories provide a rather complete and relatively uniform coverage of the phase space (see Fig. 1b). We note that the negative correlation between R_g and Q restricts the data points to the lower bottom of the plane. We further projected the simulated trajectories onto the distances between 1–3 and 2–4 nucleosomes, both of which reach their minimal values in the zigzag configuration. As shown in Fig. 1c, these two degrees of freedom are sufficiently sampled as well. The relatively complete coverage over these two new collective variables supports the efficacy of the employed sampling techniques for conformational exploration.

To gain more intuition into the simulated chromatin configurations, we used the K-means algorithm to identify the 100 most populated clusters. Examples of cluster central structures are shown in Fig. 2, and the full list is provided in Supplementary Movie 1. Stacked configurations, for which the 1–3 and 2–4 nucleosomes are in close contact as in the crystal structure, can be readily seen. However, compared to the crystal structure, these conformations appear more compact and less twisted, with a decrease in the angle between vectors formed by 1–3 and 2–4 nucleosomes. Consistent with prior studies, histone H3 (orange) and H4 (purple) terminal tails were found to mediate the close contacts between nucleosomes^{23,41}. While H4 tails primarily bridge non-neighboring nucleosomes, H3 tails form contacts with linker DNA and bring neighboring nucleosomes close.

In addition to the stacked configurations, we observed co-planar structures in which the two vectors formed by neighboring nucleosomes are either parallel, anti-parallel, or perpendicular to each other. There are also structures in which the four nucleosomes align almost along a straight line (co-linear) and along a left-handed helical path (helical). Finally, for tri-nucleosome structures, three nucleosomes adopt positions seen in the crystal structure, while the last one extends out. The disordered histone tails were again found to localize at the interface between nucleosomes to mediate contacts. Many of these structures are strikingly similar to those reported in prior studies that probe chromatin configurations *in situ*^{15,16}. Therefore, we anticipate that these structures could be useful for interpreting experimental results and provide their 3D coordinates on the group GitHub page (<https://github.com/ZhangGroup-MITChemistry/TetraNucl>).

Constructing high-dimensional free energy surface with deep learning. The diversity of structures suggests that while R_g and Q are sufficient for conformational exploration, additional collective variables are needed to fully capture the complexity of tetra-nucleosome organization. Furthermore, it would be beneficial to quantify the thermodynamic stability of the intermediate configurations. Due to the biases introduced in meta-dynamics and TAMD, such information cannot be directly obtained from the dynamical simulations.

We, therefore, determined the free energy surface as a function of the six inter-nucleosome distances, $\mathbf{d} = (d_{12}, d_{13}, d_{14}, d_{23}, d_{24}, d_{34})$. These distances are crucial for differentiating various tetra-nucleosome configurations. We note that a high-dimensional free energy surface is challenging to calculate with traditional

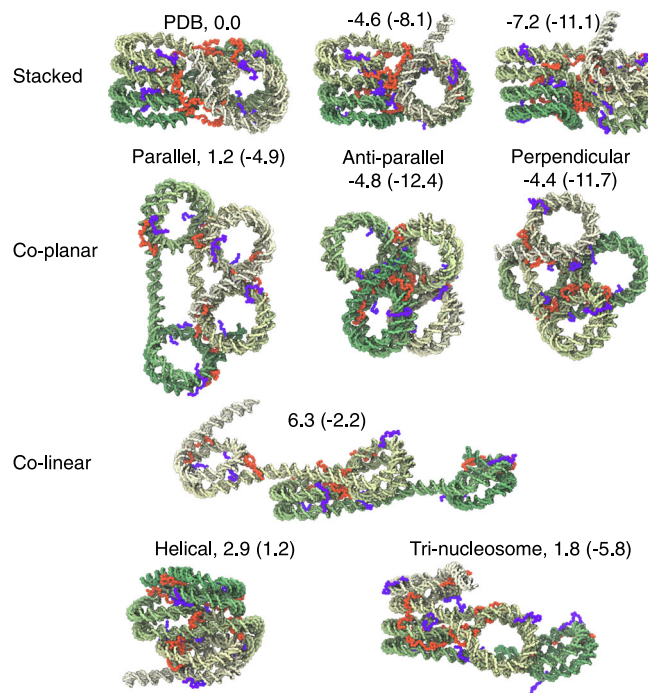


Fig. 2 Representative tetra-nucleosome configurations from K-means clustering. See Supplementary Movie 1 for the full set of cluster structures. For clarity, only the DNA molecule, the color of which varies from green to white, is shown. The N-terminal tails of histone H3 and H4 protein are drawn in purple and orange, respectively. The numbers indicate their stability relative to the PDB structure (top left) measured in the unit of kcal/mol, and were evaluated using simulations performed with the presence of protein–protein interactions between nucleosomes. The corresponding values inside the parentheses were obtained using simulations performed without such interactions.

techniques, including umbrella sampling⁴², due to their exponential increase in computational cost with dimensionality. Instead, we adopted a neural network approach to calculate the free energy from mean forces determined at a total of 10,000 preselected centers⁴³ (see Fig. 3a and Methods for details). A key advantage of this approach is that no overlap among centers is required, thus alleviating the curse of dimensionality by avoiding a uniform coverage of the phase space.

Once parameterized, the neural network can provide free energy estimations for any configuration represented by the six inter-nucleosome distances. Mean forces can be determined from the network as well via the back-propagation algorithm⁴⁴. As shown in Fig. 3b, the mean forces calculated using the trained neural network are in good agreement with the ones produced from restrained molecular dynamics (MD) simulations, and the Pearson correlation coefficient between the two is 0.87. While fine-tuning hyperparameters can lead to higher correlation coefficients, the network becomes less robust and tends to overfit the data (see Supplementary Figs. 3 and 4), partially due to the inherent uncertainty in mean force estimations using MD simulations.

To directly validate the accuracy of the free energy estimation, we integrated over the other five degrees of freedom to obtain a one-dimensional free energy profile as a function of the distance between 1–3 nucleosomes. For comparison, we carried out additional simulations to compute the same quantity using umbrella sampling⁴² and the weighted histogram analysis method (WHAM)⁴⁵. Details for these simulations are provided in the

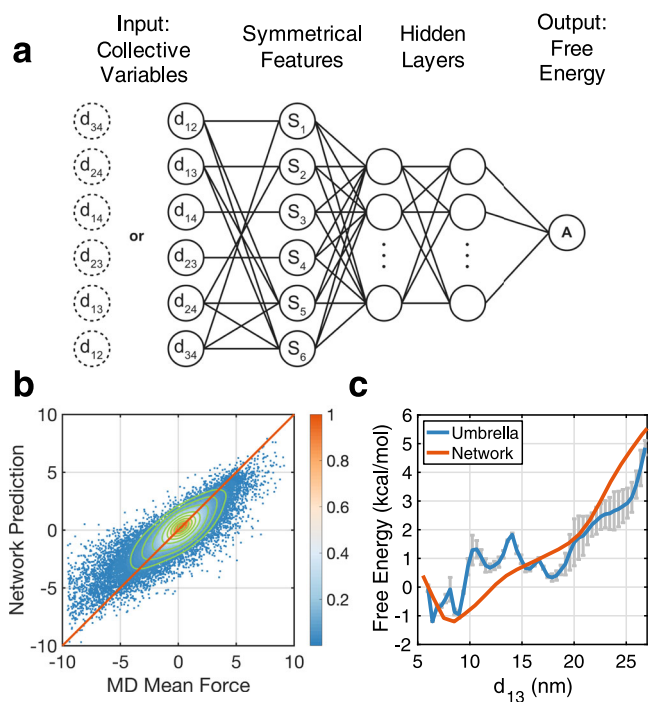


Fig. 3 Parameterization and validation of the six-dimensional free energy surface with a multilayer neural network. **a** Illustration of the neural network designed to take into account the symmetry of the structure with respect to index ordering. Input of this network is tetra-nucleosome configurations represented via inter-nucleosomal distances and output is the corresponding free energy. **b** Comparison between mean forces in the unit of kcal/mol/nm determined from restrained molecular dynamics (MD) simulations and neural network prediction. The relative density of data points increases as the color changes from blue to white to red. The green lines are constant density contours and the red diagonal line is shown as a guide for the eye. **c** Comparison between the free energy profile as a function of 1–3 nucleosome distance determined from neural network (orange) and umbrella sampling (blue). The error bars (grey shadow) represent one standard deviation from the mean.

supporting information (SI). As shown in Fig. 3c, despite the significant difference between the two methodologies, the two free energy profiles agree well with each other, and deviations between them are less than 1 kcal/mol across the entire range of distance. We note that the free energy profile from umbrella sampling is rather rugged, despite our significant simulation effort (Table S1). The ruggedness arises because the system at neighboring umbrella windows can be trapped in different tetra-nucleosome configurations with comparable d_{13} and interconversion among these configurations is slow.

We next used the neural network to quantify the stability of the example structures shown in Fig. 2. The free energy values are provided along with the structures in parentheses. We found that stacked conformations with more aligned nucleosome columns have lower free energy than the PDB structure. For example, the 1–3 nucleosome axis is more parallel to the 2–4 nucleosome axis in these conformations. Rotating the two axes relative to each other could cause twisting in the linker DNA, potentially explaining the higher free energy of the PDB structure. The rotation in the PDB structure could be a result of crystal packing as well as the high concentration of Mg^{2+} ions used in sample preparation^{10,46,47}. To ensure that the predicted structural change is not an artifact of the near-atomistic model used here, we carried out additional simulations using the SIRAH force field⁴⁸.

While molecules are still treated at a coarse-grained level, this force field includes explicit representations of water molecules and ions and is expected to describe the solvation effect and electrostatic interactions accurately. As shown in Supplementary Fig. 5, both an unbiased 4.0- μ s-long dynamical trajectory and the free energy profile estimated from a series of \sim 1.0- μ s-long umbrella sampling simulations support the stability of the less rotated conformations.

In addition, we found that some of the co-planar structures adopt comparable, or even lower, free energy values relative to the stacked configurations. Their stability potentially arises from the absence of close contacts between inter-nucleosomal DNA seen in the stacked structures while maintaining favorable interactions between histone tails and the DNA. One or both of these two features are missing or compromised in co-linear, helical, and tri-nucleosome configurations, giving rise to their higher free energy. We note that the computed free energy is a function of inter-nucleosome distances and measures the stability for an ensemble of structures, not just the ones shown in Fig. 2. For example, individual nucleosomes and histone tails can rotate and diffuse without significantly altering inter-nucleosome distances. These movements further contribute to configurational entropy and the stability of the co-planar structures as well but are severely impaired in the stacked configurations due to their tight packing.

Identifying most probable chromatin folding pathways. The high dimensional free energy surface also enables a rigorous characterization of the folding pathways to form stacked structures. Such pathways could offer insight into the stability of regular fibril configurations and their relationship to the more irregular structures often seen in situ. Our exploratory, dynamical simulations indicated a complex mechanism with multiple folding pathways, as highlighted by the two continuous trajectories in Fig. 1c (see Supplementary Movies 2 and 3). The formation of contacts along the yellow trajectory is more concerted, and there is a strong correlation between the two non-neighboring inter-nucleosome distances. On the other hand, 1–3 nucleosomes remain far apart as 2–4 nucleosomes stack on top of each other along the pink, sequential trajectory. However, since these trajectories were obtained from biased simulations, their thermodynamic relevance remains to be shown.

Starting from the parameterized free energy surface, we searched for the most probable chromatin folding pathways that connect an extended configuration with the stacked conformation (Fig. 4a) using the finite-temperature string method^{49,50} (see Methods Section for details). As shown in Fig. 4b and Supplementary Fig. 6, three sets of distinct pathways were identified. Consistent with the exploratory trajectories shown in Fig. 1c, two of the paths indeed correspond to the sequential pathway, and the third one represents the concerted pathway. Also, tetra-nucleosome configurations sampled along the paths are well separated, supporting the statistical significance of the difference between them. For simplicity, the paths were shown in the two dimensions of 1–3 and 2–4 inter-nucleosome distances. The variations of all distances along the paths are provided in Supplementary Fig. 7. Our observation of sequential and concerted pathways in both string method calculations and dynamical simulations supports their robustness and the independence of the folding mechanism from our choice of collective variables and computational techniques.

Next, we reconstructed chromatin configurations along the calculated strings using restrained MD simulations centered at the corresponding inter-nucleosome distances. Upon close examination, we found that the example structures shown in Fig. 2 emerge as intermediates along the folding pathways. In

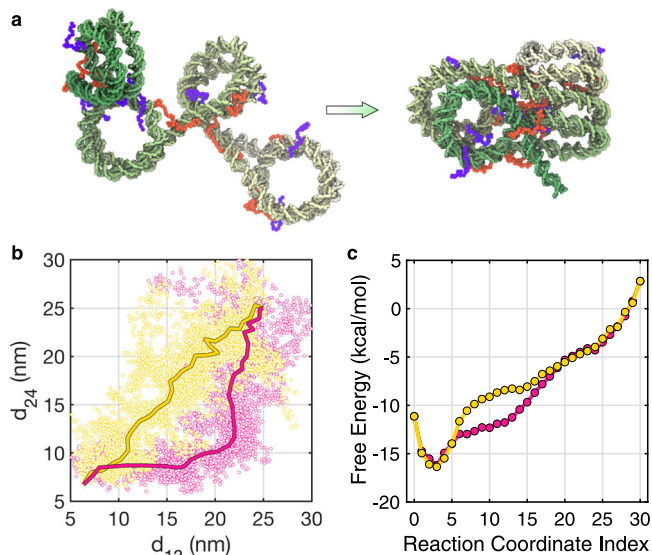


Fig. 4 Chromatin folding pathways determined using the finite temperature string method. **a** Illustration of the starting and end configurations of the chromatin folding pathways using the same coloring scheme as in Fig. 2. **b** The sequential (pink) and concerted (yellow) pathways plotted over the two inter-nucleosome distances. The dots correspond to tetra-nucleosome configurations obtained from individual simulations restricted to Voronoi cells defined using images along the path. They represent fluctuations around the path on the scale of $k_B T$. **c** Free energy profiles along the two pathways. See Supplementary Fig. 6 for corresponding results of the second sequential pathway.

particular, the concerted path goes through co-planar configurations that resemble parallel or perpendicular structures (see Supplementary Movie 4). Similarly, tri-nucleosome structures can be readily seen along the sequential pathway (Supplementary Movie 5). Therefore, these chromatin structures observed by in situ experiments can be viewed as en route to fibril conformations. We did not observe anti-parallel structures despite their apparent stability. Further investigations suggest that they might serve as kinetic traps for forming fibril structures and must be undone before folding proceeds (see Supplementary Fig. 8).

In Fig. 4c, we plotted the free energy profiles along the representative sequential and concerted pathways shown in Movies S4 and S5. These profiles compare favorably to the force-extension curves obtained by van Noort and coworkers using single-molecule force spectroscopy experiments^{35,51} (see Supplementary Fig. 9). An almost continuous decrease of the free energy supports chromatin folding as downhill transitions. The similarity of the two profiles argues for the significance of both pathways with comparable kinetics. The sequential pathway does appear to transition across more favorable chromatin configurations with lower free energy starting at reaction coordinate index 16. Such favorability arises because contacts between nucleosomes are formed earlier in the tri-nucleosome structures than in the intermediates along the concerted string.

Determinants of chromatin conformational stability. The quantitative free energy surface suggests that the fibril configurations with stacked nucleosomes are not particularly stable compared to partially open, unfolded structures. Their weak stability can also be seen from the relative flatness of the free energy profile as a function of 1–3 and 2–4 nucleosome distances

around the global minima (Fig. 5a). In the following, we investigate in detail factors that contribute to the stability of chromatin organization.

Histone proteins from different nucleosomes may form contacts to promote chromatin folding and stabilize the zigzag crystal structure. Therefore, we introduced specific interactions between proteins using the Miyazawa–Jernigan (MJ) potential that has been widely used for studying protein folding^{52,53}. A new set of simulations were carried out to estimate mean forces with the presence of protein–protein interactions, starting from the same set of configurations as those used for results shown in Fig. 3. We then trained another neural network to compute the free energy surface from these mean forces. As shown in Fig. 5b, the global minimum indeed shifts toward smaller inter-nucleosome distances. Notably, chromatin configurations with stacked nucleosomes are now favored over the co-planar structures (see Fig. 2). To our surprise, the overall quantitative change on the free energy profile from the one computed without accounting for protein–protein interactions appears minimal, suggesting the importance of additional factors for chromatin stability.

We next studied, in addition to specific interactions, the impact of non-specific contributions from configurational entropy on chromatin stability. We repeated the mean force calculations with protein–protein interactions, but at a temperature of 250 K instead of 300 K. As shown in Fig. 5c, the free energy surface parameterized from these mean forces strongly favors the crystal structure as the most stable configuration. This result suggests that the stability of unfolded, irregular chromatin structures, at least in part, arises from the configurational entropy. As mentioned before, unlike the stacked configurations, these structures place less constraint on the movement of individual nucleosomes and histone tails and are expected to be more dynamic.

Finally, since histone proteins undergo a variety of post-translational modifications that affect inter-nucleosome interactions, we explored the impact of acetylation of the H4 tail on chromatin stability. We first focused on the residue K16, the acetylation of which has implications in aging⁵⁴. Specifically, we neutralized the positive charge on the residue to mimic the chemical effect of acetylation, repeated the mean force calculations without specific protein–protein interactions at 300 K, and retrained the neural network to compute the free energy. As shown in Fig. 6a, the resulting free energy profile is somewhat similar to the wild type result shown in Fig. 5a with a different color scale. Prior studies have suggested, in addition to electrostatic interactions, acetylation could alter the secondary structure of histone tails^{23,55,56}. Via a multi-scale approach, Collepardo-Guevara et al. further showed that an increase in structural ordering of H4 tail could unfold chromatin by limiting internucleosomal interactions²³. Therefore, we carried out additional simulations that both remove the positive charge on K16 and restrain the H4 tail to a beta-sheet motif uncovered in all-atom explicit solvent simulations²³. As shown in Fig. 6b, consistent with the results by Collepardo-Guevara et al., unfolded chromatin configurations become more stable, and the free energy profile exhibits additional basins at large inter-chromosomes distances. Therefore, for the single residue acetylation, the impact of secondary structure changes appears to outweigh electrostatic interactions. We further studied an additional system in which the entire H4 tail is acetylated to evaluate the overall contribution of all the nine positive charges to chromatin stability. To focus on the role of electrostatic interactions, we did not introduce biases to the tail secondary structure. The resulting two-dimensional profile (Fig. 6c) again supports the stability of open chromatin conformations with large inter-nucleosome distances. Therefore, charged interactions,

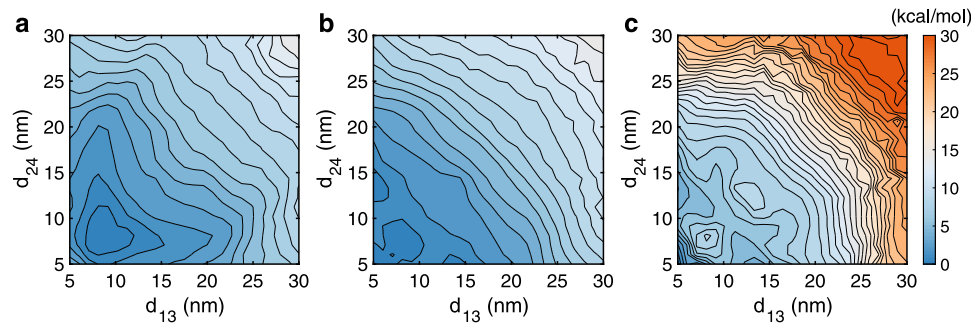


Fig. 5 Impact of specific and non-specific interactions on chromatin stability. **a, b, c** Contour of two dimensional free energy profile as a function of the distance between 1-3 (d_{13}) and 2-4 (d_{24}) nucleosomes at 1 kcal/mol interval for the system without **(a)** and with **(b)** protein-protein interactions at 300 K and with protein-protein interactions at 250 K **(c)**.

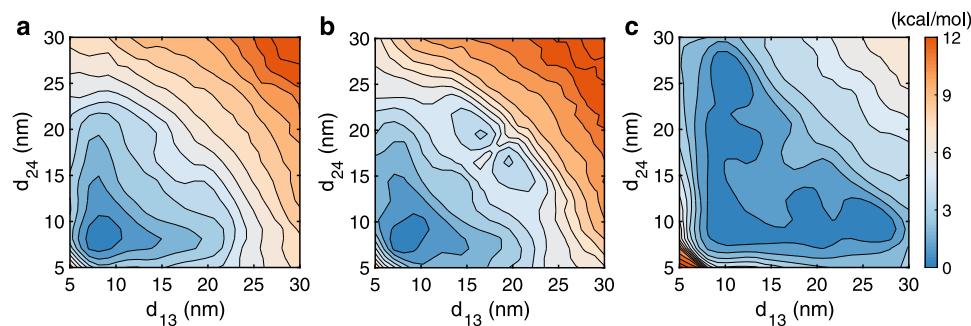


Fig. 6 Impact of histone H4 tail acetylation on the free energy landscape of chromatin folding. **a, b, c** Contour of two dimensional free energy profile as a function of the distance between 1-3 (d_{13}) and 2-4 (d_{24}) nucleosomes at 1 kcal/mol interval for the system that removes the positive charge on K16 **(a)**, that removes the positive charge on K16 and biases the tail to a beta-sheet conformation **(b)**, and that removes all the nine positive charges on the tail **(c)**.

especially when cumulated over multiple residues, contribute to H4 tails' role in stabilizing compact chromatin conformations as well^{24,57}.

Discussion

We studied tetra-nucleosome folding with a near-atomistic force field that models protein and DNA molecules at single residue and base-pair resolution. Compared to mesoscopic models, this force field gives rise to a more rugged energy landscape and a significant increase in computational cost, creating challenges for chromatin conformational sampling. We found that a technique that combines meta-dynamics with TAMD succeeded in producing continuous trajectories that connect open, extended structures to collapsed and folded ones. A deep-learning approach, together with mean force calculations, further allowed the computation of a six-dimensional free energy surface around the phase space explored by the biased simulations. The free energy surface provides unbiased evaluation on the relative stability of tetra-nucleosome configurations and its sensitivity to chemical and thermal perturbations.

Our simulations revealed a wide variety of 3D configurations that tetra-nucleosomes could adopt in addition to the zigzag structure resolved by X-ray crystallography¹⁰. Several of these simulated configurations have indeed been reported in prior experimental studies. In particular, by performing cryo-electron tomography on a HeLa cell¹⁵, Gan and coworkers directly observed tri-nucleosome configurations similar to the one shown in the bottom of Fig. 2, and found that they are more prevalent than compact structures consisting of four or more nucleosomes. On the other hand, structural motifs that resemble the parallel (β -rhombus) and helical (α -tetrahedron) have been reported in models derived from chromosome conformation capture

experiments that measure spatial proximity between nucleosomes inside the nucleus via chemical cross-linking and high-throughput sequencing¹⁶.

The relationship between in situ chromatin configurations and the zigzag crystal structure is evident from the dynamical trajectories shown in Fig. 1c (yellow and pink) and Supplementary Movies 2 and 3. These trajectories continuously bridge tri-nucleosome and β -rhombus configurations with the zigzag structure as time progresses. Therefore, the in situ configurations can be viewed as folding intermediates along the pathways towards the zigzag structure. We note that the dynamical trajectories were obtained from simulations biased over two collective variables, R_g and Q . Mechanistic insights drawn from these simulations are only reliable if the two variables represent the slowest degrees of freedom of the system. Otherwise, they may obscure the importance of true reaction coordinates that dictate chromatin folding. Importantly, however, string method calculations that track the most probable pathways for tetra-nucleosome folding support the same conclusions as the dynamical simulations. These calculations used a larger set of collective variables and an unbiased free energy surface. They provide strong support for the thermodynamic significance of the folding mechanism observed in exploratory trajectories.

Our study has significant implications on the in situ relevance of the 30 nm fiber^{58,59}. The free energy surfaces support the global stability of regular, fibril configurations for isolated chromatin. Since in situ configurations correspond to folding intermediates, they can also be viewed as local excitations or unfolding from the global minimum. Such excitations are not too rare as the energetic penalty arising from a loss of inter-nucleosome contacts can be largely compensated by an increase in the configurational entropy. Histone modifications, and potentially other factors such

as linker length, can even tilt the balance to drive the global stability towards intermediate configurations. Given the heterogeneous environment inside the nucleus, it is perhaps not too surprising to find the prevalence of partially unfolded, irregular structures.

As aforementioned, the near-atomistic model remains computationally costly for simulating long chromatin segments. Therefore, we anticipate that it will be most effective for simulating systems of tens of nucleosomes in size and evaluating the relative stability of various topological forms^{17,18,60,61}. For studying conformational arrangement of genomic regions that are 10–100 kb in length, mesoscopic models introduced by the Schlick or the de Pablo group offer a more promising route with better balanced accuracy and computational efficiency^{22,62}. At even larger scales such as a whole chromosome or the entire genome, integrative approaches that take into account the impact of the nucleus environment with experimental constraints may provide more faithful structural models for chromatin^{63,64}.

Several aspects of the near-atomistic model can be further improved for better accuracy. In particular, fine-tuning protein–protein interactions could lead to a more balanced treatment of both ordered and disordered regions of histone proteins. Introducing base pair specific protein–DNA interactions may help account for DNA sequence-dependent chromatin stability⁶⁵. Furthermore, our adoption of the Debye–Hückel theory and implicit ions may be insufficient for electrostatic interactions, especially when considering the impact of divalent ions such as Magnesium⁶⁶. Finally, as shown here and by Collepardo-Guevara et al.²³, histone tail secondary structure changes upon post-translational modifications can impact chromatin stability as well and must be approached with caution.

Methods

Near-atomistic modeling of chromatin organization. We applied a near-atomistic model to study chromatin with increased chemical details. One bead per amino acid and three sites per nucleotide were employed to describe protein and DNA molecules, respectively, leading to a system of 8058 coarse-grained beads in size. Following the same procedure as in structure-based models^{67,68}, contact potentials derived from the crystal structure¹⁰ were introduced for amino acids within individual histone octamers to stabilize the tertiary structure. No specific interactions between proteins from different nucleosomes were included unless otherwise specified. The 3SPN2.C force field developed by de Pablo and coworkers^{69–71} was used to model the DNA molecule. Protein–DNA interactions were described with the Debye–Hückel potential at a salt concentration of 150 mM and the Lennard-Jones potential for excluded volume effect. More details on the model setup and force field parameters can be found in the SI.

We anticipate the improved resolution compared to mesoscopic models to be beneficial at modeling electrostatic and protein–protein interactions within and among nucleosomes. Similar models have been used extensively to study the dynamics and stability of single nucleosomes^{41,72–76}. They were shown to reproduce the energetic cost of nucleosomal DNA unwinding^{72,74}, the dependence of the energetic cost on applied tension⁷³, and the sequence-specific DNA binding strength to the histone octamer⁷⁷. The de Pablo group recently showed that near-atomistic modeling could reproduce the binding strength between a pair of nucleosomes measured in DNA origami-based force spectrometer experiments^{41,78}.

To further evaluate the force field’s accuracy on connected nucleosomes, we simulated a pair of di-nucleosomes connected with linker DNAs of 50 and 55 bp in length and monitored the unwrapping of nucleosomal DNA. As shown in Supplementary Fig. 10, these simulations succeeded in capturing the subtle difference between the two systems. In particular, consistent with Förster resonance energy transfer (FRET) experiments⁷⁹, simulations predicted a more significant unwrapping for the di-nucleosome with 55-bp-long linker DNA. The populations of the unwrapped state for the two di-nucleosomes are in good agreement with experimental values as well.

Simulation details. To improve the near-atomistic model’s efficiency for chromatin simulations, we grouped the ordered part of histone proteins and the inner 73 bp of nucleosomal DNA together as rigid bodies (see Supplementary Fig. 11a). Positions and velocities of all the atoms within each rigid body were updated together such that the body moves and rotates as a single entity. Disordered histone

tails, outer nucleosomal DNA, and linker DNA remained flexible and no restrictions were applied on their conformational dynamics. Our partition of the rigid and flexible parts is motivated by findings from prior studies. In particular, the outer DNA does not bind tightly to histone proteins and can unwind spontaneously⁸⁰, while the inner layer is much more stable, and its unwinding is prohibited by a sizeable energetic barrier^{74,81}. Furthermore, the histone core remains relatively stable during DNA unwinding^{72,82}. As shown in Supplementary Fig. 11b, the rigid body treatment has minimal impact on the model’s accuracy in simulating inter-nucleosome distances.

Molecular dynamics simulations were carried out with the LAMMPS software package⁸³. A time step of 5 fs was used for configurational exploration and 10 fs for mean force calculations. We note that the near-atomistic model’s timescale is not carefully calibrated, and one step here can correspond to a much longer time step in all-atom simulations⁸⁴. The Nosé–Hoover thermostat was applied separately to the rigid and flexible parts of the system to maintain the simulations at a temperature of 300 K with a damping coefficient of 1 ps. All simulations were carried out in a cubic box with sides of 200 nm.

Driving chromatin folding with advanced sampling techniques. To explore the folding landscape of the tetra-nucleosome and collect a diverse set of conformations, we combined metadynamics with TAMM to bias the simulations along two collective variables R_g and Q using PLUMED⁸⁵. While meta-dynamics enhances barrier crossing by introducing memory dependent potentials to penalize the system from revisiting the sampled region of phase space, TAMM accelerates rare transitions with fictitious dynamics of collective variables performed at high temperatures. Combining the two techniques has been shown to result in superior efficiency for conformational sampling compared with simulations conducted using only one of them⁸⁶. Implementation details of the biases can be found in the SI.

We carried out 20 independent biased simulations. A total of at least 50 million steps were carried out for each simulation (see Supplementary Fig. 2), and we saved the configurations along the trajectories every 5000 steps. In total, 293,291 tetra-nucleosome configurations were recorded. Initial configurations of these simulations were sampled from a 40×10^6 step long trajectory in which the tetra-nucleosome was restricted to open configurations with 1–3 and 2–4 nucleosome distances larger than 30 nm.

Learning the six dimensional free energy surface using neural networks. We identified 10,000 representative tetra-nucleosome configurations from the biased exploratory trajectories using the K -means clustering algorithm. We then carried out additional restrained MD simulations to estimate the mean forces⁸⁷ at each one of these representative configuration, \mathbf{d}_o , as

$$\mathbf{F}(\mathbf{d}_o) = \langle k(\mathbf{d} - \mathbf{d}_o) \rangle. \quad (2)$$

The angular bracket represents averaging over the structural ensemble derived from the energy function of the restrained tetra-nucleosome $U(\mathbf{r}) + k/2 \cdot (\mathbf{d} - \mathbf{d}_o)^2$, where $U(\mathbf{r})$ is the unbiased potential. Finally, we parameterized the free energy surface $A(\mathbf{d})$ using a neural network that minimizes the difference between its gradient $-\nabla A(\mathbf{d}_o)$ and the mean forces $\mathbf{F}(\mathbf{d}_o)$. We further took advantage of the underlying symmetry of the system when designing the neural network. In particular, since the free energy is a function of internal coordinates, it should be independent of the ordering and indices of the nucleosomes. Therefore, the following two sets of distances, $\mathbf{d} = (d_{12}, d_{13}, d_{14}, d_{23}, d_{24}, d_{34})$ and $\tilde{\mathbf{d}} = (d_{34}, d_{24}, d_{14}, d_{23}, d_{13}, d_{12})$, obtained from reordering the indices of the nucleosomes, should have approximately the same free energy, i.e., $A(\mathbf{d}) \approx A(\tilde{\mathbf{d}})$. They are approximately equal because the DNA sequence used in our simulations is not palindromic. To enforce the symmetry in the neural network’s output, we converted the distances into symmetrical features such that $\mathbf{s}(\mathbf{d}) = \mathbf{s}(\tilde{\mathbf{d}})$ and parameterized the free energy as a function of \mathbf{s} (Fig. 3a). More details on the architecture and parameterization of the neural network can be found in the SI.

Uncovering folding pathways with the finite temperature string method. The finite temperature string (FTS) method⁸⁰ was used to determine chromatin folding pathways. The output of this method is a path (principal curve) that connects the reactant with the product while taking into account global features of the free energy surface and entropic effects of the transition. It is particularly useful when the free energy landscape is rugged, as is the case here, because it can average over fluctuations on the scale of $k_B T$ or smaller to reveal large scale characteristics of the landscape. The ruggedness is potentially an intrinsic feature of the tetra-nucleosome system, but it may also arise from statistical noise in mean force calculations and neural network parameterization. To avoid biases that might arise from initialization, we carried out a total of 70 string calculations starting from different initial path configurations for a comprehensive exploration of the free energy surface. Details on the implementation of the string method can be found in the SI.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

3D coordinates of representative tetra-nucleosome structures are available in the GitHub repository (DOI:10.5281/ZENODO.4446218).

Code availability

The computer code used in this study is available upon reasonable request to the corresponding author.

Received: 4 July 2020; Accepted: 22 January 2021;

Published online: 17 February 2021

References

- Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661–678 (2016).
- Schlick, T., Hayes, J. & Grigoryev, S. Toward convergence of experimental studies and theoretical modeling of the chromatin fiber. *J. Biol. Chem.* **287**, 5183–5191 (2012).
- Luger, K., Dechassa, M. L. & Tremethick, D. J. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. Cell Biol.* **13**, 436–47 (2012).
- McGinty, R. K. & Tan, S. Nucleosome structure and function. *Chem. Rev.* **115**, 2255–2273 (2015).
- Finch, J. T. & Klug, A. Solenoidal model for superstructure in chromatin. *Proc. Natl Acad. Sci. USA* **73**, 1897–1901 (2006).
- Woodcock, C. The higher-order structure of chromatin: evidence for a helical ribbon arrangement. *J. Cell Biol.* **99**, 42–52 (1984).
- Woodcock, C. L., Grigoryev, S. A., Horowitz, R. A. & Whitaker, N. A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. *Proc. Natl Acad. Sci. USA* **90**, 9021–9025 (1993).
- Horowitz, R. A., Agard, D. A., Sedat, J. W. & Woodcock, C. L. The three-dimensional architecture of chromatin in situ: electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon. *J. Cell Biol.* **125**, 1–10 (1994).
- Song, F. et al. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* **344**, 376–380 (2014).
- Schalch, T., Duda, S., Sargent, D. F. & Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436**, 138–141 (2005).
- Fierz, B. & Poirier, M. G. Biophysics of chromatin dynamics. *Annu. Rev. Biophys.* **48**, 321–345 (2019).
- Joti, Y. et al. Chromosomes without a 30-nm chromatin fiber. *Nucleus* **3**, 404–410 (2012).
- Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* **112**, E6456–E6465 (2015).
- Ou, H. D. et al. ChromEMT: visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science*. <https://doi.org/10.1126/science.aag0025> (2017).
- Cai, S., Böck, D., Pilhofer, M. & Gan, L. The in situ structures of mono-, di-, and trinucleosomes in human heterochromatin. *Mol. Biol. Cell* **29**, 2450–2457 (2018).
- Ohno, M. et al. Sub-nucleosomal genome structure reveals distinct nucleosome folding motifs. *Cell* **176**, 520–534.e25 (2019).
- Depken, M. & Schiessel, H. Nucleosome shape dictates chromatin fiber structure. *Biophys. J.* **96**, 777–784 (2009).
- Koslover, E. F., Fuller, C. J., Straight, A. F. & Spakowitz, A. J. Local geometry and elasticity in compact chromatin structure. *Biophys. J.* **99**, 3941–3950 (2010).
- Schiessel, H., Gelbart, W. M. & Bruinsma, R. DNA folding: structural and mechanical properties of the two-angle model for chromatin. *Biophys. J.* **80**, 1940–1956 (2001).
- Nishikawa, J. I. & Ohyama, T. Selective association between nucleosomes with identical DNA sequences. *Nucleic Acids Res.* **41**, 1544–1554 (2013).
- Kepper, N., Foethke, D., Stehr, R., Wedemann, G. & Rippe, K. Nucleosome geometry and internucleosomal interactions control the chromatin fiber conformation. *Biophys. J.* **95**, 3692–3705 (2008).
- Ozer, G., Luque, A. & Schlick, T. The chromatin fiber: multiscale problems and approaches. *Curr. Opin. Struct. Biol.* **31**, 124–139 (2015).
- Collepardo-Guevara, R. et al. Chromatin unfolding by epigenetic modifications explained by dramatic impairment of internucleosome interactions: a multiscale computational study. *J. Am. Chem. Soc.* **137**, 10205–10215 (2015).
- Arya, G. & Schlick, T. A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments. *J. Phys. Chem. A* **113**, 4045–4059 (2009).
- Perišić, O. & Schlick, T. Dependence of the linker histone and chromatin condensation on the nucleosome environment. *J. Phys. Chem. B* **121**, 7823–7832 (2017).
- Bascom, G. D., 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* **121**, 3882–3894 (2017).
- Bascom, G. D., Sanbonmatsu, K. Y. & Schlick, T. Mesoscale modeling reveals hierarchical looping of chromatin fibers near gene regulatory elements. *J. Phys. Chem. B* **120**, 8642–8653 (2016).
- Bascom, G. D., Myers, C. G. & Schlick, T. Mesoscale modeling reveals formation of an epigenetically driven HOXC gene hub. *Proc. Natl Acad. Sci. USA* **116**, 4955–4962 (2019).
- Parashar, M. et al. Nucleosome positioning and composition modulate in silico chromatin flexibility. *J. Phys. Condens. Matter* **27**, 064112 (2015).
- Kulaeva, O. I. et al. Internucleosomal interactions mediated by histone tails allow distant communication in chromatin. *J. Biol. Chem.* **287**, 20248–20257 (2012).
- Norouzi, D., Katebi, A., Cui, F. & B. Zhurkin, V. Topological diversity of chromatin fibers: interplay between nucleosome repeat length, DNA linking number and the level of transcription. *AIMS Biophys.* **2**, 613–629 (2015).
- Norouzi, D. & Zhurkin, V. B. Topological polymorphism of the two-start chromatin fiber. *Biophys. J.* **108**, 2591–2600 (2015).
- Bass, M. V., Nikitina, T., Norouzi, D., Zhurkin, V. B. & Grigoryev, S. A. Nucleosome spacing periodically modulates nucleosome chain folding and DNA topology in circular nucleosome arrays. *J. Biol. Chem.* <https://doi.org/10.1074/jbc.RA118.006412> (2019).
- Norouzi, D. & Zhurkin, V. B. Dynamics of chromatin fibers: comparison of Monte Carlo simulations with force spectroscopy. *Biophys. J.* **115**, 1644–1655 (2018).
- de Jong, B. E., Brouwer, T. B., 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.* **115**, 1848–1859 (2018).
- Lequieu, J., Córdoba, A., Moller, J. & De Pablo, J. J. 1CPN: a coarse-grained multi-scale model of chromatin. *J. Chem. Phys.* **150**, 215102 (2019).
- Cai, W., Shao, X., Fu, H., Chen, H. & Chipot, C. ELF: an extended-Lagrangian free energy calculation module for multiple molecular dynamics engines. *J. Chem. Inf. Model.* **58**, 1315–1318 (2018).
- Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl Acad. Sci. USA* **99**, 12562–12566 (2002).
- Maragliano, L. & Vanden-Eijnden, E. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chem. Phys. Lett.* **426**, 168–175 (2006).
- Abrams, J. B. & Tuckerman, M. E. Efficient and direct generation of multidimensional free energy surfaces via adiabatic dynamics without coordinate transformations. *J. Phys. Chem. B* **112**, 15742–15757 (2008).
- Moller, J., Lequieu, J. & De Pablo, J. J. The free energy landscape of internucleosome interactions and its relation to chromatin fiber structure. *ACS Cent. Sci.* **5**, 341–348 (2019).
- Souaille, M. & Roux, B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comp. Phys. Comm.* **135**, 40–57 (2001).
- Schneider, E., Dai, L., Topper, R. Q., Drechsel-Grau, C. & Tuckerman, M. E. Stochastic neural network approach for learning high-dimensional free energy surfaces. *Phys. Rev. Lett.* **119**, 150601 (2017).
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
- Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
- García-Saez, I. et al. Structure of an H1-bound 6-nucleosome array reveals an untwisted two-start chromatin fiber conformation. *Mol. Cell* **72**, 902–915.e7 (2018).
- Maeshima, K., Ide, S. & Babokhov, M. Dynamic chromatin organization without the 30-nm fiber. *Curr. Opin. Cell Biol.* **58**, 95–104 (2019).
- Brandner, A., Schüller, A., Melo, F. & Pantano, S. Exploring DNA dynamics within oligonucleosomes with coarse-grained simulations: SIRAH force field extension for protein-DNA complexes. *Biochem. Biophys. Res. Commun.* **498**, 319–326 (2018).
- E, W., Ren, W. & Vanden-Eijnden, E. Finite temperature string method for the study of rare events. *J. Phys. Chem. B* **109**, 6688–6693 (2005).
- Vanden-Eijnden, E. & Venturoli, M. Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.* **130**, 194103 (2009).

51. Kruihof, M. et al. Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. *Nat. Struct. Mol. Biol.* **16**, 534–540 (2009).
52. Miyazawa, S. & Jernigan, R. L. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* **256**, 623–644 (1996).
53. Latham, A. P. & Zhang, B. Maximum entropy optimized force field for intrinsically disordered proteins. *J. Chem. Theory Comput.* **16**, 773–781 (2020).
54. Dang, W. et al. Histone H4 lysine 16 acetylation regulates cellular lifespan. *Nature* **459**, 802–807 (2009).
55. Potoyan, D. A. & Papoian, G. A. Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics. *J. Am. Chem. Soc.* **133**, 7405–7415 (2011).
56. Potoyan, D. A. & Papoian, G. A. Regulation of the H4 tail binding and folding landscapes via Lys-16 acetylation. *Proc. Natl Acad. Sci. USA.* **109**, 17857–17862 (2012).
57. Dhar, S., Gursoy-Yuzugullu, O., Parasuram, R. & Price, B. D. The tale of a tail: histone H4 acetylation and the repair of DNA breaks. *Philos. Trans. R. Soc. B Biol. Sci.* <https://doi.org/10.1098/rstb.2016.0284> (2017).
58. Tremethick, D. J. Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* **128**, 651–654 (2007).
59. Maeshima, K., Hihara, S. & Eltsov, M. Chromatin structure: does the 30-nm fibre exist in vivo? *Curr. Opin. Cell Biol.* **22**, 291–297 (2010).
60. Robinson, P. J. J., Fairall, L., Huynh, V. A. T. & Rhodes, D. EM measurements define the dimensions of the “30-nm” chromatin fiber: evidence for a compact, interdigitated structure. *Proc. Natl Acad. Sci. USA* **103**, 6506–6511 (2006).
61. Nikitina, T., Norouzi, D., Grigoryev, S. A. & Zhurkin, V. B. DNA topology in chromatin is defined by nucleosome spacing. *Sci. Adv.* **3**, e1700957 (2017).
62. Moller, J. & de Pablo, J. J. Bottom-up meets top-down: the crossroads of multiscale chromatin modeling. *Biophys. J.* **118**, 2057–2065 (2020).
63. Qi, Y. & Zhang, B. Predicting three-dimensional genome organization with chromatin states. *PLoS Comput. Biol.* **15**, e1007024 (2019).
64. Qi, Y. et al. Data-driven polymer model for mechanistic exploration of diploid genome organization. *Biophys. J.* **119**, 1905–1916 (2020).
65. Widom, J. Role of DNA sequence in nucleosome stability and dynamics. *Q. Rev. Biophys.* **34**, 269–324 (2001).
66. Hayes, R. L. et al. Generalized manning condensation model captures the RNA ion atmosphere. *Phys. Rev. Lett.* **114**, 1–6 (2015).
67. Whitford, P. C. et al. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins* **75**, 430–441 (2009).
68. Clementi, C., Nymeyer, H. & Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and ‘en-route’ intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **298**, 937–953 (2000).
69. Freeman, G. S., Hinckley, D. M. & De Pablo, J. J. A coarse-grain three-site-per-nucleotide model for DNA with explicit ions. *J. Chem. Phys.* <https://doi.org/10.1063/1.3652956> (2011).
70. Hinckley, D. M., Freeman, G. S., Whitmer, J. K. & De Pablo, J. J. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* <https://doi.org/10.1063/1.4822042> (2013).
71. Freeman, G. S., Lequieu, J. P., Hinckley, D. M., Whitmer, J. K. & De Pablo, J. J. DNA shape dominates sequence affinity in nucleosome formation. *Phys. Rev. Lett.* <https://doi.org/10.1103/PhysRevLett.113.168101> (2014).
72. Zhang, B., Zheng, W., Papoian, G. A. & Wolynes, P. G. Exploring the free energy landscape of nucleosomes. *J. Am. Chem. Soc.* **138**, 8126–8133 (2016).
73. Lequieu, J., Córdoba, A., Schwartz, D. C. & De Pablo, J. J. Tension-dependent free energies of nucleosome unwrapping. *ACS Cent. Sci.* **2**, 660–666 (2016).
74. Parsons, T. & Zhang, B. Critical role of histone tail entropy in nucleosome unwinding. *J. Chem. Phys.* **150**, 185103 (2019).
75. Kenzaki, H. & Takada, S. Partial unwrapping and histone tail dynamics in nucleosome revealed by coarse-grained molecular simulations. *PLoS Comput. Biol.* **11**, 1–20 (2015).
76. Watanabe, S., Mishima, Y., Shimizu, M., Suetake, I. & Takada, S. Interactions of HP1 bound to H3K9me3 dinucleosome by molecular simulations and biochemical assays. *Biophys. J.* **114**, 2336–2351 (2018).
77. Freeman, G. S., Lequieu, J. P., Hinckley, D. M., Whitmer, J. K. & De Pablo, J. J. DNA shape dominates sequence affinity in nucleosome formation. *Phys. Rev. Lett.* **113**, 168101 (2014).
78. Funke, J. J., Ketterer, P., Lieleg, C., Schunter, S., Korber, P. & Dietz, H. Uncovering the forces between nucleosomes using DNA origami. *Sci. Adv.* <https://doi.org/10.1126/sciadv.1600974> (2016).
79. Buning, R., Kropff, W., Martens, K. & van Noort, J. spFRET reveals changes in nucleosome breathing by neighboring nucleosomes. *J. Phys. Condens. Matter* **27**, 064103 (2015).
80. Li, G., Levitus, M., Bustamante, C. & Widom, J. Rapid spontaneous accessibility of nucleosomal DNA. *Nat. Struct. Mol. Biol.* **12**, 46–53 (2005).
81. Brower-Toland, B. D. et al. Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA. *Proc. Natl Acad. Sci. USA* **99**, 1960–1965 (2002).
82. Böhm, V. et al. Nucleosome accessibility governed by the dimer/tetramer interface. *Nucleic Acids Res.* **39**, 3093–3102 (2011).
83. Plimpton, S. & National, L. S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–42 (1995).
84. Tan, C. et al. Modeling structural dynamics of biomolecular complexes by coarse-grained molecular simulations. *Acc. Chem. Res.* **48**, 3026–3035 (2015).
85. Bonomi, M. et al. PLUMED: a portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **180**, 1961–1972 (2009).
86. Chen, M., Cuendet, M. A. & Tuckerman, M. E. Heating and flooding: a unified approach for rapid generation of free energy surfaces. *J. Chem. Phys.* **137**, 024102 (2012).
87. Maragliano, L. & Vanden-Eijnden, E. Single-sweep methods for free energy calculations. *J. Chem. Phys.* **128**, 184110 (2008).

Acknowledgements

This work was supported by the National Institutes of Health (Grant 1R35GM133580). We thank Dr. Collepardo-Guevara for sharing the atomistic structure of histone H4 tail, Dr. van Noort for providing the FRET data on di-nucleosomes, and Dr. Pantano for sharing the simulation trajectory of tetra-nucleosome.

Author contributions

X.D., X.L., and B.Z. designed and performed research, analyzed data, and wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-21377-z>.

Correspondence and requests for materials should be addressed to B.Z.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021