ACS central science

http://pubs.acs.org/journal/acscii

Research Article

Tetranucleosome Interactions Drive Chromatin Folding

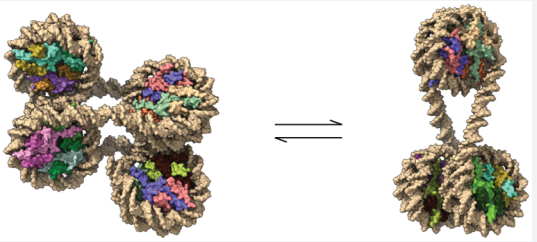
Walter Alvarado, Joshua Moller, Andrew L. Ferguson,* and Juan J. de Pablo*

Cite This: *ACS Cent. Sci.* 2021, 7, 1019−1027

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The multiscale organizational structure of chromatin in eukaryotic cells is instrumental to DNA transcription, replication, and repair. At mesoscopic length scales, nucleosomes pack in a manner that serves to regulate gene expression through condensation and expansion of the genome. The particular structures that arise and their respective thermodynamic stabilities, however, have yet to be fully resolved. In this study, we combine molecular modeling using the 1CPN mesoscale model of chromatin with nonlinear manifold learning to identify and characterize the structure and free energy of metastable states of short chromatin segments comprising between 4- and 16-nucleosomes. Our results reveal the formation of two previously characterized tetranucleosomal conformations, the "α-tetrahedron" and the "β-rhombus", which have been suggested to play an important role in the accessibility of DNA and, respectively, induce local chromatin compaction or elongation. The spontaneous formation of these motifs is potentially responsible for the slow nucleosome dynamics observed in experimental studies. Increases of the nucleosome repeat length are accompanied by more pronounced structural irregularity and flexibility and, ultimately, a dynamic liquid-like behavior that allows for frequent structural reorganization. Our findings indicate that tetranucleosome motifs are intrinsically stable structural states, driven by local internucleosomal interactions, and support a mechanistic picture of chromatin packing, dynamics, and accessibility that is strongly influenced by emergent local mesoscale structure.



INTRODUCTION

Chromatin is the complex of DNA, RNA, and proteins found in eukaryotic cell nuclei. Chromatin's dynamic, multiscale structure is central to the regulation of transcription, replication, and DNA repair.[1,2] The basic building block of eukaryotic chromatin is the nucleosome, a disk-like DNA−protein complex of approximately 146 basepairs (bp) of DNA wrapped around a protein complex known as the histone octomer.[2] These small and positively charged proteins bind strongly to the negatively charged DNA. Each nucleosome contains four core histone proteins (H2A, H2B, H3, and H4) that are found in equal proportions in cells.[3] Nucleosomes connect to adjacent nucleosomes through a segment of linker DNA whose combined length with core DNA is referred to as a nucleosome repeat length (NRL). Nucleosome repeat lengths exhibit a distribution centered around 180 bp, depending on organism, cell type, or loci in a given cell type.[1]

It has long been argued that nucleosomes condense into a 30 nm thick chromatin fiber that progressively folds to form mitotic chromosomes and interphase nuclei.[4] The character and properties of higher-order chromatin structures *in vivo* also continue to be the source of debate.[4] Experimental techniques such as cryo-electron microscopy, X-ray scattering, scanning transmission electron microscopy, chromosome conformation capture (Hi-C), and super-resolution STORM (stochastic optical reconstruction microscopy) indicate that chromatin in living cells is predominantly comprised of highly irregular structures.[5−10] The emerging view that chromatin organization is dominated by intermediate scale assemblies (∼3−10-nucleosomes) is supported by results from various chromosome conformation capture techniques and electron cryotomography, which have revealed the existence of clusters comprising only a few nucleosomes that may play a role in chromatin biology.[6,11−13]

Tetranucleosomes are proposed to be functional and structural units that regulate gene expression.[6,13,14] They have been crystallizable at NRLs of 157 and 167 bp and observed in longer chromatin fibers of 12 tandem repeats of 177 and 187 bp from cryo-EM.[15,16] Cryo-EM experimental evidence of nucleosome stacking, with two-start tetranucleosomes stacked on top of each other, has also been reported in human cell lines.[17] Two tetranucleosome motifs have recently been proposed to serve regulatory functions: the α-tetrahedron and the β-rhombus.[13] The all-atom structure of the α-tetrahedron and β-rhombus motifs are shown in Figure 1A,B, respectively.[15,18] In Figure 1C,D, we provide a representation of these two motifs that relies on the 1CPN model. By combining nucleosome-resolved Hi-C measurements with simulated annealing−molecular dynamics (SA-MD), the

Received: January 18, 2021
Published: May 7, 2021

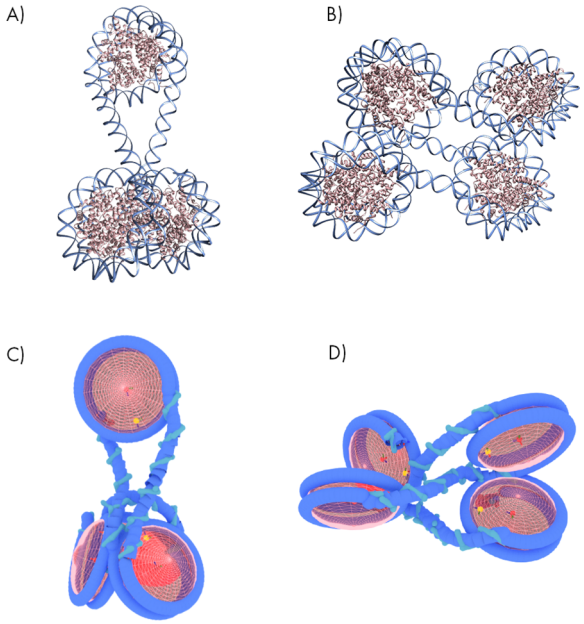


Figure 1. Recent experimental studies have suggested the existence of two basic secondary structural motifs in chromatin involved in epigenetic regulatory function: α-tetrahedron (A, C) and β-rhombus (B, D).[13] α-Tetrahedron motifs have been suggested to contribute to local chromatin compaction. β-Rhombus conformations promote an open-ladder-like chromatin structure that forms linear elongated aggregates. (A, B) Representative PDB structures from cryo-EM and X-ray crystallography of tri- and tetranucleosomes.[15,18] (B, C) Representative structures from 1CPN simulations of tetranucleosome fibers.

orientation of nucleosome pairs were analyzed by focusing on which nucleosome ends were ligated to one another in yeast genome.[13] Given their location near regulatory regions, within genes (α-tetrahedron) and at gene ends (β-rhombus), these specific folding motifs may serve a regulatory function and seem consistent with what is already established about chromatin organization, such as nucleosomal "clutches" and "topologically associated domains" (TADs).[13] It is unclear, however, whether these motifs arise due to an orchestrated set of interactions governed by protein partners or if they are intrinsically stable structural states of the nucleosome. Chemical fixation with compounds such as glutaraldehyde and methanol, which are employed in EM and super-resolution light microscopy, have benefited the study of chromatin but introduce the formation of artificial structures not present under biological conditions.[19] The adherence of free molecules, facilitated by the increase in hydrophobic interactions, forms long and continuous artificial structures which have become an issue in newer imaging experiments such as STORM.[19] Furthermore, long imaging times may mask important short-time nucleosome dynamics and mischaracterize small nucleosome domains.[19] Resolving the emergent structure and stability of tetranucleosome motifs as a function of NRL is of critical biological interest in understanding chromatin packing, DNA accessibility, and epigenetics.

In this work we relied on large-scale simulations of various chromatin fiber lengths using the recently developed 1CPN mesoscale model of chromatin (1-cylinder-per-nucleosome).[20] This model is structured around a rigorous multiscale approach where free energies are derived from an established and extensively validated model of the nucleosome and mapped onto a single anisotropic topology.[21,22] The 1CPN model has accurately reproduced a range of chromatin properties, including nucleosome−nucleosome free energy interactions, nucleosome unwrapping free energies, and sedimentation coefficients of short chromatin fibers,[20] and its computational efficiency enables access to length and time scales not currently possible with atomistic models.

To investigate the organizational properties of chromatin, we used nonlinear manifold learning, and identified the metastable states and collective motions that mediate transitions between such states (see Diffusion Maps in Materials and Methods). We investigated the formation of two previously characterized tetranucleosomal conformations (α-tetrahedron and β-rhombus motifs) that play an important role in the accessibility of DNA. Our results show that the relative thermodynamic stability of these two motifs are intrinsic properties of the chromatin chain independent of external cellular factors. Our findings and trends with NRL are consistent with experimental observations and provide new mechanistic insight into mesoscale structure formation and the hierarchical structure and conformational transitions of chromatin.

RESULTS AND DISCUSSION

We present the results of 1CPN simulations of three representative chromatin fibers with NRL values of 157, 187, and 197. These repeat lengths were selected due to their structures being available from X-ray and cyro-EM, thereby allowing direct comparisons to experiment.[14,15,23] For each NRL system we considered chromatic fibers containing 4-, 8-, and 16-nucleosomes and identified the metastable states and oligonucleosome organization using diffusion map dimensionality reduction. Simulations were conducted from an initial elongated conformation; convergence was considered achieved when the root-mean-square deviation (RMSD) with respect to the initial frame approached a steady state (see Simulating the Chromatin Fiber in Material and Methods.) Where possible, we make contact with experimental measurements.

NRL 157. Diffusion maps are a nonlinear dimensionality reduction technique that extracts collective variables (CVs) characterizing the large-scale collective motions of molecular systems.[24] These CVs are computed as the eigenvectors of a particular transition matrix characterizing the similarity of various system configurations. The appropriate number of eigenvectors to retain in the low-dimensional diffusion map embeddings can be determined by identifying a gap in the eigenvalue spectrum of the transition matrix that demarcate the leading modes characterizing the important large-scale collective motions from those associated with smaller amplitude conformational changes. Often the spectrum may exhibit multiple gaps corresponding to a hierarchical splitting into groups of collective modes ordered by decreasing amplitude. In all systems studied in this work we observe the leading spectral gap to occur after the first nontrivial eigenvalue (Supporting Information Figure S1). This suggests that the conformational dynamics of the system are dominated by a single CV, and we invariably correlate the leading eigenvector with global contraction or expansion of the fiber. In many cases, we also find the second eigenvector to also be informative and we can often correlate it with some additional aspect of the fiber conformational motions as discussed below. Many of the eigenvalue spectra exhibit additional spectral gaps indicative of a hierarchical partitioning of CVs, but our

https://doi.org/10.1021/acscentsci.1c00085
ACS Cent. Sci. 2021, 7, 1019−1027

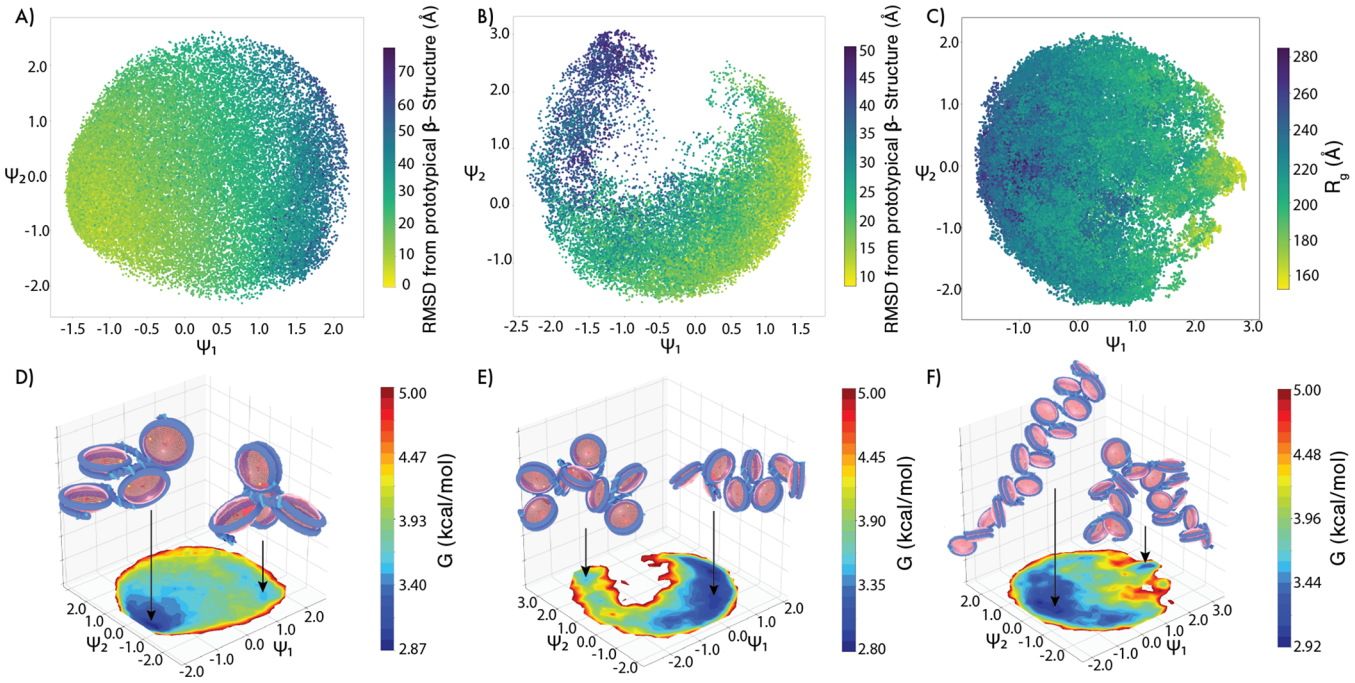


Figure 2. Two tetranucleosome motifs, the α-tetrahedron and the β-rhombus, respectively induce local chromatin compaction or the formation of elongated aggregates at short nucleosome repeat lengths of 157. (A, D) In 4-nucleosome fibers we observe a global free energy minimum at ($\psi_1 \approx -1.8$; $\psi_2 \approx -0.7$) and a metastable minimum at ($\psi_1 \approx 1.3$; $\psi_2 \approx -1.5$) that lies ∼2.21 kcal/mol higher in free energy. The global minimum contains β-rhombus-like structures, while the local minimum contains compact α-tetrahedron-like packings. (B, E) The 8-nucleosome fiber possesses two metastable states with a large basin containing the global free energy minimum residing at ($\psi_1 \approx 0.2$; $\psi_2 \approx -0.8$) and a weak local minimum at ($\psi_1 \approx -1.5$; $\psi_2 \approx 2.8$). The more compact structures reside in the local minimum and contain predominantly α-tetrahedron-like packings, whereas the more extended in the global free energy minimum are β-rhombus-like. (**C**, **F**) The 16-nucleosome fiber possesses a global minimum at ($\psi_1 \approx -0.99$; $\psi_2 \approx -0.23$) containing the β-rhombus-like fiber that is ∼0.83 kcal/mol more stable than the local minimum at ($\psi_1 \approx 2.44$; $\psi_2 \approx -0.01$) containing the α-tetrahedron-like fiber.

exploration of the higher-order eigenvectors did not show any obvious correlation with putative physical order parameters or expose any additional interpretability beyond what we were able to extract from the 2D embeddings. For this reason, in addition to visual clarity and accessibility, all embeddings in this work are 2D in nature.

In Figure 2A−C we present 2D embeddings of the 95,000 simulation snapshots of the chromatin fibers with NRL 157 into the two leading diffusion map CVs (ψ_1,ψ_2). To provide some physical interpretability of the diffusion map embeddings, we computed the RMSD of each contiguous stretch of 4-nucleosomes relative to a prototypical α-/β-structure motif by passing a 4-nucleosome sliding window along the fiber. Dividing by the number of windows defines an order parameter by which to measure the propensity of the complete fiber for either motif. In the 4- and 8-nucleosome fibers, a Pearson correlation analysis reveals the α-/β-structure preference to be strongly correlated with the leading diffusion map CV ($\rho^{(4)}_{\alpha,\psi_1} = -0.75$, $\rho^{(4)}_{\beta,\psi_1} = 0.87$; $\rho^{(8)}_{\alpha,\psi_1} = 0.68$, $\rho^{(8)}_{\beta,\psi_1} = -0.72$). For the 16-nucleosome fiber, the leading CV is strongly correlated with the radius of gyration, R_g ($\rho^{(16)}_{R_g,\psi_1} = -0.74$). The strong correlation with radius of gyration indicates that the leading CV extracted by diffusion maps is strongly associated with the compaction and expansion of the 16-nucleosome fiber. Inspection of the molecular configurations of the elongated and compact fibers in Figure 2C reveals that the elongated fiber contains a high degree of β motif character whereas the compact fiber is composed of primarily α motifs. For the systems studied in this work, it is a general trend that the leading CVs identified by diffusion maps tend to correspond to local α/β motif character for shorter NRL values and fewer nucleosomes *versus* more global descriptors such as the fiber radius of gyration or end-to-end distance for longer NRL values and more nucleosomes. This transition can be understood as a result of the increased length and flexibility of the fiber, leading to the emergence of more global collective variables that subsume and contain the local packing effects.

To resolve the metastable states of the system, in Figure 2D−F we present the corresponding free energy surfaces (FESs) in the Gibbs free energy, $G(\Psi)$, constructed over the 2D intrinsic manifolds and decorated with representative molecular renderings. In the 4-nucleosome system (Figure 2D), a global free energy minimum is observed at ($\psi_1 \approx -1.8$, $\psi_2 \approx -0.7$), and a metastable minimum is found at ($\psi_1 \approx 1.3$, $\psi_2 \approx -1.5$) lying ∼2.21 kcal/mol higher in free energy. Visualization of the simulation snapshots contained within the local minima of the free energy landscape confirms the results of the Pearson correlation analysis that ψ_1 should separate configurations on the basis of α/β character, with the global minimum at low-ψ_1 containing β-rhombus-like structures of the 4-nucleosomes and the local minimum at high-ψ_1 containing compact α-tetrahedron-like packings. These two motifs were previously reported in several chromosome conformation capture experiments, which have provided evidence for the existence of tri- and tetranucleosome folding motifs as the primary level of organization in yeast genome.[6,13] As mentioned earlier, oligonucleosome motifs may serve an important role as functional elements for processes such as

https://doi.org/10.1021/acscentsci.1c00085
ACS Cent. Sci. 2021, 7, 1019−1027

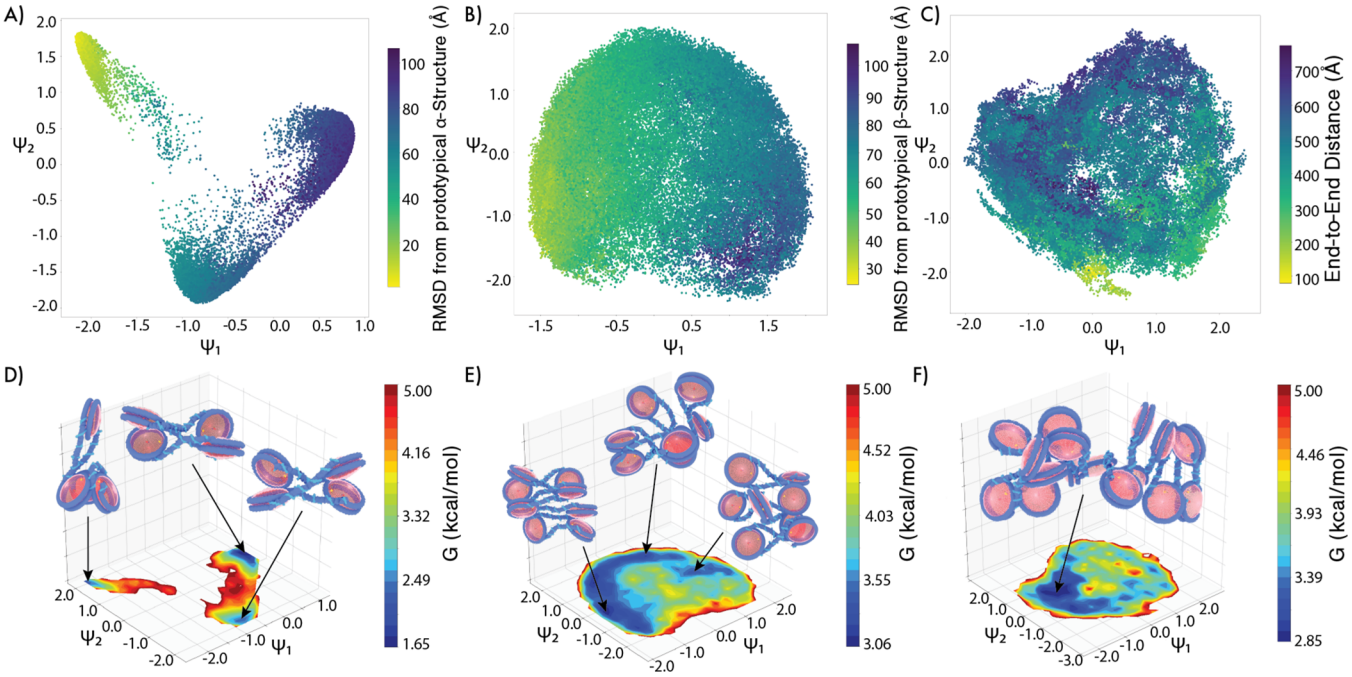


Figure 3. Chromatin fibers of 187 NRL showing a high propensity for small α-tetrahedron and β-rhombus clusters. Nucleosomes engage in short-lived stacking interactions that form distinct tetranucleosome motifs. (A, D) In 4-nucleosome fibers, we observe two β-rhombus clusters corresponding to the two distinct ways in which nucleosomes can arrange themselves to form the motif. The leading diffusion CV correlated with α-structure preference ($\rho^{(4)}_{\alpha,\psi_1} = 0.95$) and the second leading CV correlated with β-preference ($\rho^{(4)}_{\beta,\psi_2} = 0.97$). (B, E) In the 8-nucleosome fibers, α-tetrahedron motifs contribute to local chromatin compaction, while β-rhombus structures resemble the more ladder-like chromatin structure. β-preference is highly correlated with the leading diffusion CV ($\rho^{(8)}_{\beta,\psi_1} = 0.91$). (C, F) A 16-nucleosome fiber exhibits structural heterogeneity comprised of several α-tetrahedron and β-rhombus motifs. The second CV is moderately correlated with the end-to-end distance of the fiber ($\rho_{R_{end}}$, $\psi^{(16)}_2 = 0.56$).

transcription, replication, and DNA repair.[12] In addition, the rearrangement of tetranucleosomes from open/closed states may serve as a potential target for cellular regulation. For example, previous studies have shown many chromatin regulators have a preference to bind to multinucleosome structures *in vitro*.[25] We emphasize that our simulations were conducted from an initial elongated conformation, and we have verified that our sampling of the conformational ensemble is converged with the fiber making multiple transitions over the intrinsic manifold and between the two basins. As such, these structures represent inherent metastable states of the 4-nucleosome NRL 157 system.

In the larger 8-nucleosome system we again observe two metastable states with a large basin for the global free energy minimum at ($\psi_1 \approx 0.2$, $\psi_2 \approx -0.8$), and a weak local minimum at ($\psi_1 \approx -1.5$, $\psi_2 \approx 2.8$) lying ~2.51 kcal/mol higher in free energy (Figure 2E). As for the 4-nucleosome fiber, the global minimum consists of more extended β-rhombus-like arrangements and the local minimum and contain more compact α-tetrahedron-like packings. This analysis reveals that the prototypical α-tetrahedron and β-rhombus tetranucleosome motifs are preserved within the longer 8-nucleosome fiber.

The 16-nucleosome fiber exhibits a similar trend, with a global minimum at ($\psi_1 \approx -0.99$, $\psi_2 \approx -0.23$) consisting of a β-rhombus-like fiber that is ~0.83 kcal/mol more stable than the local minimum at ($\psi_1 \approx 2.44$, $\psi_2 \approx -0.01$) that includes the α-tetrahedron-like fiber (Figure 2F). As was the case for the 4- and 8-nucleosome fibers, the 16-nucleosome β-rhombus-like fiber is less compact, more elongated, and more stable than the α-tetrahedron-like fiber. This is due to the condensation and ordered linear packing of the β-rhombus subunits compared to the relatively poorly packed and disordered chain of α-tetrahedron motifs.

Our results confirm the existence of two tetranucleosome motifs, α-tetrahedron and β-rhombus, that induce local chromatin compaction or form elongated aggregates at short nucleosome repeat lengths of 157, respectively. These computational findings are consistent with experimental Hi-C studies that have shown α-tetrahedron motifs to contribute to local chromatin compaction.[13] Tetranucleosomes that take on an α-tetrahedron conformation introduce kinks and bring disorder into the fiber. β-Rhombus conformations, on the other hand, form an open-ladder-like chromatin structure that packs to form linear elongated aggregates. Our simulations of longer chromatin fibers at an NRL of 157 in the absence of H1, a linker histone protein considered important in maintaining higher-order chromatin structure, adopted a zigzag conformation in agreement with experiments from reconstituted arrays of nucleosomes without H1.[16,26] The crystallization of 157 NRL fibers may be due to the stabilizing interactions between neighboring nucleosomes which intrinsically form these tetranucleosome motifs.[16]

NRL 187. In Figure 3A–C we present the 2D embeddings into the two leading diffusion map CVs for chromatin fiber with NRL 187. For a 4-nucleosome fiber, the leading diffusion CV is strongly correlated with α-structural preference ($\rho^{(4)}_{\alpha,\psi_1} = 0.95$) and the second leading CV correlated with β-preference ($\rho^{(4)}_{\beta,\psi_2} = 0.97$). For an 8-nucleosome fiber, β-preference highly correlated with the leading diffusion CV ($\rho^{(8)}_{\beta,\psi_1} = 0.91$). For the

https://doi.org/10.1021/acscentsci.1c00085
ACS Cent. Sci. 2021, 7, 1019–1027

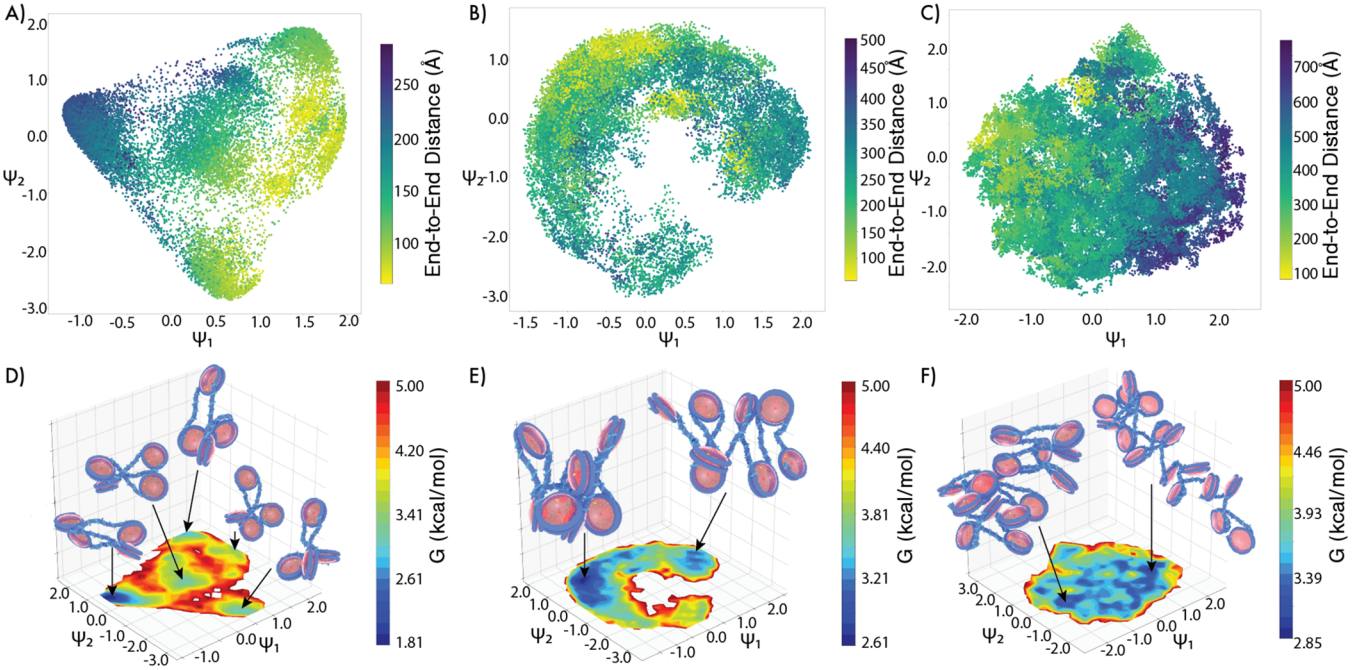


Figure 4. Chromatin fibers at NRL 197 showing high irregularity and flexibility and resembling a "sea of nucleosomes" model. An increase in fiber length is accompanied by an increase in structural irregularity and flexibility. (A, D) In 4-nucleosome fibers, the increase in nucleosome repeat length allows for more complex nucleosome arrangements of α-/β-structures. The leading diffusion map CV is strongly correlated with end-to-end distance of the fiber R_{end} ($\rho^{(4)}_{R_{end},\psi_1} = 0.88$). (B, E) In 8-nucleosome fibers, the α-tetrahedron motifs contribute to local chromatin compaction while β-rhombus structures resemble the more ladder-like chromatin structure. The second CV is moderately correlated with the end-to-end distance ($\rho^{(8)}_{R_{end},\psi_2} = 0.42$). (C, F) In 16-nucleosome fibers, local chromatin motion is isotropic and largely driven by thermal fluctuations. As for the 8-nucleosome case, the second CV is moderately correlated with end-to-end distance ($\rho^{(16)}_{R_{end},\psi_2} = 0.69$).

fiber comprised of 16-nucleosomes, the second CV is moderately correlated with the end-to-end distance of the fiber R_{end} ($\rho_{R_{end}}$, $\psi_2^{(16)} = 0.56$). In Figure 3D−F we present the corresponding FESs and their representative molecular renderings derived from simulation.

In contrast to the 4-nucleosome system with 157 NRL (Figure 2D), the 187 NRL system exhibits two distinct β-rhombus wells in the free energy landscape for the 187 NRL (Figure 3D). This is a consequence of the increased degree of freedom permitted by the longer NRL chain that opens up the availability of additional metastable structures. Specifically, the two wells correspond to two distinct alternative packings of the 4-nucleosomes into the β-rhombus motif. The global minimum at ($\psi_1 \approx 1.0$, $\psi_2 \approx 0.5$) corresponds to a β-rhombus where the N- and (N + 1)-nucleosomes lie diagonally from each other. The local minimum located at ($\psi_1 \approx -0.82$, $\psi_2 \approx -1.62$) contains a second β-rhombus conformation where the N- and (N + 1)-nucleosomes lie directly across from one another and exists ∼3.55 kcal/mol higher in free energy. The local minimum at ($\psi_1 \approx 1.3$, $\psi_2 \approx -1.5$) contains the compact α-tetrahedron motif and lies ∼1.75 kcal/mol higher in free energy. The gap between clusters suggests limited sampling in the interstitial region between the metastable α-/β-states. Comparing our results to simulations of NRLs of 157 (Figure 2D−F), the transition region is destabilized relative to the metastable states for longer NRL fibers, reflecting their larger degree of freedom to stably pack the nucleosomes into the α/β motifs.

The 8- and 16-nucleosome fibers exhibit a relatively broad free energy landscape comprising metastable α/β motifs (Figure 3E,F and Figures S2 and S3). As in the case for chromatin fibers of NRL 157, α-tetrahedron motifs contribute to local chromatin compaction while β-rhombus structures resemble ladder-like chromatin structure. The presence of several β-rhombus structures is consistent with the idea of a tri- or tetranucleosome motif in chromatin fiber folding proposed from Micro-C experiments where the decay of nucleosome−nucleosome interactions as a function of distance revealed no evidence for long-range periodicity in internucleosomal interactions but of short-range structures.[6] In addition, the results of our unbiased simulations where the only driver of the dynamics is thermal fluctuations, suggest that local chromatin motion is spontaneous, as observed from single-particle tracking experiments.[27] The observed local motion of chromatin at longer NRLs are characteristic of fluid-like behavior, suggesting a more irregular and dynamic view of chromatin. Interestingly, variability in motion seems to be a product of the constraints imposed by physical or geometrical factors. Experimental single-nucleosome tracking data in living human cells have suggested a nucleosome's motion dependence on nucleosome−nucleosome interactions.[28] The fiber makes multiple transitions over the intrinsic manifold and between the two metastable states. As such, these nucleosomes engage in short-lived stacking interactions, relative to total simulation time, that form distinct tetranucleosome motifs which we propose may be responsible for slowing chromatin dynamics through transient trapping in metastable states.

NRL 197. We present in Figure 4A−C the 2D embeddings into the two leading diffusion map CVs for chromatin fibers of NRL 197. For the 4-nucleosome fiber, the end-to-end distance R_{end} is strongly correlated with the leading diffusion CV ($\rho^{(4)}_{R_{end},\psi_1}$ = 0.88). In the case of 8- and 16-nucleosomes, each leading

https://doi.org/10.1021/acscentsci.1c00085
ACS Cent. Sci. 2021, 7, 1019−1027

CVs lacked simple physical interpretation but the second CV correlated moderately with end-to-end distance ($\rho^{(8)}_{R_{end},\psi_2} = 0.42$, $\rho^{(16)}_{R_{end},\psi_2} = 0.69$). We observed that for longer NRLs, fibers are highly irregular and flexible, resembling a "sea of nucleosomes" model.[5,8] In this view, nucleosomes are able to interact with local and distant partners, leading to a more dynamic, liquid-like behavior that supports structural plasticity and frequent reorganization.

The FES for a 4-nucleosome 197 NRL fiber in Figure 4D presents the various ways in which either motif can assemble. As in the case of a 187 NRL fiber, the increase in nucleosome repeat length allows for more complex nucleosome arrangements of α-/β-structures. As previous studies have suggested, shorter linker lengths form less flexible fibers, which have the advantage of exposing their DNA to transcription and replication machinery by the displacement of only a few nucleosomes.[29] The lack of a cohesive structure for longer linker lengths suggests that they are less stable and more prone to opening as observed experimentally.[30] At 8- and 16-nucleosomes (Figure 4E,F), the liquid-like behavior of chromatin becomes more apparent. Representative structures for both fibers exhibit structural heterogeneity comprised of several α-tetrahedron and β-rhombus motifs. Our results are in line with ChromEMT experiments which have confirmed the presence of small clusters of ∼2−10-nucleosomes, and a lack of larger organized structures.[31] Our results suggest that these motifs arise due to an orchestrated set of interactions that are mainly driven by nucleosome−nucleosome interactions.

The liquid-like behavior of chromatin is critical to gene expression since these dynamic changes directly affect DNA accessibility. Hi-C experiments have shown a correlation between the structural changes induced by the liquid-like behavior of chromatin and gene expression levels.[32] In addition, the changes in replication timing during cell differentiation have been suggested to involve the fluctuations of TAD structures.[33] Modeling studies using a simple polymer model of chromatin have shown that these TAD domains are susceptible to the liquid-like movement of chromatin.[34] The regulatory role of these intrinsically stable oligonucleosome motifs remains to be determined.

CONCLUSION

By analyzing long simulation trajectories generated by a coarse-grained multiscale chromatin model using nonlinear manifold learning, we resolved the spontaneous and intrinsic formation within the chromatin fiber of α-tetrahedron and β-rhombus motifs—two previously characterized tetranucleosomal conformations that play an important role in the accessibility of DNA. Our analysis showed that both motifs represent metastable states that correspond to local free energy minima and whose formation is not dependent on external factors such as post-translational modifications or protein remodeling complexes. We also characterized the effects that varying linker DNA length and nucleosome orientation have on the formation of these motifs and, by extension, local chromatin compaction. We observed that local chromatin compaction is induced by α-tetrahedron motifs that allow sterically unfavorable conformations to form such as kinks along the fiber. β-Rhombus conformations, on the other hand, were observed to form an open ladder-like chromatin structure that can facilitate DNA accessibility by external machinery such as transcription factors. Longer linker lengths are accompanied by an increase in structural irregularity and flexibility and, ultimately, a dynamic liquid-like "sea of nucleosomes"[5,8] behavior that allows for constant structural reorganization. The lack of a cohesive structure for longer linker lengths suggests that they are less stable and more prone to opening and would require linker histones or some other chromatin architectural protein to fold into either conformation. This finding is in line with experimental observations which have shown that linker histone levels correlate positively with linker DNA length.[35] Here, we considered homogeneous linker lengths and were able to identify the formation of these two unique folding motifs. Hi-C methods, however, have revealed that these conserved motifs exist within heterogeneous chromatin fibers as well.[13] The effects by which variable linker lengths and tetranucleosome motifs influence the free energy landscape of chromatin folding require further investigation.

Taken together, our results suggest a two-state model for the local chromatin structure, wherein extended β-rhombus motifs are more stable than compact α-tetrahedron motifs. It is possible that chromatin remodeling proteins and/or epigenetic status may be responsible for lowering the free energy barrier between the observed states. Future studies will require examining the mechanism by which linker histones and remodeling protein complexes, such as Polycomb, affect the emergence of metastable states. In addition, we are investigating the kinetic rates between each state using Markov-state models (MSM) to elucidate dynamical links between chromatin structure and gene expression.

MATERIALS AND METHODS

Simulating the Chromatin Fiber. Coarse-grained molecular dynamics simulations of chromatin fibers were conducted using the 1CPN model (Figure 1).[20] The model was parametrized by explicit experimental measurements and finer-scaled models of DNA and proteins[20] and incorporates and preserves molecular-level nucleosome physics while enabling kilobase-scale simulations of genomic DNA. Each nucleosome is represented by a single anisotropic site and DNA as a spherical particle at a three basepair-per-bead resolution. We refer interested readers to the detailed descriptions found in prior publications.[20,36,37]

Simulations of 4-, 8-, and 16-nucleosomes at nucleosome repeat lengths 157, 187, and 197 were conducted in a periodic box at a temperature of 300 K and a simulation time step of 60 fs. Solvent was represented implicitly with a Langevin thermostat and salt concentration of 150 mM. Nucleosome interactions were represented by the Zewdie potential, which has been shown to be well-suited for representing internucleosomal interactions.[38] To account for the stabilizing effects of the histone H3 tail of DNA entering and exiting the nucleosome, an additional pairwise interaction was introduced between the dyad and DNA sites. Electrostatic interactions were incorporated using Debye−Hückel theory.[39] Simulations were carried out using the LAMMPS molecular dynamics package.[40] Chromatin fibers were initialized in an extended state with the angle between entering and exiting DNA and the nucleosome at 180°. The root-mean-square deviation with respect to the first given frame achieved a steady state after a ∼100 μs equilibration period. The final configuration from the relaxation procedure was used as the initial starting structure for our production run and analysis. Five 30 μs replicas were conducted totaling 150 μs of simulation time for each

https://doi.org/10.1021/acscentsci.1c00085
ACS Cent. Sci. 2021, 7, 1019−1027

permutation and amounting to a total 1.35 ms of overall simulation time.

Diffusion Maps. Diffusion maps are a nonlinear manifold learning technique that have found extensive applications in generating low-dimensional embeddings of high-dimensional molecular trajectories.[41−43] Assuming that the distance metric used to compare pairs of configurational microstates is a good proxy for short-time kinetic distance and that the conformational dynamics over the state space may be approximated as a diffusion process, the leading collective variables of the diffusion map correspond to the large-scale, high-variance collective motions of the system, and kinetically close configurational microstates are embedded close together.[24] We employ the density-adaptive variant of diffusion maps, which we find to be particularly useful for handling the large inhomogeneities in sampling densities observed in our chromatin simulations.[44] We provide a brief summary of the approach below, but direct the reader to prior publications for mathematical and algorithmic details.[24,41−43]

Pairwise distances, d_{ij}, are calculated between data points in our set, x_i and x_j, which correspond to the RMSD between translationally and rotationally aligned nucleosomal coordinates in frames i and j of the simulation. A Gaussian kernel is applied to d_{ij} to construct a threshold pairwise distance matrix $\mathbf{A}$,

$$A_{ij} = \exp\left(\frac{-d_{ij}^{2\alpha}}{2\epsilon}\right) \tag{1}$$

where ϵ is the kernel bandwidth and defines the local neighborhood of each point and α is a parameter that globally rescales pairwise distances to smooth out large density fluctuations between densely and sparsely sampled regions of configurational state space.[44] Matrix $\mathbf{A}$ is then row-normalized to form the transition matrix,

$$\mathbf{M} = \boldsymbol{D}^{-1}\mathbf{A} \tag{2}$$

where $\boldsymbol{D}$ is a diagonal matrix with elements,

$$D_{ii} = \sum_j A_{ij} \tag{3}$$

The transition matrix, $\mathbf{M}$, is then diagonalized to calculate its eigenvectors ψ_i and eigenvalues λ_i. By the Markov property, the top eigenvalue−eigenvector pair ($\psi_0 = \vec{1}$, $\lambda_0 = 1$) is trivial, corresponding to the steady-state distribution of a random walk. A gap in the eigenvalue spectrum after the kth nontrivial eigenvalue identifies the k-leading eigenvectors corresponding to the leading high-variance nonlinear collective modes of the system. Snapshot i of the molecular simulation trajectory is embedded into these collective variables spanning the so-called intrinsic manifold of the system under the mapping,

$$x_i \mapsto [\psi_1(i), \psi_2(i), \ldots, \psi_k(i)] \tag{4}$$

The ψ_k are the leading nonlinear collective variables identified by the diffusion map that correspond to the high-variance dynamical modes of the system and are responsible for large-scale conformational rearrangements.

Free energy surfaces over the intrinsic manifold $G(\Psi)$ are computed by collecting histogram approximations $\hat{P}$ to the observed distribution of configurational microstates projected into the leading k-eigenvectors $\Psi = \{\psi_i\}_{i=1}^{k}$ and then inverting this distribution using the relation

$$\beta G(\Psi) = -\ln \hat{P}(\Psi) + C \tag{5}$$

where $\beta = 1/(k_B T)$ is the inverse temperature and C is an arbitrary additive constant that sets an absolute free energy scale.[45] By virtue of the interpretability of the eigenvectors as the leading collective modes of the system, the free energy surface constructed over the intrinsic manifold can resolve both the metastable macrostates of the chromatin structure and the interconversion pathways between them.[24] Diffusion maps have already been used successfully to examine the dynamics of DNA around histone proteins, thereby providing precedent for our approach,[43] but we note that we could have employed tICA, VAMPnets, or SRVs in conjunction with Markov-state models to identify kinetic microstates and macrostates.[46−50] These approaches have the benefit of furnishing kinetic networks without requiring that the assumption of diffusive dynamics be made. In the present work, it is the structure and thermodynamics of the metastable states that are of primary interest, as opposed to the kinetic transition rates, and for this reason we favor the smooth, continuous, and more structurally interpretable free energy surfaces furnished by diffusion maps.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscentsci.1c00085.

(Figure S1) Eigenvalue spectrum of the transition matrix; (Figure S2) representative tetranucleosomes motifs along 187 NRL chromatin fiber; (Figure S3) fiber of 16-nucleosomes transitioning between tetranucleosomes motifs (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Andrew L. Ferguson − *Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0002-8829-9726; Phone: (773) 702-7791; Email: andrewferguson@uchicago.edu

Juan J. de Pablo − *Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0002-3526-516X; Phone: (773) 702-7791; Email: depablo@uchicago.edu

Authors

Walter Alvarado − *Biophysical Sciences, University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0002-9027-9951

Joshua Moller − *Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acscentsci.1c00085

Notes

The authors declare the following competing financial interest(s): Andrew L. Ferguson is a co-founder and consultant of Evozyne, Inc. and a co-author of U.S. Provisional Patents 62/853,919 and 62/900,420 and International Patent Applications PCT/US2020/035206 and PCT/US20/50466.

https://doi.org/10.1021/acscentsci.1c00085
ACS Cent. Sci. 2021, 7, 1019−1027

■ ACKNOWLEDGMENTS

We thank Dr. Tobin Sosnick and Dr. Vadim Backman for their generous feedback of this work. We also thank Aria Coraor, Soren Kyhl, Mike Jones, Kirill Shmilovich, Wei Chen, and Yutao Ma for their helpful discussions and Xinran Lian for her assistance in preparing artwork. This study was supported by NSF Grant EFRI EEC 1830969. This work was completed in part with resources provided by the University of Chicago Research Computing Center. We gratefully acknowledge computing time on the University of Chicago high-performance GPU-based cyberinfrastructure (NSF Grant No. DMR-1828629).

■ REFERENCES

(1) van Holde, K. E. Chromatin structure and transcription. In *Chromatin*; Springer Series in Molecular Biology; Springer, 1989; pp 355−408, DOI: 10.1007/978-1-4612-3490-6_8.
(2) Grandy, R. A.; Lian, J. B.; Stein, J. L.; van Wijnen, A. J.; Stein, G. S. Chromatin. In *Brenner's Encyclopedia of Genetics*, 2nd ed.; Elsevier, 2013; pp 538−541.
(3) Joffe, J.; Keene, M.; Weintraub, H. Histones H2a, H2b, H3, and H4 Are Present in Equimolar Amounts in Chick Erythroblasts. *Biochemistry* **1977**, *16*, 1236−1238.
(4) Tremethick, D. J. Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* **2007**, *128*, 651−654.
(5) Eltsov, M.; MacLellan, K. M.; Maeshima, K.; Frangakis, A. S.; Dubochet, J. Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 19732−19737.
(6) Hsieh, T.-H. S.; Weiner, A.; Lajoie, B.; Dekker, J.; Friedman, N.; Rando, O. J. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **2015**, *162*, 108−119.
(7) Roth, E.; Agrawal, V.; Eshein, A.; Fredrick, J.; Almassalha, L.; Shim, A.; Bleher, R.; Dravid, V. P.; Backman, V. Quantifying Three-Dimensional Chromatin Organization Utilizing Scanning Transmission Electron Microscopy: ChromSTEM. *bioRxiv Preprint (Genomics)*. 2019. 636209. https://www.biorxiv.org/content/10.1101/636209v1
(8) Nishino, Y.; Eltsov, M.; Joti, Y.; Ito, K.; Takata, H.; Takahashi, Y.; Hihara, S.; Frangakis, A. S; Imamoto, N.; Ishikawa, T.; Maeshima, K. Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *EMBO J.* **2012**, *31*, 1644−1653.
(9) Ricci, M. A.; Manzo, C.; Garcia-Parajo, M. F.; Lakadamyali, M.; Cosma, M. P. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **2015**, *160*, 1145−1158.
(10) Rao, S. S. P.; Huang, S.-C.; Durand, N. C.; Huntley, M. H.; Jewett, A. I.; Bochkov, I. D.; Chinnappan, D.; Cutkosky, A.; Li, J.; Geeting, K. P.; et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E6456−E6465.
(11) Chen, C.; Lim, H. H.; Shi, J.; Tamura, S.; Maeshima, K.; Surana, U.; Gan, L. Budding yeast chromatin is dispersed in a crowded nucleoplasm in vivo. *Mol. Biol. Cell* **2016**, *27*, 3357−3368.
(12) Moraru, M.; Schalch, T. Chromatin fiber structural motifs as regulatory hubs of genome function? *Essays Biochem.* **2019**, *63*, 123−132.
(13) Ohno, M.; Ando, T.; Priest, D. G.; Kumar, V.; Yoshida, Y.; Taniguchi, Y. Sub-nucleosomal Genome Structure Reveals Distinct Nucleosome Folding Motifs. *Cell* **2019**, *176*, 520−534.e25.
(14) Song, F.; Chen, P.; Sun, D.; Wang, M.; Dong, L.; Liang, D.; Xu, R.-M.; Zhu, P.; Li, G. Cryo-EM Study of the Chromatin Fiber Tetranucleosomal Units. *Science* **2014**, *344*, 376−380.
(15) Ekundayo, B.; Richmond, T. J.; Schalch, T. Capturing Structural Heterogeneity in Chromatin Fibers. *J. Mol. Biol.* **2017**, *429*, 3031−3042.
(16) Schalch, T.; Duda, S.; Sargent, D. F.; Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **2005**, *436*, 138−141.
(17) Eltsov, M.; Grewe, D.; Lemercier, N.; Frangakis, A.; Livolant, F.; Leforestier, A. Nucleosome conformational variability in solution and in interphase nuclei evidenced by cryo-electron microscopy of vitreous sections. *Nucleic Acids Res.* **2018**, *46*, 9189−9200.
(18) Takizawa, Y.; Ho, C.-H.; Tachiwana, H.; Matsunami, H.; Kobayashi, W.; Suzuki, M.; Arimura, Y.; Hori, T.; Fukagawa, T.; Ohi, M. D.; et al. Cryo-EM structures of centromeric tri-nucleosomes containing a central CENP-A nucleosome. *Structure* **2020**, *28*, 44−53.
(19) Maeshima, K.; Tamura, S.; Hansen, J. C.; Itoh, Y. Fluid-like chromatin: Toward understanding the real chromatin organization present in the cell. *Curr. Opin. Cell Biol.* **2020**, *64*, 77−89.
(20) Lequieu, J.; Cordoba, A.; Moller, J.; de Pablo, J. J. 1CPN: A coarse-grained multi-scale model of chromatin. *J. Chem. Phys.* **2019**, *150*, 215102.
(21) 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.* **2014**, *113*, 168101.
(22) Lequieu, J.; Cordoba, A.; Schwartz, D. C.; de Pablo, J. J. Tension-dependent free energies of nucleosome unwrapping. *ACS Cent. Sci.* **2016**, *2*, 660−666.
(23) Routh, A.; Sandin, S.; Rhodes, D. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 8872−8877.
(24) Ferguson, A. L.; Panagiotopoulos, A. Z.; Kevrekidis, I. G.; Debenedetti, P. G. Nonlinear dimensionality reduction in molecular simulation: The diffusion map approach. *Chem. Phys. Lett.* **2011**, *509*, 1−11.
(25) Martino, F.; Kueng, S.; Robinson, P.; Tsai-Pflugfelder, M.; van Leeuwen, F.; Ziegler, M.; Cubizolles, F.; Cockell, M. M.; Rhodes, D.; Gasser, S. M. Reconstitution of Yeast Silent Chromatin: Multiple Contact Sites and O-AADPR Binding Load SIR Complexes onto Nucleosomes In Vitro. *Mol. Cell* **2009**, *33*, 323−334.
(26) Schalch, T.; Kulangara, A.; Duda, S.; Schroeder, R. R.; Richmond, T. J.; Dorigo, B. Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* **2004**, *306*, 1571−1573.
(27) Marshall, W.F; Straight, A; Marko, J.F; Swedlow, J; Dernburg, A; Belmont, A; Murray, A.W; Agard, D.A; Sedat, J.W Interphase chromosomes undergo constrained diffusional motion in living cells. *Curr. Biol.* **1997**, *7*, 930−939.
(28) Ashwin, S. S.; Nozaki, T.; Maeshima, K.; Sasai, M. Organization of fast and slow chromatin revealed by single-nucleosome dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 19939−19944.
(29) Perisic, O.; Collepardo-Guevara, R.; Schlick, T. Modeling studies of chromatin fiber structure as a function of DNA linker length. *J. Mol. Biol.* **2010**, *403*, 777−802.
(30) Poirier, M. G; Oh, E.; Tims, H. S; Widom, J. Dynamics and function of compact nucleosome arrays. *Nat. Struct. Mol. Biol.* **2009**, *16*, 938−944.
(31) Ou, H. D.; Phan, S.; Deerinck, T. J.; Thor, A.; Ellisman, M. H.; O'Shea, C. C. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **2017**, *357*, eaag0025.
(32) Baù, D.; Sanyal, A.; Lajoie, B. R.; Capriotti, E.; Byron, M.; Lawrence, J. B.; Dekker, J.; Marti-Renom, M. A. The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. *Nat. Struct. Mol. Biol.* **2011**, *18*, 107.
(33) Pope, B. D.; Ryba, T.; Dileep, V.; Yue, F.; Wu, W.; Denas, O.; Vera, D. L.; Wang, Y.; Hansen, R. S.; Canfield, T. K.; et al. Topologically associating domains are stable units of replication-timing regulation. *Nature* **2014**, *515*, 402−405.
(34) Barbieri, M.; Chotalia, M.; Fraser, J.; Lavitas, L.-M.; Dostie, J.; Pombo, A.; Nicodemi, M. Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 16173−16178.

(35) Woodcock, C. L.; Skoultchi, A. I.; Fan, Y. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res.* **2006**, *14*, 17−25.
(36) 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.* **2013**, *139*, 144903.
(37) Lequieu, J.; Schwartz, D. C.; de Pablo, J. J. In silico evidence for sequence-dependent nucleosome sliding. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E9197−E9205.
(38) Stehr, R.; Kepper, N.; Rippe, K.; Wedemann, G. The Effect of Internucleosomal Interaction on Folding of the Chromatin Fiber. *Biophys. J.* **2008**, *95*, 3677−3691.
(39) Niedermeier, C.; Schulten, K. Molecular Dynamics Simulations In Heterogeneous Dielectrica And Debye-Huckel Media - Application To The Protein Bovine Pancreatic Trypsin Inhibitor. *Mol. Simul.* **1992**, *8*, 361−387.
(40) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1−19.
(41) Boninsegna, L.; Gobbo, G.; Noe, F.; Clementi, C. Investigating Molecular Kinetics by Variationally Optimized Diffusion Maps. *J. Chem. Theory Comput.* **2015**, *11*, 5947−5960.
(42) Coifman, R. R.; Lafon, S. Diffusion maps. *Appl. Comput. Harmonic Anal.* **2006**, *21*, 5−30.
(43) Guo, A. Z.; Lequieu, J.; de Pablo, J. J. Extracting collective motions underlying nucleosome dynamics via nonlinear manifold learning. *J. Chem. Phys.* **2019**, *150*, 054902.
(44) Wang, J.; Gayatri, M. A.; Ferguson, A. L. Mesoscale Simulation and Machine Learning of Asphaltene Aggregation Phase Behavior and Molecular Assembly Landscapes. *J. Phys. Chem. B* **2017**, *121*, 4923−4944.
(45) Shell, M. S. *Thermodynamics and Statistical Mechanics: An Integrated Approach*; Cambridge University Press, 2015; DOI: 10.1017/CBO9781139028875.
(46) Chen, W.; Sidky, H.; Ferguson, A. L. Nonlinear discovery of slow molecular modes using state-free reversible VAMPnets. *J. Chem. Phys.* **2019**, *150*, 241703.
(47) Mardt, A.; Pasquali, L.; Wu, H.; Noe, F. VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* **2018**, *9*, 5.
(48) Noe, F.; Clementi, C. Kinetic Distance and Kinetic Maps from Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2015**, *11*, 5002−5011.
(49) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **2010**, *52*, 99−105.
(50) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Perez-Hernandez, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noe, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11*, 5525−5542.

https://doi.org/10.1021/acscentsci.1c00085
ACS Cent. Sci. 2021, 7, 1019−1027