

Nucleosome condensate and linker DNA alter chromatin folding pathways and rates

Yunrui Qiu,^{†,‡,§} Shuming Liu,^{¶,§} Xingcheng Lin,[¶] Ilona Christy Unarta,^{†,‡} Xuhui Huang,^{*,†,‡} and Bin Zhang^{*,¶}

[†]*Department of Chemistry, Theoretical Chemistry Institute, University of Wisconsin-Madison, Madison, WI, USA*

[‡]*Data Science Institute, University of Wisconsin-Madison, Madison, WI, USA*

[¶]*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA*

[§]*Contributed equally to this work*

E-mail: xhuang@chem.wisc.edu; binz@mit.edu

Abstract

Chromatin organization is essential for DNA packaging and gene regulation in eukaryotic genomes. While significant progresses have been made, the exact atomistic arrangement of nucleosomes remains controversial. Using a well-calibrated residue-level coarse-grained model and advanced dynamics modeling techniques, particularly the non-Markovian dynamics model, we map the free energy landscape of tetra-nucleosome systems, identify both metastable conformations and intermediate states in folding pathways, and quantify the folding kinetics. Our findings show that chromatin with $10n$ base pairs (bp) DNA linker lengths favor zigzag fibril structures. However, longer linker lengths destabilize this conformation. When the linker length is $10n + 5$ bp, chromatin loses unique conformations, favoring a dynamic ensemble of structures resembling folding intermediates. Embedding the tetra-nucleosome in a nucleosome con-

condensate similarly shifts stability towards folding intermediates as a result of the competition of inter-nucleosomal contacts. These results suggest that chromatin organization observed *in vivo* arises from the unfolding of fibril structures due to nucleosome crowding and linker length variation. This perspective aids in unifying experimental studies to develop atomistic models for chromatin.

Significance. Atomic structures of chromatin have become increasingly accessible, largely through cryo-EM techniques. Nonetheless, these approaches often face limitations in addressing how intrinsic *in vivo* factors influence chromatin organization. We present a structural characterization of chromatin under the combined effects of nucleosome condensate crowding and linker DNA length variation—two critical *in vivo* features that have remained challenging to capture experimentally. This work leverages a novel application of non-Markovian dynamical modeling, providing accurate mapping of chromatin folding kinetics and pathways. Our findings support a hypothesis that *in vivo* chromatin organization arises from folding intermediates advancing toward a stable fibril configuration, potentially resolving longstanding questions surrounding chromatin atomic structure.

Introduction

Chromatin organization, essential for packaging DNA in eukaryotic genomes, plays a crucial role in various genetic functions.¹⁻⁵ Atomistic structures of chromatin, where available, have proven invaluable in constructing mechanistic models of gene regulation and other processes.^{6,7} However, the detailed organization at the atomic level has been the subject of contentious debate. The existence of 30-nm fibers, characterized by nucleosomes following a zig-zag path and stacking on top of each other to form a twisted two-column fibril, has long been documented.⁶⁻¹⁰ However, several *in vivo* experimental techniques, including cryo-electron microscopy,⁶ Micro-C,¹¹ and ChromEMT,¹² have indicated a lack of ordered fibril-like structures. Instead, these techniques suggest the presence of 10-nm disordered arrays with prevailing local oligomer motifs, such as trimers, α -tetrahedron, and β -rhombus tetramers.¹¹⁻¹⁶ These contradictory observations have left the folding principles of chromatin, particularly at the short length scale of tens of nucleosomes, unresolved.

A more dynamic perspective on chromatin organization has proven to be insightful. Ding et al.¹⁷ characterized the folding landscape of a tetra-nucleosome using a residue-level coarse-grained (CG) model, proposing that the unfolding of fibril structures leads to the irregular conformations observed in the nucleus, thus bridging *in vitro* and *in vivo* configurations. Liu et al.¹⁸ further demonstrated that unfolding can generate clutches, consistent with observations from super-resolution imaging of the cell nucleus.¹⁹ A recent cryo-electron tomography (cryo-ET) study further supports the presence of similar folding principles for chromatin both *in vitro* and *in situ*.¹² However, the computational studies employed biased simulations to accelerate chromatin folding and unfolding. While thermodynamic predictions can be made, extracting accurate kinetic information from such simulations proves challenging. Consequently, it is significant to more directly determine whether the irregular *in vivo* chromatin configurations represent intermediate folding stages leading to the zigzag fibril structure.

Markov state models (MSMs) offer an effective approach to extract interpretable kinetic information, facilitating the prediction of transition rates and the identification of

reaction pathways from the molecular dynamics (MD) data.^{20–35} Specifically, MSMs could coarse-grain complex MD configurations into conformational states and model continuous dynamics as Markovian jumps between these states at discrete time interval (i.e., lag time). The construction of such models and prediction of long-timescale dynamics require only distributed, relatively short MD trajectories as inputs, which can be efficiently generated in parallel using supercomputing facilities. To aid the interpretation of complex dynamics, it is often beneficial to build a model with only a handful of metastable and interpretable states. However, accurately capturing the slow dynamics between these states is challenging, as this requires a Markovian lag time that is typically much longer than the length of the affordable MD simulation. Our recently developed non-Markovian dynamic model demonstrates significant advantages in addressing this challenge. By employing the generalized master equation (GME) with memory kernel functions, it could effectively model non-Markovian transition dynamics, greatly enhancing modeling accuracy.^{36–39}

In this study, we conduct extensive residual-level CG MD simulations to examine chromatin folding. Specifically, we investigate the folding of four systems: three isolated tetranucleosomes with varying nucleosome repeat lengths (NRL) of 167, 172, and 177 base pairs (bp), and a tetra-nucleosome with an NRL of 167 bp embedded within a nucleosome condensate. We construct microstate-MSM and interpretable non-Markovian dynamics model for each system to comprehensively characterize the folding pathways and kinetics. In every system, transition path analysis consistently reveals numerous parallel folding pathways with comparable fluxes. This finding contrasts with typical protein folding and more closely resembles multi-body self-assembly processes. Notably, intermediate states along these pathways resemble *in vivo* chromatin organization. Furthermore, while the tetra-nucleosome with NRL = 167 bp folds stably into the zigzag fibril structure, both the nucleosome condensate and a 5 bp longer DNA linker destabilize the fibril conformation, promoting the formation of folding intermediates through different molecular mechanisms. Extending the linker DNA length by 10 or 5 bp has markedly different effects on the chromatin folding landscape: while

the 10 bp longer linker favors the zigzag fibril configuration, the 5 bp longer linker does not result in any specific preferred structure, instead favoring an ensemble of dynamic conformations. These results reinforce the idea that the absence of regular fibril configurations inside cell nuclei could result from chromatin unfolding driven by various factors such as crowding environments and linker length variation.

Results

Construction of dynamics models for chromatin folding from extensive MD simulations

Our study focuses on tetra-nucleosomes, the fundamental units of a zigzag fibril configuration,^{6,7} to explore chromatin organization⁶ in various biological contexts. Specifically, we constructed a series of tetra-nucleosome structures with 20, 25, and 30 bp linkers, corresponding to NRL of 167, 172, and 177 bp (Figure 1A-C). We define nucleosomal DNA as 147 bp, and refer to the DNA connecting two adjacent nucleosomes as the linker DNA. To simulate a biologically relevant scenario, we further embedded the tetra-nucleosome with the 20 bp linker within a solution of single nucleosomes to account for nuclear crowding effects. Our setup results in an overall nucleosomal concentration of 0.3 mM (Figure 1D), consistent with concentrations estimated from *in vitro* nucleosome array condensates.^{40,41}

We performed MD simulations to investigate conformational dynamics of chromatin with a residue-level CG model. Coarse-graining is essential because all-atom simulations are computationally prohibitive for extensive sampling of multiple nucleosome structures.^{42–46} Specifically, we employed a one-bead-per-amino-acid model⁴⁷ and a three-bead-per-nucleotide model⁴⁸ to maintain sufficient resolution for an accurate description of specific protein-protein and protein-DNA interaction with physical chemistry potentials. These models have been effectively utilized to study various protein-DNA systems, accurately reproducing experimental results and providing mechanistic insights.^{17,18,49–54} Further details on the CG

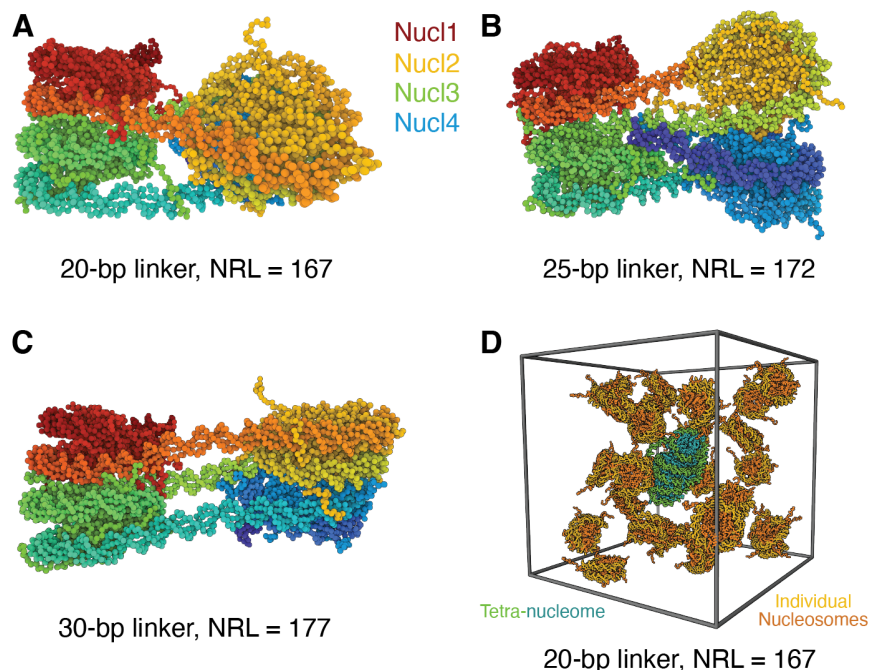


Figure 1: **Representative configurations for the four tetra-nucleosome systems studied.** The three isolated systems feature tetra-nucleosomes of 20-bp (A), 25-bp (B), and 30-bp (C) DNA linker. The corresponding NRL is 167, 172, and 177 bp, respectively. In the fourth system, the tetra-nucleosome with NRL=167 is embedded into a nucleosome condensate.

force field are provided in the *Methods* section.

For each of these four systems, we conducted multiple independent, unbiased MD simulations to explore various regions of the chromatin configurational space. As illustrated in Figure 2A, these simulations were initiated from distinct conformations, capturing various degrees of chromatin compaction and folding, as identified in a previous study.¹⁷ For isolated tetra-nucleosomes with varying linker lengths, we executed 4,643 simulations for each system. In the case of the condensate system, due to its longer equilibration time, we reduced the number of trajectories to 530 and extended each simulation length. All simulations are sufficiently long to allow for the relaxation of chromatin conformations and the attainment of local equilibrium (Figures S1 and S2).

Subsequently, we constructed dynamics models for each system to study chromatin folding over timescales much longer than individual MD simulation trajectories. We projected

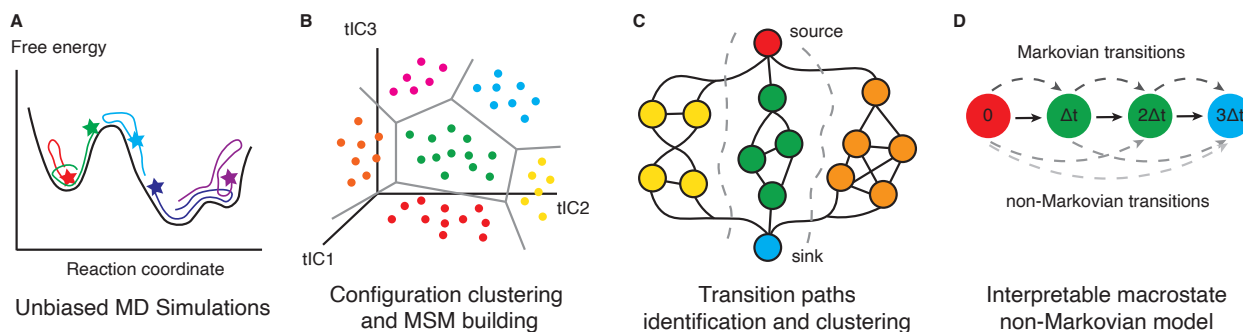


Figure 2: Overview of the computational pipeline for determining chromatin folding kinetics and pathways. The process begins with extensive unbiased MD simulations, initiated from a variety of configurations (A). Configurations collected from these simulations are projected onto collective variables identified by tICA and then clustered into microstates to construct MSMs. (B). Subsequently, chromatin folding pathways and reaction channels are identified using transition path theory and a path clustering algorithm (C). Finally, the microstates are lumped into a few metastable macrostates, and the transition dynamics between these macrostates are modeled using the generalized master equation that incorporates time-dependent memory kernels (D).

tetra-nucleosome configurations explored by unbiased trajectories onto collective variables using time-lagged independent component analysis (tICA),^{55–57} and clustered them into microstates by the K-Means algorithm (Figure 2B and Figures S3-S6). Microstate-MSMs were then constructed and validated to model chromatin folding dynamics (Figures S7-S10). Further employing the transition path theory (TPT)^{20,21,58} and path clustering algorithm,²⁷ we identified the kinetic pathways and reaction channels for tetra-nucleosome folding. (Figure 2C and Figures S11-S17).

To facilitate the interpretation of folding dynamics, we also built non-Markovian dynamics models with six metastable macrostates lumped from microstates using the Integrated Generalized Master Equation (IGME) method³⁶ (Figure 2D and Figures S18-S25). Unlike microstate-MSMs, which typically require hundreds of states to capture the slowest dynamics and are thereby challenging to interpret, IGME models can accurately recover the slowest dynamics with only a few metastable states. This advantage arises from IGME’s incorporation of the time-integration of memory kernels to account for non-Markovian dynamics. We demonstrated that for all systems reported in this study, IGME models consistently show sig-

nificantly higher accuracy in modeling long-term dynamics compared to MSMs constructed with similar or longer MD segments. More details regarding system setup, simulations, and constructions of MSMs and IGME models are available in the *Methods* section and *Supporting Information*.

We note that CG models are frequently limited in their ability to accurately represent the timescales of dynamical motion. Nevertheless, their demonstrated accuracy in quantifying protein-DNA interactions and mapping the free energy landscape of nucleosome unwinding and chromatin folding^{17,18,59–61} suggests that they are well-suited to capture the sequence of events occurring over slow timescales, governed by the topology of the energy landscape. This capability to model the qualitative mechanisms driving large-scale conformational changes supports the comprehensive kinetic analyses presented above.

Elucidation of tetra-nucleosome folding pathways and kinetics

We initiated our investigation with the isolated $NRL = 167$ tetra-nucleosome. As shown in Figure 3A, we first plotted the two-dimensional free energy landscape projected onto internucleosomal distances, d_{13} and d_{24} , based on the equilibrium populations obtained from the microstate-MSM. The resulting landscape, featuring an approximate 10 kcal/mol difference between unfolded and folded regions, indicates the stability of the native tetra-nucleosome structures and is consistent with previous estimations from neural network fitting of mean forces.¹⁷

By conducting transition path analysis with the microstate-MSM, we identified over 20,000 pathways connecting the unfolded to folded states. Strikingly, the most dominant pathway only contributes 0.09% of the total reactive flux for the folding. This contrasts with typical protein folding, such as NTL9 folding, where the top 10 pathways can account for 25% of the total flux.²⁹ In the case of tetra-nucleosome folding, approximately 600 pathways are required to reach the same flux level (Figure S13A). The observation of the downhill landscape and numerous parallel pathways with comparable fluxes suggests that

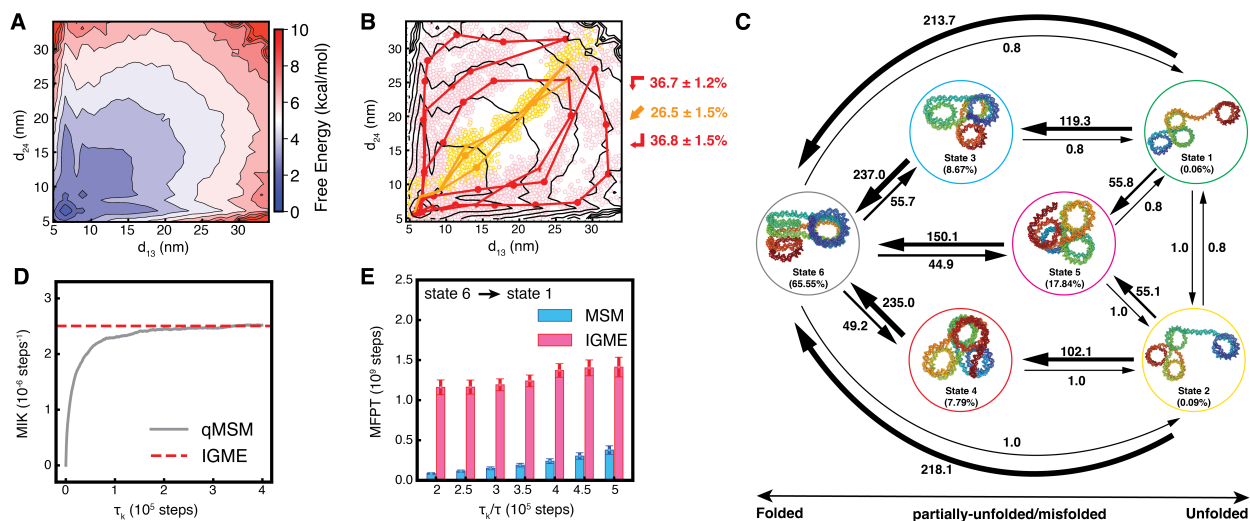


Figure 3: **Folding pathways and kinetics for the NRL = 167 tetra-nucleosome.** (A) The free energy profile along the center-of-geometry distance between 1-3 (d_{13}) and 2-4 (d_{24}) nucleosomes. (B) The three reaction channels for tetra-nucleosome folding. Top three transition pathways with most reactive flux from each one reaction channel are drawn as lines. The filled and open circles represent the centers of, and the MD configurations belonging to, the microstates along the pathways, respectively. The total reactive flux of each reaction channel is provided on the side. (C) Diagram of the IGME model with six macrostates. The numbers represent rates estimated as inverse MFPT labeled in unit of $(10^9 \text{ steps})^{-1}$ for the corresponding transitions. Histone proteins are not shown in the representative configurations of each state. (D) The mean integral of memory kernels (MIK) across different τ_k for the six-macrostate model, calculated using the quasi-MSM (qMSM) and IGME. (E) The mean first passage time (MFPT) from State 6 to State 1, computed from MSMs and IGME with different lag times, is compared. The error bars show standard deviations, estimated from bootstrapping of data 20 times with replacement.

tetra-nucleosome folding is more akin to heterogeneous aggregation in self-assembly systems than to typical protein folding.³²

To aid in interpreting the multiple parallel pathways with similar fluxes, we grouped the pathways into a small set of metastable reaction channels using the Latent-space Path Clustering algorithm.²⁷ This method identified three metastable reaction channels: two sequential channels (up and down sequential channels) and one concerted channel (Figure 3B and S17A). These reaction channels exhibit distinct folding mechanisms: the sequential channels correspond to processes where one pair of nucleosomes stacks before the other, while the concerted channel represents processes where both pairs of nucleosomes stack simultane-

ously. These findings align with a previous study that characterized tetra-nucleosome folding pathways using the string method.¹⁷ The overall reactive flux in the sequential channels is slightly higher than in the concerted channel, indicating a preference for reactions. This preference may result from more stable intermediates encountered along sequential pathways with increased internucleosomal contacts.

To further understand chromatin folding dynamics, we constructed an IGME model with six metastable macrostates (Figure 3C). The sparsely populated states 1 and 2 correspond to unfolded structures. Their high transition rates toward the folded state (state 6), coupled with low rates for backward transitions, align with a downhill free energy landscape. States 3 and 4 present partially unfolded structures, where one nucleosome extrudes from the other three. These configurations resemble those observed in sequential chromatin folding pathways (Figure S27). A cryo-EM study directly observed tri-nucleosome configurations,⁶² demonstrating the IGME model’s capability to identify metastable states.

The model also identifies a misfolded state (state 5), marked by non-canonical stacking between the first and fourth nucleosomes (i.e., i & $i + 3$ contacts). For the transition to the zigzag conformation of state 6, these interactions must be disrupted, resulting in slow transition rates. We classify this state as misfolded, as transition path analysis indicates that none of the top 3,000 most reactive pathways pass through it. Future cryo-EM studies may clarify whether this state can be resolved.

Specifically, we found that our IGME model provides significant improvements over MSMs in accurately quantifying equilibrium populations and transition rates among metastable states using shorter MD simulation segments, as the memory kernel’s relaxation timescale is generally much shorter than the Markovian lag time. By evaluating memory kernel integrations over various lag times using different approaches (see *Methods* for more details), we observed that the integrations converge at approximately $\tau_k = 2 \times 10^5$ steps (Figure 3D), beyond which the IGME models can be constructed. An example IGME model built from MD simulation segments of $\sim 3 \times 10^5$ steps has been shown to accurately predict dynamics

compared to MD simulations (Figure S19). Conversely, the MSM constructed with a longer lag time exhibited substantially higher prediction errors, failed the Chapman–Kolmogorov test, and predicted much faster state relaxation dynamics (Figures S18–S19). Across a wide range of lag times, IGME models consistently outperformed MSMs in capturing the slowest mean first passage time (MFPT) for unfolding dynamics (Figure 3E). Matching the performance of the IGME model would require constructing the MSM with significantly longer lag times, exceeding the length of our MD simulations.

Nucleosome condensate promotes chromatin unfolding

Our study on the isolated tetra-nucleosome does not consider the nuclear environment. The presence of additional nucleosomes may not only act as a crowding agent but also affect the stability of chromatin conformations through direct interactions. Inter-nucleosomal interactions could destabilize the chromatin fiber by forming interdigitated configurations.^{18,63} In this section, we examine the effects of immersing the tetra-nucleosome in a nucleosome condensate environment on chromatin folding.

In simulations of this condensate system, individual free nucleosomes were initially uniformly distributed throughout the cubic simulation box with the tetra-nucleosome at the center (Figure 4A). Over time, they rapidly and spontaneously aggregated around the tetra-nucleosome. This aggregation resulted in contacts between free nucleosomes and those within the tetra-nucleosome. Such contacts can be clearly observed in the end configuration of a typical simulation trajectory provided in Figure 4A. They are also evident in the distributions of free nucleosomes around the tetra-nucleosome. As shown in Figure S28, the radial distributions indicate significant populations at distances as small as 5 nm, a value achievable only with stacked nucleosomes.

We integrated individual unbiased simulations to construct microstate-MSM for the tetra-nucleosome in the condensate. By using the equilibrium populations of microstates, we estimated the two-dimensional free energy landscape of the tetra-nucleosome. As shown

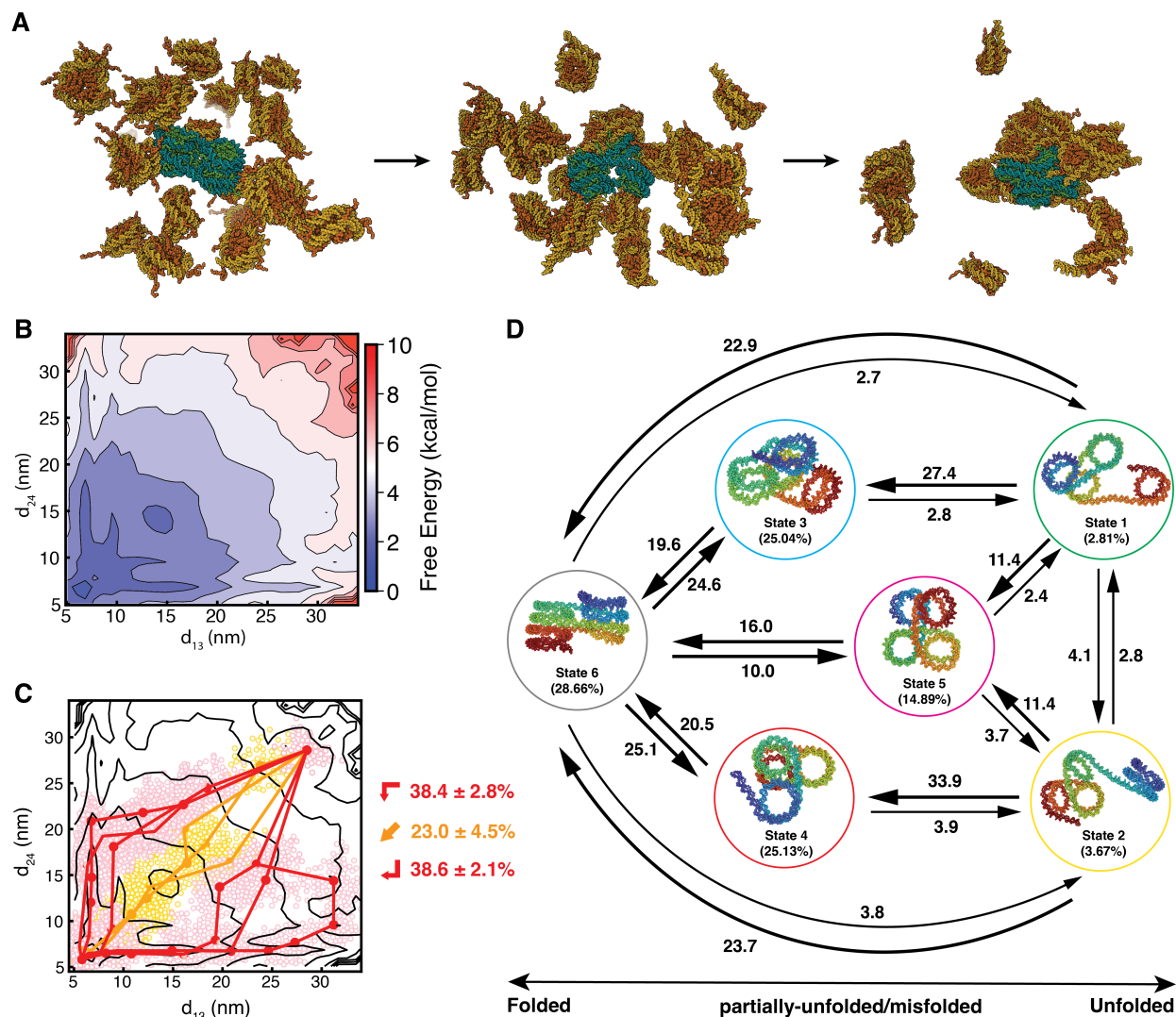


Figure 4: **Folding pathways and kinetics for the $NRL = 167$ tetra-nucleosome embedded in nucleosome condensate.** (A) Illustration of the starting, middle, and end configurations of the condensate system along a 70 million step long simulation trajectory. The tetra-nucleosome is shown in cyan and green, and individual nucleosomes are shown in yellow and orange. (B) The free energy profile along the center-of-geometry distance between 1-3 (d_{13}) and 2-4 (d_{24}) nucleosomes from the tetra-nucleosome. (C) The three reaction channels for tetra-nucleosome folding. Top three transition pathways with most reactive flux from each one reaction channel are drawn as lines. The filled and open circles represent the centers of, and all the MD configurations belonging to, the microstates along the pathways, respectively. The total reactive flux of each reaction channel is provided on the side. (D) Diagram of the IGME model with six macrostates. The numbers represent rates estimated as inverse MFPT labeled in unit of $(10^9 \text{ steps})^{-1}$ for the corresponding transitions. Histone proteins are not shown in the representative configurations of each state.

in Figure 4B, the overall shape of the free energy landscape remains similar to that of the isolated system, but the free energy values of the folding intermediate structures have decreased (see also Figure S29). This reduction is attributed to interactions between the tetra-nucleosome and free nucleosomes, which stabilize partially folded structures. However, the overall change in the free energy landscape is relatively small, on the order of 1-2 kcal/mol.

We then identified tetra-nucleosome folding pathways for the condensate system and clustered them into reaction channels. Similar to the isolated system, we found thousands of pathways with comparable fluxes were categorized into sequential or concerted channel. The distribution of fluxes among these reaction channels also remained largely unchanged, indicating a conservation of the folding mechanism. (Figures 4C and S17). The relatively small changes in the free energy landscape and folding reaction channels prompt questions about the role of the nucleosome condensate in tetra-nucleosome folding.

To more directly and accurately examine chromatin folding kinetics, we constructed a six-macrostate IGME model (Figure 4D). Due to the complex interactions introduced by nucleosome condensation, we found that the required lag time to build a Markovian MSM is significantly prolonged (Figure S25). With limited-length MD simulations, the IGME model shows distinct advantages, capturing the slow dynamics with much less error (Figure S24-S25). Based on the optimal IGME model, we observed an obvious increase in the populations of partially unfolded states 3 and 4, which are comparable to the population of the folded state 6. This finding underscores the changes in the free energy observed in Figure 4B and the impact of nucleosome condensate on chromatin stability. Notably, the unfolding rates from state 6 to these intermediate states are faster than the folding rates. Therefore, the nucleosome condensate does not fundamentally alter the folding modes of the tetra-nucleosome but quantitatively modulates the free energy landscape, favoring unfolding dynamics driven by inter-nucleosomal contacts.

The role of DNA linker length on chromatin folding

Our study thus far has focused on a tetra-nucleosome with an NRL of 167 bp (i.e., 20-bp linkers). However, linker length is known to influence chromatin organization significantly.^{40,64–68} A DNA linker of $10n$ bp facilitates well-aligned stacking between i and $i + 2$ nucleosomes to form the zigzag fibril structure.^{6,7,69,70} Extending the linker by 5 bp induces a half-turn twist in the DNA, disrupting the stacking between i and $i + 2$ nucleosomes and destabilizing the zigzag conformation.⁷¹ These $10n + 5$ linkers are significant and have been observed in certain mammalian cells.⁷² However, structural characterization of chromatin with $10n + 5$ linker DNA is scarce due to the irregularity and instability of their organization. We further investigate how increasing the linker by 5 and 10 bp impacts chromatin stability and folding dynamics using the computational pipelines mentioned earlier.

The folding of the tetra-nucleosome with an NRL of 177 resembles that of an NRL of 167 but exhibits quantitative differences in reactive fluxes of folding. Its free energy landscape also shows a global minimum at small values of d_{13} and d_{24} (Figure S26A), indicating stacked two-column structures. However, the stability of this folded state is reduced, as evidenced by its lower equilibrium populations in the six-macrostate IGME model (Figure S26B). The folded conformations bring linker DNA into close proximity, increasing electrostatic repulsion for longer linkers, which decreases stability. Intermediate states along the sequential pathway also feature stacked nucleosomes with closely positioned linker DNA, resulting in similar electrostatic penalties. As a result, unlike the NRL = 167 tetra-nucleosome, the NRL = 177 one favors the concerted folding pathway over the sequential ones (Figure S17D). Notably, our findings align with those of Correll et al.⁶⁵, who observed an expansion of chromatin with longer $10n$ -bp linkers.

Unlike systems with $10n$ -bp linker DNA, the NRL = 172 tetra-nucleosome system explores distinct chromatin conformations over a different folding landscape. As shown in Figure 5A, the free energy profile has a broader global minimum and a smaller free energy gap between the folded and unfolded states. This broad basin is due to the absence of a

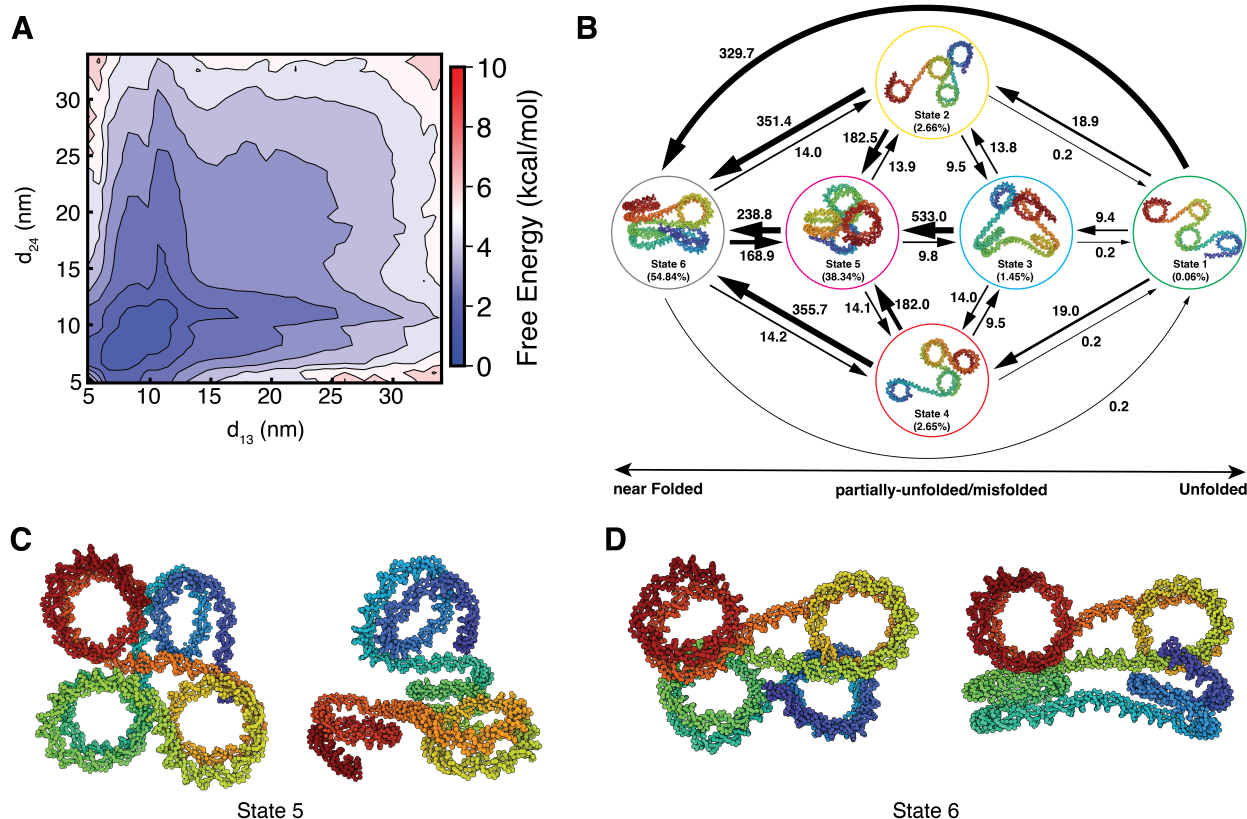


Figure 5: Folding pathways and kinetics for the $NRL = 172$ tetra-nucleosome. (A) The free energy profile along the center-of-geometry distance between 1-3 (d_{13}) and 2-4 (d_{24}) nucleosomes from the tetra-nucleosome. (B) Diagram of the IGME model with six macrostates. The numbers represent rates estimated as inverse MFPT labeled in unit of $(10^9 \text{ steps})^{-1}$ for the corresponding transitions. Histone proteins are not shown in the representative configurations of each state. (C,D) Additional representative structures the compact stable states 5 and 6.

single stable conformation. Instead, an ensemble of irregular structures are equally probable. The six-macrostate IGME model (Figure 5B) reveals that two states, 5 and 6, emerge as “folded” states with comparable populations. Unlike the folded states in $10n$ bp linker systems, these states do not feature a dominant stable configuration but rather an ensemble of conformations (Figures 5C and 5D). While the structures in states 5 and 6 are compact and promote protein-DNA interactions, they do not exhibit perfect stacking between i and $i + 2$ nucleosomes due to the topological constraints imposed by the additional 5 bp. Instead, they resemble intermediate states observed along the folding pathways for $10n$ bp linker systems. The high transition rates between states 5 and 6 further emphasize the dynamic nature of

the tetra-nucleosome with $10n + 5$ bp linkers.

The unfolded states, 1, 2, and 4, which have more extended structures in the $NRL = 172$ tetra-nucleosome, also reveal distinct topology. Unlike the extended states seen in $10n$ bp linkers, where consecutive nucleosomes reside on the same side of the linker DNA, the half-turn linker DNA in $NRL = 172$ tetra-nucleosome introduces additional twist, causing consecutive nucleosomes to reside on opposite sides of the linker.

Conclusions and Discussion

We combined CG MD simulations, Markov state modeling, and non-Markovian dynamic modeling techniques to comprehensively characterize chromatin stability and folding mechanisms. The implementation of the non-Markovian IGME model allows us to accurately investigate long-timescale folding dynamics, which are challenging to capture through direct simulations or conventional MSMs. Our study of a single tetra-nucleosome reveals multiple folding pathways, with intermediate states resembling *in vivo* chromatin organization.^{15,73,74} For instance, intermediates along the concerted pathway often adopt a coplanar geometry, with all four nucleosomes in the same plane, resembling the β -rhombus configuration.¹⁵ In contrast, intermediates along the sequential pathway have at least one nucleosome out of the plane, similar to the α -tetrahedron structure.¹⁵ These findings suggest that the absence of regular fibril configurations inside cell nuclei may result from chromatin unfolding driven by factors such as the condensate environment and linker length.

Our examination of the tetra-nucleosome within the nucleosome condensate supports the hypothesis that a crowded environment, typical inside cell nuclei, can drive chromatin unfolding.¹⁸ The stacking interactions that stabilize the fibril configuration in a single chromatin chain may be replaced by contacts between nucleosomes from different chains. The replacement leads to the formation of irregular conformations, even for DNA linker length that favor fibril structures.

Additionally, we found that $10n+5$ bp linkers destabilize the zigzag fibril configurations, favoring structures that resemble *in vivo* configurations observed by cryo-electron tomography.^{73,74} Chromatin with these DNA linkers tends to adopt compact conformations without perfect stacking between i and $i + 2$ nucleosomes. These conformations are irregular and form a dynamic ensemble, similar to the behavior of intrinsically disordered proteins.

Our study demonstrates that residue-level CG models enable predictive modeling of chromatin organization at near-atomistic resolution, particularly in conditions challenging for experimental techniques, while the novel non-Markovian IGME models facilitate in-depth qualitative analysis of long-term chromatin conformational dynamics. Although our focus was on the impact of nucleosome condensate on chromatin organization, our approach can be generalized to study other protein condensates, such as HP1 α for constitutive heterochromatin. Recent advancements in force fields and software highlight the potential for exciting future research directions.^{61,75,76}

Methods

Coarse-grained molecular dynamics simulations of chromatin

Following previous studies,^{17,18} we utilized residue-level CG models to examine chromatin conformational dynamics. Specifically, we combined the 3SPN.2C model,^{48,77,78} which represents DNA with three sites per nucleotide, and the SMOG model,⁷⁹ which represents histone proteins with one bead per amino acid, to model chromatin. Protein-DNA interactions included electrostatic interactions and non-specific Lennard-Jones potentials. Electrostatic interactions are calculated at 300 K and 150 mM ionic strength using the Debye-Hückel potential. This setup has been shown to replicate the energetics of DNA unwinding in single nucleosomes⁶⁰ and the force-extension curves of chromatin fibers.¹⁸ To reduce computational cost and prevent nucleosome sliding, we rigidified the histone core and the inner 73 bp layer of core DNA during all simulations. Ding et al.¹⁷ demonstrated that such rigidification has

negligible effects on the chromatin folding landscape.

We studied three isolated tetra-nucleosome systems with NRL of 167, 172, and 177 bp, respectively. For each system, we performed 4,643 independent, unbiased MD simulations starting from distinct initial configurations. Initial configurations for these simulations were prepared using 10,000 tetra-nucleosome structures from a previous study that applied enhanced sampling methods to comprehensively cover the configuration space.¹⁷ We computed the inter-nucleosome distances for the 10,000 structures and selected those with $d_{13} \geq d_{24}$ (4,643 in total) to serve as reference values for further MD simulations. From these reference inter-nucleosome distances, we generated the initial configurations using 0.2 million-step-long restrained MD simulations to bias the d_{13} and d_{24} toward reference values. From these initial configurations, we conducted unbiased NVT simulations for production. For the NRL = 167 system, each trajectory lasted at least 0.8 million steps, while for the NRL = 172 and NRL = 177 systems, the trajectory length was consistently set to 1 million steps. All simulations were performed at 300 K using LAMMPS on CPUs,⁸⁰ and the pairwise distances between each pair of nucleosome geometric centers (i.e., $(d_{12}, d_{13}, d_{14}, d_{23}, d_{24}, d_{34})$) were recorded every 500 steps.

Additionally, we conducted simulations for a system where the NRL = 167 tetra-nucleosome is embedded in a nucleosome solution. Due to the higher computational cost of this system, we limited the number of independent simulations to 530. Initial configurations of the tetra-nucleosome for these simulations were prepared by selecting structures from the centers of the 530 microstates constructed for the isolated NRL = 167 tetra-nucleosome. Then we placed each of the tetra-nucleosomes in the center of a cubic box with a 55 nm edge, and randomly placed 26 additional single nucleosomes to achieve a total nucleosome concentration of 0.3 mM. We relaxed each of these initial configurations with a 0.2 million-step NVT simulation, during which the tetra-nucleosome was fixed as a rigid body while single nucleosomes were free to move. From the relaxed configurations, we conducted production NVT simulations of at least 7 million steps, allowing all nucleosomes to evolve freely. Simulations

were performed at 300 K with OpenMM and OpenABC packages on GPUs.^{76,81} The pairwise distances between nucleosomes within the tetra-nucleosome, as well as the positions of all nucleosomes, were recorded every 500 steps for analysis.

Additional details on the system setup and MD simulations are provided in the *Supporting Information*.

Markov state modeling and transition path analysis for chromatin folding

We constructed independent microstate-MSMs and performed transition path analysis for all four systems following a consistent protocol, briefly outlined below. Additional details are provided in the *Supporting Information*.

(a) Select the converged segments of the trajectories and duplicate the converged trajectories according to the reflection symmetry of the nucleosome indices. All the following analyses are applied to the converged and duplicated trajectories.

(b) With six inter-nucleosomal distances d_{ij} as input features, apply the time-lagged independent component analysis (tICA) method with kinetic mapping algorithm to further reduce dimensionality and uncover independent collective variables (CVs) that better represent slow timescale dynamics.^{55–57,82}

(c) Group MD conformations into microstates by the K-Means algorithm according to their kinetic similarities based on the geometric distances in the tICA CV space. The hyperparameters of the tICA and clustering (i.e., tICA relaxation time, number of tICs, and number of microstates) are optimized using the cross-validation with the generalized matrix Rayleigh quotient (GMRQ) score.⁸³

(d) Construct and validate the microstate-MSMs by the Chapman-Kolmogorov (CK) test and implied time scale (ITS) analysis.^{25,26}

(e) Employ the Transition Path Theory (TPT)^{20,21,58} to elucidate the folding kinetic pathways from unfolded states to folded states and their corresponding fluxes.

(f) Lump multiple parallel kinetic pathways into a small set of metastable and representative reaction channels using Latent-space Path Clustering (LPC) algorithm to facilitate the understanding of folding mechanisms.^{27,32,84}

All analyses were conducted using in-house Python codes and codes based on MSM-Builder version 3.8.1.⁸⁵

IGME modeling of chromatin folding

To facilitate the interpretation of chromatin folding, for each simulation system, we modeled its folding dynamics with a six-macrostates IGME model³⁶ (Figure 2D). These six macrostates were constructed by lumping hundreds of microstates using the Robust Perron Cluster Analysis algorithm.^{86,87} The IGME model encodes non-Markovian dynamics through time integrations of memory kernels based on the GME.³⁶ Specifically, IGME model describes the evolution of transition probabilities over time using $\mathbf{T}(t \geq \tau_k) = \mathbf{A}\hat{\mathbf{T}}^t$, where $\mathbf{T}(t)$ is the transition probability matrix (TPM) at lag time t , and τ_k is the relaxation time of the memory kernels. The matrices \mathbf{A} and $\hat{\mathbf{T}}$ are constant matrices estimated from the MD simulation data.

We determined the relaxation time of the memory kernels, τ_k , by examining the convergence of the mean integral memory kernels. This can be quantified using two approaches: the first involves our previously developed quasi-MSM (qMSM)³⁸ to compute discretized kernels across various lag times, while the second directly utilizes the IGME to approximate the time-integrated memory kernels.³⁶ The \mathbf{A} and $\hat{\mathbf{T}}$ matrices were estimated using a set of TPMs at short lag times through least squares fitting with a Lagrangian approach. The optimal fitting range was selected by a systematic scanning to minimize the deviation between the predictions from IGEM model and raw MD data. The optimal IGME models were further validated and compared to MSMs using the Chapman-Kolmogorov test. All reported thermodynamic and kinetic properties of the metastable macrostates were derived from the optimal IGME models, which exhibit the smallest deviations from the raw MD data. More

details and results about IGME models are presented in the *Supporting Information*.

Acknowledgement

This work was supported by the National Institutes of Health (Grant R35GM133580) and the National Science Foundation (Grant MCB-2042362). X.H. acknowledge the support from the Hirschfelder Professorship Fund from University of Wisconsin-Madison and the Research Forward Fund from the University of Wisconsin-Madison Office of the Vice Chancellor for Research with funding from the Wisconsin Alumni Research Foundation. The computational resource support from Center for High Throughput Computing at the University of Wisconsin-Madison is also acknowledged.⁸⁸

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: Data presented in this study is available upon reasonable request to the corresponding author.

References

- (1) Luger, K.; Dechassa, M. L.; Tremethick, D. J. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nature reviews Molecular cell biology* **2012**, *13*, 436–447.
- (2) Rowley, M. J.; Corces, V. G. Organizational principles of 3D genome architecture. *Nature Reviews Genetics* **2018**, *19*, 789–800.
- (3) Jerkovic, I.; Cavalli, G. Understanding 3D genome organization by multidisciplinary methods. *Nature Reviews Molecular Cell Biology* **2021**, *22*, 511–528.
- (4) Lin, X.; Qi, Y.; Latham, A. P.; Zhang, B. Multiscale modeling of genome organization with maximum entropy optimization. *The Journal of chemical physics* **2021**, *155*.
- (5) Liu, S.; Athreya, A.; Lao, Z.; Zhang, B. From Nucleosomes to Compartments: Physicochemical Interactions Underlying Chromatin Organization. *Annual Review of Biophysics* **2024**, *53*.
- (6) 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 reveals a double helix twisted by tetranucleosomal units. *Science* **2014**, *344*, 376–380.
- (7) 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.
- (8) Williams, S.; Athey, B.; Muglia, L.; Schappe, R.; Gough, A.; Langmore, J. Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length. *Biophysical journal* **1986**, *49*, 233–248.
- (9) Dorigo, B.; Schalch, T.; Kulangara, A.; Duda, S.; Schroeder, R. R.; Richmond, T. J. Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* **2004**, *306*, 1571–1573.

- (10) Robinson, P. J.; Fairall, L.; Huynh, V. A.; Rhodes, D. EM measurements define the dimensions of the “30-nm” chromatin fiber: evidence for a compact, interdigitated structure. *Proceedings of the National Academy of Sciences* **2006**, *103*, 6506–6511.
- (11) 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.
- (12) 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.
- (13) 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. *Proceedings of the National Academy of Sciences* **2008**, *105*, 19732–19737.
- (14) Nishino, Y.; Eltsov, M.; Joti, Y.; Ito, K.; Takata, H.; Takahashi, Y.; Hihara, S.; Frangakis, A. S.; Imamoto, N.; Ishikawa, T., et al. Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *The EMBO journal* **2012**, *31*, 1644–1653.
- (15) 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.
- (16) Alvarado, W.; Agrawal, V.; Li, W. S.; Dravid, V. P.; Backman, V.; de Pablo, J. J.; Ferguson, A. L. Denoising Autoencoder Trained on Simulation-Derived Structures for Noise Reduction in Chromatin Scanning Transmission Electron Microscopy. *ACS Central Science* **2023**,

- (17) Ding, X.; Lin, X.; Zhang, B. Stability and folding pathways of tetra-nucleosome from six-dimensional free energy surface. *Nature communications* **2021**, *12*, 1091.
- (18) Liu, S.; Lin, X.; Zhang, B. Chromatin fiber breaks into clutches under tension and crowding. *Nucleic Acids Research* **2022**, *50*, 9738–9747.
- (19) Ricci, M. A.; Manzo, C.; García-Parajo, M. F.; Lakadamyali, M.; Cosma, M. P. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **2015**, *160*, 1145–1158.
- (20) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proceedings of the National Academy of Sciences* **2009**, *106*, 19011–19016.
- (21) Vanden-Eijnden, E., et al. Towards a theory of transition paths. *Journal of statistical physics* **2006**, *123*, 503–523.
- (22) 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.
- (23) Husic, B. E.; Pande, V. S. Markov state models: From an art to a science. *Journal of the American Chemical Society* **2018**, *140*, 2386–2396.
- (24) Konovalov, K. A.; Unarta, I. C.; Cao, S.; Goonetilleke, E. C.; Huang, X. Markov state models to study the functional dynamics of proteins in the wake of machine learning. *JACS Au* **2021**, *1*, 1330–1341.
- (25) Wang, W.; Cao, S.; Zhu, L.; Huang, X. Constructing Markov State Models to elucidate the functional conformational changes of complex biomolecules. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*, e1343.
- (26) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.;

- Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation. *The Journal of chemical physics* **2011**, *134*.
- (27) Qiu, Y.; O'Connor, M. S.; Xue, M.; Liu, B.; Huang, X. An Efficient Path Classification Algorithm Based on Variational Autoencoder to Identify Metastable Path Channels for Complex Conformational Changes. *Journal of Chemical Theory and Computation* **2023**, *19*, 4728–4742.
- (28) Chodera, J. D.; Noé, F. Markov state models of biomolecular conformational dynamics. *Current opinion in structural biology* **2014**, *25*, 135–144.
- (29) Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S. Molecular simulation of ab initio protein folding for a millisecond folder NTL9 (1- 39). *Journal of the American Chemical Society* **2010**, *132*, 1526–1528.
- (30) Silva, D.-A.; Bowman, G. R.; Sosa-Peinado, A.; Huang, X. A role for both conformational selection and induced fit in ligand binding by the LAO protein. *PLoS computational biology* **2011**, *7*, e1002054.
- (31) Shukla, D.; Meng, Y.; Roux, B.; Pande, V. S. Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nature communications* **2014**, *5*, 3397.
- (32) Liu, B.; Qiu, Y.; Goonetilleke, E. C.; Huang, X. Kinetic network models to study molecular self-assembly in the wake of machine learning. *MRS Bulletin* **2022**, *47*, 958–966.
- (33) Zeng, X.; Li, B.; Qiao, Q.; Zhu, L.; Lu, Z.-Y.; Huang, X. Elucidating dominant pathways of the nano-particle self-assembly process. *Physical Chemistry Chemical Physics* **2016**, *18*, 23494–23499.
- (34) Bowman, G. R.; Huang, X.; Pande, V. S. Network models for molecular kinetics and their initial applications to human health. *Cell research* **2010**, *20*, 622–630.

- (35) Da, L.-T.; Pardo Avila, F.; Wang, D.; Huang, X. A two-state model for the dynamics of the pyrophosphate ion release in bacterial RNA polymerase. *PLoS computational biology* **2013**, *9*, e1003020.
- (36) Cao, S.; Qiu, Y.; Kalin, M. L.; Huang, X. Integrative generalized master equation: A method to study long-timescale biomolecular dynamics via the integrals of memory kernels. *The Journal of Chemical Physics* **2023**, *159*.
- (37) Yik, A. K.-h.; Qiu, Y.; Unarta, I. C.; Cao, S.; Huang, X. *A Practical Guide to Recent Advances in Multiscale Modeling and Simulation of Biomolecules*; AIP Publishing LLC Melville, New York, 2023; pp 10–1.
- (38) Cao, S.; Montoya-Castillo, A.; Wang, W.; Markland, T. E.; Huang, X. On the advantages of exploiting memory in Markov state models for biomolecular dynamics. *The Journal of Chemical Physics* **2020**, *153*.
- (39) Qiu, Y.; Wiewiora, R. P.; Izaguirre, J. A.; Xu, H.; Sherman, W.; Tang, W.; Huang, X. Non-Markovian Dynamic Models Identify Non-Canonical KRAS-VHL Encounter Complex Conformations for Novel PROTAC Design. *JACS Au* **2024**,
- (40) Gibson, B. A.; Doolittle, L. K.; Schneider, M. W.; Jensen, L. E.; Gamarra, N.; Henry, L.; Gerlich, D. W.; Redding, S.; Rosen, M. K. Organization of chromatin by intrinsic and regulated phase separation. *Cell* **2019**, *179*, 470–484.
- (41) Zhang, M.; Díaz-Celis, C.; Onoa, B.; Cañari-Chumpitaz, C.; Requejo, K. I.; Liu, J.; Vien, M.; Nogales, E.; Ren, G.; Bustamante, C. Molecular organization of the early stages of nucleosome phase separation visualized by cryo-electron tomography. *Molecular cell* **2022**, *82*, 3000–3014.
- (42) Armeev, G. A.; Kniazeva, A. S.; Komarova, G. A.; Kirpichnikov, M. P.; Shaytan, A. K. Histone dynamics mediate DNA unwrapping and sliding in nucleosomes. *Nature communications* **2021**, *12*, 2387.

- (43) Peng, Y.; Li, S.; Onufriev, A., et al. Binding of regulatory proteins to nucleosomes is modulated by dynamic histone tails. *Nat Commun* **12**: 5280. 2021.
- (44) Li, S.; Wei, T.; Panchenko, A. Histone variant H2A. Z modulates nucleosome dynamics to promote DNA accessibility. *Biophysical Journal* **2023**, *122*, 218a.
- (45) Woods, D. C.; Rodríguez-Roperero, F.; Wereszczynski, J. The Dynamic Influence of Linker Histone Saturation within the Poly-Nucleosome Array. *Journal of Molecular Biology* **2021**, *433*, 166902.
- (46) Winogradoff, D.; Aksimentiev, A. Molecular Mechanism of Spontaneous Nucleosome Unraveling. *Journal of Molecular Biology* **2019**, *431*, 323–335.
- (47) 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. *Journal of molecular biology* **2000**, *298*, 937–953.
- (48) 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. *The Journal of chemical physics* **2013**, *139*.
- (49) Lequieu, J.; Córdoba, A.; Schwartz, D. C.; de Pablo, J. J. Tension-dependent free energies of nucleosome unwrapping. *ACS central science* **2016**, *2*, 660–666.
- (50) Moller, J.; Lequieu, J.; de Pablo, J. J. The free energy landscape of internucleosome interactions and its relation to chromatin fiber structure. *ACS central science* **2019**, *5*, 341–348.
- (51) Leicher, R.; Ge, E. J.; Lin, X.; Reynolds, M. J.; Xie, W.; Walz, T.; Zhang, B.; Muir, T. W.; Liu, S. Single-molecule and in silico dissection of the interaction between

- Polycomb repressive complex 2 and chromatin. *Proceedings of the National Academy of Sciences* **2020**, *117*, 30465–30475.
- (52) Latham, A. P.; Zhang, B. On the Stability and Layered Organization of Protein-DNA Condensates. *Biophysical Journal* **2022**, *121*, 1727–1737.
- (53) Lin, X.; Leicher, R.; Liu, S.; Zhang, B. Cooperative DNA Looping by PRC2 Complexes. *Nucleic Acids Research* **2021**, *49*, 6238–6248.
- (54) Farr, S. E.; Woods, E. J.; Joseph, J. A.; Garaizar, A.; Collepardo-Guevara, R. Nucleosome Plasticity Is a Critical Element of Chromatin Liquid–Liquid Phase Separation and Multivalent Nucleosome Interactions. *Nature Communications* **2021**, *12*, 1–17.
- (55) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *The Journal of chemical physics* **2013**, *139*, 07B604_1.
- (56) Noé, F.; Clementi, C. Kinetic distance and kinetic maps from molecular dynamics simulation. *Journal of chemical theory and computation* **2015**, *11*, 5002–5011.
- (57) Schwantes, C. R.; Pande, V. S. Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *Journal of chemical theory and computation* **2013**, *9*, 2000–2009.
- (58) Metzner, P.; Schütte, C.; Vanden-Eijnden, E. Transition path theory for Markov jump processes. *Multiscale Modeling & Simulation* **2009**, *7*, 1192–1219.
- (59) Zhang, B.; Zheng, W.; Papoian, G. A.; Wolynes, P. G. Exploring the free energy landscape of nucleosomes. *Journal of the American Chemical Society* **2016**, *138*, 8126–8133.
- (60) Parsons, T.; Zhang, B. Critical role of histone tail entropy in nucleosome unwinding. *The Journal of chemical physics* **2019**, *150*.

- (61) Lin, X.; Zhang, B. Explicit Ion Modeling Predicts Physicochemical Interactions for Chromatin Organization. *12*, RP90073.
- (62) Dombrowski, M.; Engholm, M.; Dienemann, C.; Dodonova, S.; Cramer, P. Histone H1 binding to nucleosome arrays depends on linker DNA length and trajectory. *Nature structural & molecular biology* **2022**, *29*, 493–501.
- (63) Zhurkin, V. B.; Norouzi, D. Topological polymorphism of nucleosome fibers and folding of chromatin. *Biophysical Journal* **2021**, *120*, 577–585.
- (64) Szerlong, H. J.; Hansen, J. C. Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. *Biochemistry and Cell Biology* **2011**, *89*, 24–34.
- (65) Correll, S. J.; Schubert, M. H.; Grigoryev, S. A. Short nucleosome repeats impose rotational modulations on chromatin fibre folding. *The EMBO journal* **2012**, *31*, 2416–2426.
- (66) Collepardo-Guevara, R.; Schlick, T. Chromatin fiber polymorphism triggered by variations of DNA linker lengths. *Proceedings of the National Academy of Sciences* **2014**, *111*, 8061–8066.
- (67) Kenzaki, H.; Takada, S. Linker DNA length is a key to tri-nucleosome folding. *Journal of Molecular Biology* **2021**, *433*, 166792.
- (68) Alvarado, W.; Moller, J.; Ferguson, A. L.; de Pablo, J. J. Tetranucleosome interactions drive chromatin folding. *ACS Central Science* **2021**, *7*, 1019–1027.
- (69) Garcia-Saez, I.; Menoni, H.; Boopathi, R.; Shukla, M. S.; Soueidan, L.; Noirclerc-Savoye, M.; Le Roy, A.; Skoufias, D. A.; Bednar, J.; Hamiche, A., et al. Structure of an H1-bound 6-nucleosome array reveals an untwisted two-start chromatin fiber conformation. *Molecular cell* **2018**, *72*, 902–915.

- (70) Lewis, T. S.; Sokolova, V.; Jung, H.; Ng, H.; Tan, D. Structural basis of chromatin regulation by histone variant H2A. *Z. Nucleic Acids Research* **2021**, *49*, 11379–11391.
- (71) Brandani, G. B.; Gopi, S.; Yamauchi, M.; Takada, S. Molecular dynamics simulations for the study of chromatin biology. *Current Opinion in Structural Biology* **2022**, *77*, 102485.
- (72) Voong, L. N.; Xi, L.; Sebeson, A. C.; Xiong, B.; Wang, J.-P.; Wang, X. Insights into nucleosome organization in mouse embryonic stem cells through chemical mapping. *Cell* **2016**, *167*, 1555–1570.
- (73) Li, Y.; Zhang, H.; Li, X.; Wu, W.; Zhu, P. Cryo-ET study from in vitro to in vivo revealed a general folding mode of chromatin with two-start helical architecture. *Cell Reports* **2023**, *42*.
- (74) Hou, Z.; Nightingale, F.; Zhu, Y.; MacGregor-Chatwin, C.; Zhang, P. Structure of Native Chromatin Fibres Revealed by Cryo-ET in Situ. *Nature Communications* **2023**, *14*, 6324.
- (75) Latham, A. P.; Zhang, B. Unifying Coarse-Grained Force Fields for Folded and Disordered Proteins. *Current Opinion in Structural Biology* **2022**, *72*, 63–70.
- (76) Liu, S.; Wang, C.; Latham, A. P.; Ding, X.; Zhang, B. OpenABC enables flexible, simplified, and efficient GPU accelerated simulations of biomolecular condensates. *PLOS Computational Biology* **2023**, *19*, e1011442.
- (77) Freeman, G. S.; Hinckley, D. M.; Lequieu, J. P.; Whitmer, J. K.; De Pablo, J. J. Coarse-grained modeling of DNA curvature. *The Journal of chemical physics* **2014**, *141*, 10B615_1.
- (78) Lu, W.; Bueno, C.; Schafer, N. P.; Moller, J.; Jin, S.; Chen, X.; Chen, M.; Gu, X.; Davtyan, A.; de Pablo, J. J., et al. OpenAWSEM with Open3SPN2: A fast, flexible,

- and accessible framework for large-scale coarse-grained biomolecular simulations. *PLoS computational biology* **2021**, *17*, e1008308.
- (79) Noel, J. K.; Levi, M.; Raghunathan, M.; Lammert, H.; Hayes, R. L.; Onuchic, J. N.; Whitford, P. C. SMOG 2: a versatile software package for generating structure-based models. *PLoS computational biology* **2016**, *12*, e1004794.
- (80) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D., et al. LAMMPS—a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications* **2022**, *271*, 108171.
- (81) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D., et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology* **2017**, *13*, e1005659.
- (82) Naritomi, Y.; Fuchigami, S. Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis. *The Journal of Chemical Physics* **2013**, *139*, 12B605_1.
- (83) McGibbon, R. T.; Pande, V. S. Variational cross-validation of slow dynamical modes in molecular kinetics. *The Journal of chemical physics* **2015**, *142*, 03B621_1.
- (84) Meng, L.; Sheong, F. K.; Zeng, X.; Zhu, L.; Huang, X. Path lumping: an efficient algorithm to identify metastable path channels for conformational dynamics of multi-body systems. *The Journal of Chemical Physics* **2017**, *147*, 044112.
- (85) Beauchamp, K. A.; Bowman, G. R.; Lane, T. J.; Maibaum, L.; Haque, I. S.; Pande, V. S. MSMBuilder2: modeling conformational dynamics on the picosecond to millisecond scale. *Journal of chemical theory and computation* **2011**, *7*, 3412–3419.

- (86) Deuffhard, P.; Weber, M. Robust Perron cluster analysis in conformation dynamics. *Linear algebra and its applications* **2005**, *398*, 161–184.
- (87) Röblitz, S.; Weber, M. Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. *Advances in Data Analysis and Classification* **2013**, *7*, 147–179.
- (88) Center for High Throughput Computing, Center for High Throughput Computing. 2006; <https://chtc.cs.wisc.edu/>.