Biophysical Reviews
https://doi.org/10.1007/s12551-024-01195-5

REVIEW

Regulation of chromatin architecture by protein binding: insights from molecular modeling

Stephanie Portillo-Ledesma[1,4] · **Tamar Schlick**[1,2,3,4]

Received: 14 November 2023 / Accepted: 22 April 2024

Abstract

Histone and non-histone proteins play key roles in the activation and repression of genes. In addition to experimental studies of their regulation of gene expression, molecular modeling at the nucleosome, chromatin, and chromosome levels can contribute insights into the molecular mechanisms involved. In this review, we provide an overview for protein-bound chromatin modeling, and describe how our group has integrated protein binding into genome systems across the scales, from all-atom to coarse-grained models, using explicit to implicit descriptions. We describe the associated applications to protein binding effects and biological mechanisms of genome folding and gene regulation. We end by illustrating the application of machine learning tools like AlphaFold2 to proteins relevant to chromatin systems.

Keywords Histone and non-histone proteins · Molecular modeling · Nucleosome · Chromatin · Chromosome

Introduction

Histone and non-histone proteins are a diverse group of biomolecules that bind to eukaryotic DNA to regulate genome folding and gene expression (Lobbia et al. 2021). These biopolymers play crucial roles in many nuclear processes, such as nucleosome remodeling, chromatin folding and looping, and gene activation and repression. Through post-translational modifications, they add an additional layer of complexity in the regulation of nuclear activity.

Histone proteins include the core histones H2A, H2B, H3, and H4, which assemble together to form the nucleosome core, and the linker histone (LH) H1 that binds to the linker DNA at the nucleosome entry/exit sites to further compact chromatin. LH proteins are also involved in DNA replication and repair, as well as genome stability. Multiple variants of H1 exist in higher eukaryotes, and they can bind to nucleosomes with different stoichiometries (from less than 1 to more than 1 LH per nucleosome) (Woodcock et al. 2006; Bates and Thomas 1981) and binding modes (on/off-dyad binding) (Zhou et al. 2015; Öztürk et al. 2018; Zhou et al. 2013, 2016; Cutter and Hayes 2017).

Non-histone proteins include transcription factors (TF) and repressors like the heterochromatin protein 1 (HP-1) and polycomb group proteins. Also involved are structural proteins like the high mobility group (HMG) proteins, CTCF, and cohesin. TFs bind to linker or nucleosomal DNA to regulate the transcription of genes. Some TFs bind to promoter sequences near the transcription start sites and help initiate transcription, and others bind to regulatory sequences, such as enhancers, to activate or repress transcription. HP-1 and polycomb group proteins bind to chromatin to limit DNA accessibility and thus mediate transcriptional repression. HP1 recognizes histone H3 methylated at Lys9 (Bannister et al. 2001; Lachner et al. 2001), and polycomb group proteins recognize genomic sequences termed polycomb responsive elements (PRE) located near target genes and mediate the methylation of histone H3 at Lys27 (Plath et al. 2003; Czermin et al. 2002; Kuzmichev et al. 2002; Müller et al. 2002), a repressive mark. Structural non-histone proteins bind to DNA and nucleosomes to induce structural changes in the chromatin fiber, influencing chromatin dynamics. HMG pro-

✉ Tamar Schlick
schlick@nyu.edu

1 Department of Chemistry, 100 Washington Square East, Silver Building, New York University, New York, NY 10003, USA

2 Courant Institute of Mathematical Sciences, New York University, 251 Mercer St., New York, NY 10012, USA

3 New York University-East China Normal University Center for Computational Chemistry, New York University Shanghai, Shanghai 200122, China

4 Simons Center for Computational Physical Chemistry, 24 Waverly Place, Silver Building, New York University, New York, NY 10003, USA

Published online: 09 May 2024

teins bind preferentially to AT-rich sequences in the DNA and induce conformational changes that promote the recruitment of additional binding proteins (Elton et al. 1987; Reeves and Nissen 1990). CTCF and cohesin act together to execute loop extrusion and create topologically associated domains (TADs) (Fudenberg et al. 2016, 2017).

Molecular modeling studies have helped explore the mechanisms of chromatin and gene regulation by protein binding. In the next section, we review the state-of-the-art of protein-bound chromatin modeling at different genomic scales, from nucleosomes to Mb fibers. Then, we illustrate protein modeling by our group to provide mechanistic information on the role of proteins in chromatin fiber architecture and dynamics at nucleosome resolution. We end with a discussion on how machine learning tools like AlphaFold2 can help integrate protein effects into chromatin structures and processes.

State-of-the-art of protein-bound chromatin systems

While core histones have been extensively studied, the role of linker histones (LHs) in regulating chromatin architecture is less appreciated and somewhat nuanced. The first crystal structure of the chromatosome revealed the LH globular head binding to the nucleosome dyad (Zhou et al. 2015). Other experiments showed that LHs can also bind off-dyad (Zhou et al. 2013), and that the LH C-terminal domain (CTD) interacts mostly with one of the DNA linkers (Bednar et al. 2017; White et al. 2016), believed to direct LH assembly and chromatin higher-order. Many experiments demonstrated the importance of LH in establishing compact 30-nm fibers (Fyodorov et al. 2018). However, the architecture of these fibers and mechanisms of folding directed by LH were not well understood.

Numerous modeling approaches have explored the mechanisms of LH-induced chromatin folding, the effects of different LH binding modes and densities, and the role of LH variants.

From the prediction of the first chromatosome structure by docking and molecular modeling (Bharath et al. 2003), several chromatosomes have been modeled with different LH isoforms and binding modes, at all-atom and coarse-grained levels, and with conventional or enhanced sampling simulations (Öztürk et al. 2020). Studies that focused on the LH binding mode and variant types demonstrated conformational variability (Öztürk et al. 2016), thermodynamic basis for on-dyad versus off-dyad binding (Woods and Wereszczynski 2020), and a mixture of nucleosomal DNA and linker DNA interactions (Zhou et al. 2021). When focusing on the role of the flexible and disordered LH N- and C-terminal domains, studies have shown that the CTD remains disordered upon binding (Sridhar et al. 2020a), leading to an asymmetric and dynamical nucleosome conformation, while the N-terminal domain transitions to a more ordered structure (Sridhar et al. 2020b), affecting LH binding affinity. These disordered domains help nucleosome compaction as the binding of LH variants with and without the domains produces deferentially compacted chromatosomes (Wu et al. 2021).

At the kb level, LH is coarse-grained as one (Lin and Zhang 2024) or several amino acids per bead (Arya and Schlick 2009; Luque et al. 2014; Lequieu et al. 2019; Sridhar et al. 2020a), modeled implicitly by modifying the linker DNA geometry (Kepper et al. 2008; Stehr et al. 2008), or derived from crystal or CryoEM structures of chromatosomes or nucleosomes (Wong et al. 2007; Izadi et al. 2016). Associated nucleosome arrays sampled by Monte Carlo (Arya and Schlick 2009; Sridhar et al. 2020a), Brownian Dynamics (Lequieu et al. 2019; Li et al. 2023), or Molecular Dynamics (Izadi et al. 2016; Woods et al. 2021) simulations have revealed LH effects on fiber architecture.

Studies of kb fibers with LH bound showed that: different fiber configurations are formed based on the orientation of the LH and length of the linker DNA (Wong et al. 2007), LH binding reduces fiber irregularity due to the formation of DNA stem as well as linker-length dependent local stem geometry (Stehr et al. 2008; Arya and Schlick 2009; Perišić et al. 2010; Portillo-Ledesma et al. 2022), LH limits DNA flexibility and stabilizes repeating tetra-nucleosomal units (Woods et al. 2021), and LH induces fiber compaction by changing the path for the entering and exiting linker DNA (Kepper et al. 2008; Arya and Schlick 2009). Salt-dependent simulations show that LH interacts preferentially with one linker DNA at high salt, but with both linkers at low salt (Luque et al. 2014). The disordered nature of the LH CTD favors looping and long-range interactions needed for genome regulation (Sridhar et al. 2020a). In the presence of explicit ions, LHs also interact with the linker DNA of neighboring nucleosomes, contributing to the overall chromatin compaction (Lin and Zhang 2024).

Overall, these studies helped define the mechanisms of fiber compaction by LH and the effect of LH binding on fiber architecture, which can be directly related to the regulation of gene expression. In the next section, we detail our implementation of coarse-grained LH within mesoscale chromatin fibers.

Other common chromatin structure modulators are transcription factors (TFs) and repressors. The binding of these proteins to nucleosomes and chromatin fibers has been studied with all-atom (MacCarthy et al. 2022; Huertas et al. 2020; Azzaz et al. 2014), coarse-grained (Tan and Takada 2020; Watanabe et al. 2018; Bajpai et al. 2017; Portillo-Ledesma et al. 2024), and polymer (Nicodemi and Prisco 2009; Brackley et al. 2013; Buckle et al. 2018) models to determine

mechanisms of genome activation and repression. Mechanisms of chromatin opening have been determined for the pioneer TF Oct4. The binding of Oct4 to a single nucleosome revealed that nucleosome motions, such as breathing and twisting, mediate nucleosome recognition by the pioneer TF (Huertas et al. 2020), and that Oct4 binding enhances nucleosome structural flexibility to favor open nucleosome conformations (MacCarthy et al. 2022). In the presence of another TF, Sox2, an allosteric effect has been shown: the binding of a first TF changes the nucleosome conformation, affecting the binding of a second TF (Tan and Takada 2020). For mechanisms of chromatin repression, the binding of HP1 to a H3K9 trimethylated dinucleosome revealed compaction by bridging adjacent nucleosomes (Watanabe et al. 2018), possibly mediated by large range motions of HP1's N-terminal chromo domain (Azzaz et al. 2014).

Studies of kb fibers that incorporate protein binding implicitly have illuminated mechanisms of chromatin compaction and regulation. For example, the binding of TFs creates microdomains that are dependent on the binding position and are regulated by LH and tail acetylation (Portillo-Ledesma et al. 2024). The remodeler protein HMG bends the linker DNA, destabilizing the regular 30-nm chromatin structure (Bajpai et al. 2017).

When focusing on chromosomes or Mb systems, polymer models incorporating TFs like NFκB or CTCF have shed light on protein-induced chromosome organization. Studies showed that chromosome architecture is dictated by protein binding based on thermodynamic preferences (Nicodemi and Prisco 2009), proteins organize the genome by clustering along the chromosomes based on entropic contributions (Brackley et al. 2013), and TF binding induces chromosome loops to help form TADs and compartments (Brackley et al. 2016). For smaller Mb regions, active epigenetic marks for CTCF can predict the overall folding (Buckle et al. 2018).

Although many advances have been made at the nucleosome and kb levels (Brandani et al. 2022; Portillo-Ledesma et al. 2023; Huertas et al. 2022), few studies have provided detailed mechanistic information on the role of proteins in chromatin fiber architecture and dynamics at nucleosome resolution. Next, we summarize the efforts made in our group to bridge this gap and discuss the additional modeling advances needed to better integrate protein effects on chromatin processes (see Figs. 1 and 2).

Linker histones and antibodies increase chromatin fiber compaction

Rigid 3-bead LH model As a first-order approach to incorporate LH in our nucleosome resolution chromatin model (Bascom and Schlick 2017; Portillo-Ledesma and Schlick 2020), we first introduced LH as a 3-bead rigid element (Arya and Schlick 2009). This coarse-grained LH model first demonstrated that LH reduces the DNA entry/exit angle, which promotes the formation of a rigid linker DNA "stem" and brings nucleosomes closer to each other. As a result, internucleosome interactions between i and i $\pm$ 2 nucleosomes increase, leading to the formation of a more regular and rigid chromatin fiber (Arya and Schlick 2009). For fibers with linker DNA typical of eukaryotes, it also showed that when combined with Mg^{2+} ions, it favors heteromorphic and highly compact configurations with high DNA bending and sequential nucleosome interactions (i $\pm$ 1) (Grigoryev et al. 2009). By further studying nucleosome arrays with increasing linker DNA lengths, we determined that the effect of LH is optimal, namely a tight DNA stem is formed, when the length of the linker DNA is similar to the LH length (Perišić et al. 2010).

In particular, the coarse-grained LH model is based on the rat H1.4. The first bead is derived from the atomic structure of the globular head (GH) (76 residues) (Bharath et al. 2003), and the other two beads represent the CTD (111 residues). The short N-terminal domain of 33 residues is not included in the model. The three beads are positioned in a straight line and each bead has a charge in its center parameterized for the Debye-Hückel potential so it reproduces the Poisson-Boltzmann electric field of the atomistic LH. LH beads interact with all other chromatin elements, namely the nucleosome cores, histone tails H2A, H2B, H3, and H4, and linker DNA, through excluded volume (Lennard–Jones potential) and electrostatic interactions (Debye-Hückel potential).

Dynamic LH binding To improve the description of LH, we developed a dynamic model for LH binding in which LHs were allowed to detach and reattach from the nucleosome cores based on binding and unbinding probabilities (P_{bind} and P_{unbind}) (Collepardo-Guevara and Schlick 2011). By changing the probabilities of attaching/detaching, we can simulate different binding affinities or fast/slow binding, and different effective LH densities (ρ). For example:

$$P_{bind} \gg P_{unb}: \text{ Very high affinity and very fast binding,} \tag{1a}$$

$$P_{bind} > P_{unb}: \text{ High affinity and fast binding,} \tag{1b}$$

$$P_{bind} = P_{unb}: \text{ Moderate affinity,} \tag{1c}$$

$$P_{bind} < P_{unb}: \text{ Low affinity and slow binding,} \tag{1d}$$

where LH density $\rho = P_{bind} / (P_{bind} + P_{unbind})$ (Portillo-Ledesma et al. 2022).

With this simple but effective model of LH dynamic binding, we determined that fast and slow LH binding, which occurs simultaneously in vivo, might act cooperatively to help control fiber unfolding, especially when the linker DNA length is typical of eukaryotes (Collepardo-Guevara and

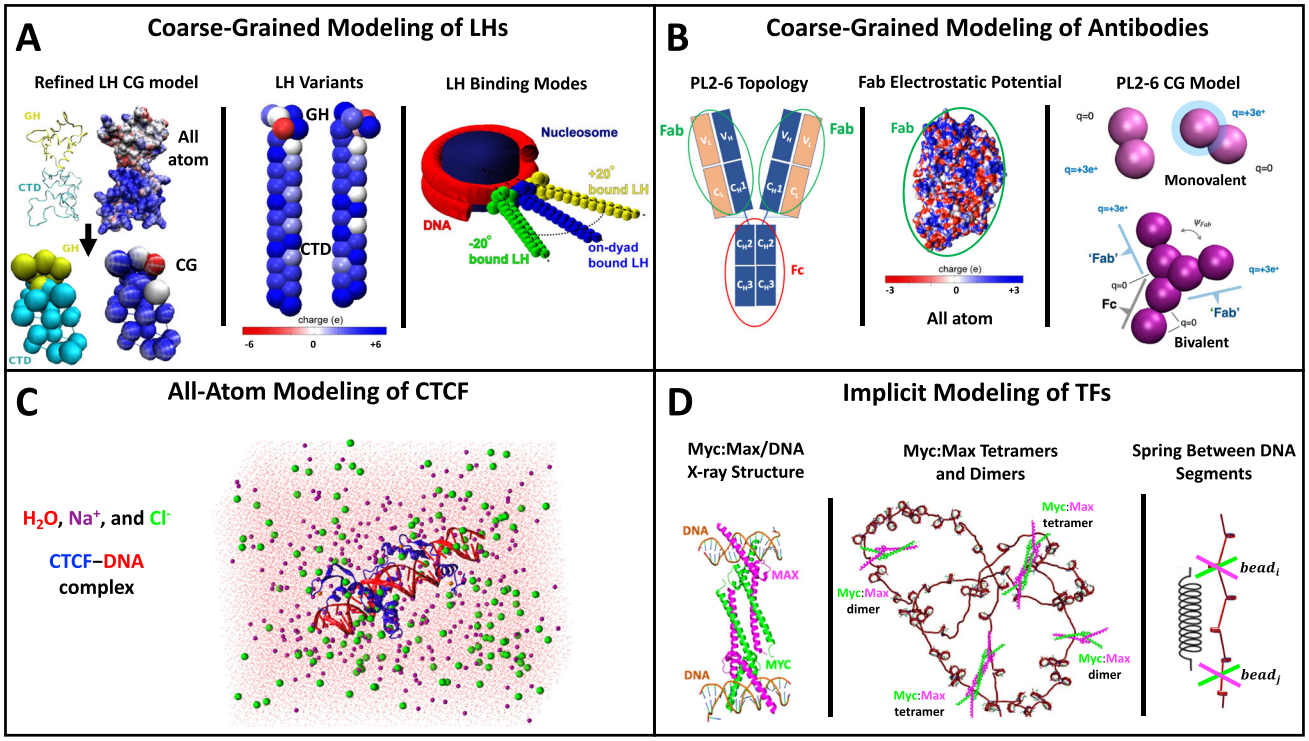

Fig. 1 Modeling protein binding in genome structures. **A** Current implementation of flexible LH binding. Left: Coarse-grained model of the flexible 28-bead LH H1E from the atomistic structure with 6 GH beads in yellow and 22 CTD beads in cyan (Luque et al. 2014). Middle: Coarse-grained model for the two LH variants H1E and H1C (27 beads) (Perisic et al. 2019) with beads colored based on their charges. Right: Chromatosome with different LH biding modes: off-dyad −20° (green), on-dyad (blue), and off-dyad +20° (yellow) (Perisic et al. 2019). **B** Coarse-grained modeling of diffusive monovalent (2 magenta beads) and bivalent (6 magenta beads) antibodies based on the Fab all-atom structure and electrostatic potential (Myers et al. 2020). **C** All-atom model containing explicit water (red dots) and ions (magenta and green balls) for the wildtype CTCF–DNA complex based on its experimental crystal structure (Mao et al. 2023). **D** Implicit modeling of TF binding to chromatin fibers at nucleosome resolution and kb chromatin arrays based on the crystal structure of Myc:Max binding to DNA (Portillo-Ledesma et al. 2024). Shown from left to right are: Myc:Max–DNA crystal structure; binding of Myc:Max complexes to chromatin fibers forming tetramers; and implicit modeling of Myc:Max binding through a spring between two DNA beads

Schlick 2011). Fibers with fixed LH behave more stiffly and unfold based on interactions between linker DNA and LH, whereas dynamic LH softens the unfolding dynamics due to destabilization of the DNA stem and increase of the DNA–DNA repulsion. Our later study of dynamic LH binding in the presence of Mg^{2+} ions (Collepardo-Guevara and Schlick 2012) showed reduction of the Mg^{2+} stiffening effect. Both dynamic LH binding and divalent ions act together to promote heteromorphic superbeads-on-a-string structures during fiber unfolding, which helps decondense the chromatin fiber.

Flexible 28-bead LH model Next, our group improved the LH coarse-grained model (Luque et al. 2014) to describe the interaction between the intrinsically disordered CTD and the linker DNA, and the CTD folding dynamics upon nucleosome binding. Our refined LH model consists on 6 rigid beads for the GH, also based on the atomic structure of LH H1.4 (H1E) (Bharath et al. 2003), and 22 flexible beads for the CTD, with a resolution of 5 residues per bead, similar to histone tails (Fig. 1A). The CTD initial configuration was based on experimental FRET data showing a compressed but elongated configuration of length 10 nm. Because the CTD 22 beads are flexible, we associated stretching and bending terms as:

$$E_{LH_S} = \sum_{i=1}^{N_{LH}} \sum_{j=1}^{N_i} \frac{k_j}{2}(l_{ij} - l_{j0})^2, \tag{2}$$

where N_{LH} is the number of nucleosome cores with LH bound and N_i is the number of beads used to calculate the stretching energy. The constants are $k_j = 0.1\ kcal/mol/\text{Å}$ and $l_{j0} = 15$ Å for the CTD beads, and $l_{j0} = 0$ for the globular head beads connected to the CTD.

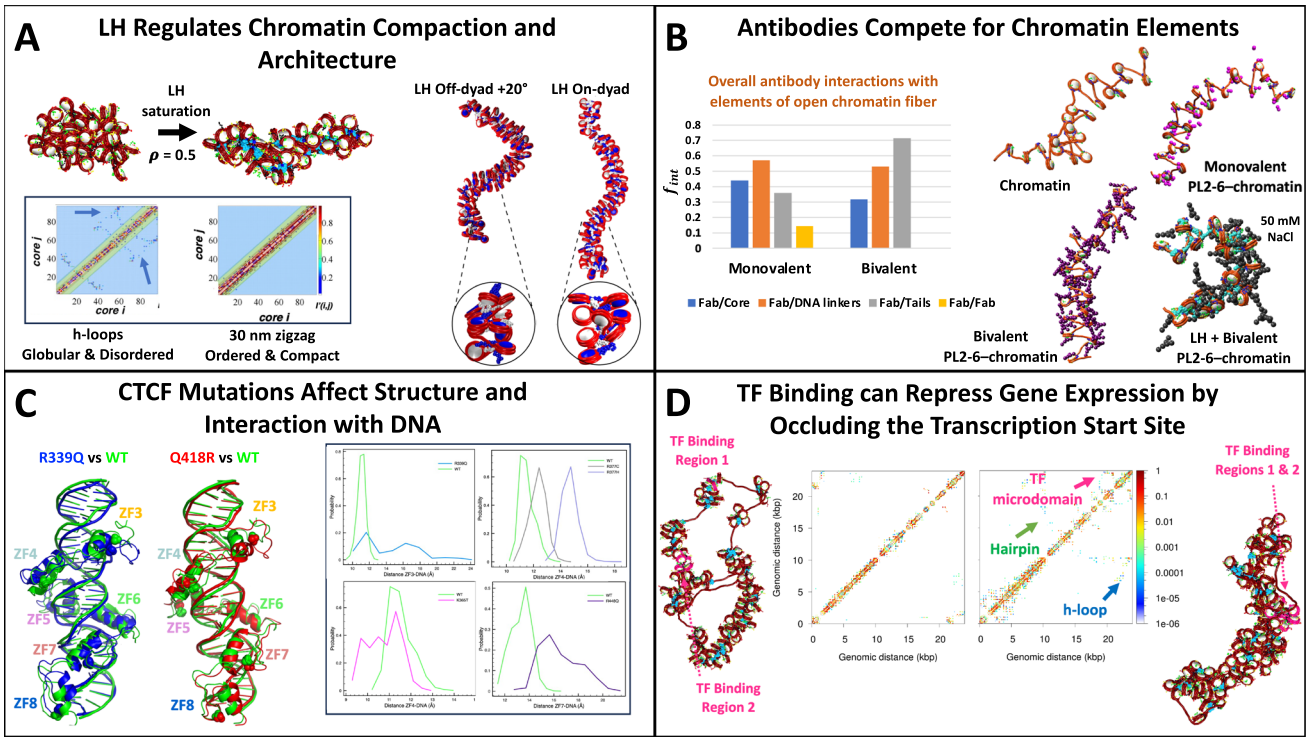

Fig. 2 Biophysical insights obtained from modeling protein-bound genome systems. **A** Binding of LH to chromatin fibers produces a transition from a globular disordered fiber with hierarchical loops to an ordered zigzag compact fiber (Grigoryev et al. 2016; Portillo-Ledesma et al. 2022). Shown are chromatin structures in the absence (left) and presence (right) of LH and their corresponding nucleosome contact maps depicting long and medium-range interactions in the fiber without LH. LH off-dyad +20° binding mode produces more compact fibers than when bound on-dyad (Perisic et al. 2019). **B** Coarse-grained modeling of diffusive antibody binding to chromatin fibers reveals that bivalent antibodies are better compactors than monovalent antibodies and that bivalent antibodies act together with LH to compact the fibers. As shown by the plot of interaction frequencies, such compaction is driven by the competition for interaction with tails (Myers et al. 2020). **C** All-atom MD simulations show that single-residue mutations of CTCF affect its secondary structure and interaction with DNA, as seen here by the CTCF structure of the wildtype and mutants R339Q and Q418R, and by the distances between CTCF's zinc fingers (ZFs) and DNA (Mao et al. 2023). **D** Implicit modeling of TF binding to chromatin fibers reveals the mechanism of gene silencing in the EED gene locus by occlusion of the transcription start site driven by higher folding motifs like hairpins and hierarchical loops (Portillo-Ledesma et al. 2024)

The corresponding bending term is:

$$E_{LH_B} = \sum_{i=1}^{N_{LH}} \sum_{j=1}^{N_i - 1} \frac{k_{\beta_j}}{2} (\beta_{ij} - \beta_{j0})^2, \tag{3}$$

where $k_{\beta_j} = 1$ and $\beta_{j0} = 110^o$.

Similar to the initial 3-bead rigid model, each LH bead interacts with other chromatin elements through an electrostatic Debye-Hückel (E_{DH}) term of the form:

$$E_{DH}(q_i, q_j, r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}} \exp(-\kappa r_{ij}), \tag{4}$$

where q_i, q_j are the charges on the two beads, r_{ij} the distance between the two beads, ε the dielectric constant, and κ the inverse Debye length. Also included is a Lennard–Jones term (E_{LJ}) defined as:

$$E_{LJ}(\sigma, k_{ev}, r_{ij}) = k_{ev}\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right], \tag{5}$$

where k_{ev} is the excluded volume interaction energy parameter, and σ is the effective diameter of the two interacting beads.

The LH CTD 22 beads are sampled with translational moves during the Monte Carlo simulation. This model accurately reproduces the CTD condensation upon LH binding and its dependency on chromatin fiber size (Luque et al. 2014).

With this refined LH model, we showed that upon binding, the CTD folds in the nucleosome together with other nucleosome elements, which is controlled by a charge neutralization

mechanism. We also showed that LH density controls chromatin higher-order folding producing well-defined 30-nm zigzag fibers when the fiber is LH-saturated and flexible and looped structures when saturation levels are reduced (Grigoryev et al. 2016) (Fig. 2A). Such a hierarchical looping mechanism for fiber folding establishes an efficient way to package DNA that facilitates DNA unwrapping without knotting. LH also affects the persistence length of the looped structures; higher LH densities decrease DNA bending, suggesting that LH can help regulate gene expression by modulating long-range interactions (Bascom et al. 2016). We finally determined that the critical LH density associated with the formation of the well-defined 30-nm zigzag fibers increases with the linker DNA length (Luque et al. 2016), indicating that a balance between the negatively charged DNA and positively charged LH determines optimal stems and nucleosome array structures.

In our initial studies with this updated LH model, the globular head was considered to be bound on-dyad, thus symmetrically centered on the nucleosome dyad axis. However, LH can bind asymmetrically, or off-dyad (Zhou et al. 2015; Öztürk et al. 2018; Zhou et al. 2013; Cutter and Hayes 2017; Zhou et al. 2016), and LH densities higher than 1 are possible (Woodcock et al. 2006; Bates and Thomas 1981). To better address these factors, and the fact that in mammals 11 LH variants exits (Fyodorov et al. 2018), we further introduced two off-dyad binding modes (Fig. 1A), chromatosomes with two LHs, and derived a model for the shorter LH variant H1.2 (Fig. 1A) (Perisic et al. 2019). In particular, H1.2 (H1C) was modeled similarly to H1.4 (Luque et al. 2014), with off-dyad binding modes. These were modeled by rotating the LH around that axis, $-20°$ or $+20°$, based on the experimental structure of a tetranucleosome with off-dyad LH (Song et al. 2014). A chromatosome with 2 LHs was formulated by binding 2 LHs with different binding modes.

This study showed that chromatin structure depends on LH variant and binding mode and that the off-dyad binding of H1.4 is advantageous for higher compaction due to strong interactions with nonparental DNA and promotion of high tail/nonparental core interactions (Fig. 2A). We also found that when two LHs are bound to the same nucleosome, the most favorable organization is the combination of an on-dyad with an off-dyad binding mode. Overall, these results show how LH variants, binding modes, and densities orchestrate chromatin condensation and DNA accessibility, adding an additional level of epigenetic regulation.

Similarly, we also introduced a dynamic binding for the updated H1.2 and H1.4 LHs (Portillo-Ledesma et al. 2022). Because this LH model is larger than the initial 3-bead model and is flexible, we introduced an additional LH Monte Carlo move to reorganize the LHs upon binding and thus avoid clashing with other chromatin elements. We used this model to determine the critical LH density that triggers a structural transition from a globular, open, and disordered state to a compact, ordered, and straight state, as found in connection with B-cell lymphoma development (Yusufova et al. 2021). A density of 0.5 triggered this transition in fibers with linker DNA length typical of eukaryotes due to the optimal DNA stem formation (Fig. 2A).

Antibody/chromatin modeling Antibody proteins can bind to linker DNA or nucleosomes, which is used to probe chromatin states and sub-structures (Olins et al. 2011) with many biological applications (Seredkina et al. 2013). To gain a better understanding of how antibodies interact with chromatin and compete with other chromatin elements such as linker DNA, histone tails, and LH, we introduced coarse-grained models of the monovalent (fragment crystallizable, Fc, region removed) and bivalent (antigen binding region, Fab, and Fc regions present) forms of the PL2-6 antibody into our chromatin mesoscale model (Myers et al. 2020). Experiments indicate that whereas the bivalent PL2-6 interacts with mitotic chromatin, the monovalent form interacts with chromatin throughout the cell cycle, indicating different binding modes (Olins et al. 2011).

The modeling of diffusive antibodies required incorporating periodic boundary conditions to examine crowding effects. Our study of the electrostatic surface of the atomistic PL2-6 showed a basic region at the binding end of Fab subunits (Fig. 1B), indicating that the interaction with chromatin occurs through the linker DNA or the nucleosomal acidic path. Thus, our coarse-grained antibody model includes two rigid spherical beads for each Fab subunit and two rigid spherical beads for the Fc subunit, totaling 6 beads for the bivalent antibody and 2 beads for the monovalent antibody (Fig. 1B). The implementation of rigid beads was rationalized by data showing that the angle between the two Fab domains (Ψ_{fab} in Fig. 1B) does not affect antibody properties (Calero-Rubio et al. 2016); the model is also simpler (Calero-Rubio et al. 2018). The charge on the end bead of each Fab subdomain is defined as +3q to mimic the basic electrostatics, whereas the other beads have a charge of zero. Similar to the LH modeling, each antibody bead interacts with LHs, tails, cores, linker DNA, and other antibody beads through electrostatic terms (Debye-Hückel potential) and excluded volume terms (Lennard–Jones potential) as:

$$E_{Fab}(\sigma, k_{ev}, r_{ij}, q_i, q_j) = k_{ev}\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon r_{ij}} \exp(-\kappa r_{ij}). \quad (6)$$

The antibodies move in the Monte Carlo simulation with two additional translational and rotational moves. As men-

tioned above, a new development for this type of study between antibodies and chromatin requires consideration of periodic boundary conditions. For that purpose, a 400-nm box is centered around the chromatin fiber.

Our study of antibody binding to chromatin compared simulations of free chromatin fibers, monovalent PL2-6–chromatin, and bivalent PL2-6–chromatin complexes at low and high salt concentrations. We found that bivalent antibodies have more intense interactions than the monovalent counterparts, increasing fiber compaction (Fig. 2B). These interactions result from antibodies competing with cores and linker DNA for histone tails (Fig. 2B). Additionally, we found that LH and PL2-6 act cooperatively to compact chromatin fibers. Thus, overall we showed that dynamic interactions of proteins with chromatin fibers depend on the fiber's internal structure and the interactions already established with other chromatin elements. This competition dynamically alters the internal chromatin structure and can in turn modulate the interaction with other proteins like LHs, HP1, or HMG.

Overall, we have demonstrated that at the coarse-grained level, the interaction of proteins like LHs or antibodies with chromatin fibers is possible to model explicitly. With simple, like the 3-bead rigid LH, or more sophisticated models, like the 28-bead flexible LH or 2–6-bead rigid but diffusive antibodies, insights into the impact of proteins on chromatin architecture and gene expression regulation can be obtained.

Transcription factors shape the chromatin fiber

Transcription factors (TFs) bind to chromatin regions to exert a variety of functions that are related to the regulation of gene expression. In particular, CTCF, an 11 zinc-finger (ZF) protein, acts together with cohesin to perform loop extrusion and create TADs. Because of CTCF's role in genome organization, single-residue mutations are associated with cancer development (Marshall et al. 2014; Filippova et al. 2002).

All-atom CTCF–DNA model To better understand how such single-residue mutations affect CTCF function, we compared the wildtype CTCF–DNA complex to mutant complexes (Mao et al. 2023). We focused on the effect that mutations had on CTCF structure, its interaction with DNA, and complex stability. Based on the experimental crystal structure of the wildtype CTCF–DNA complex containing ZFs 3 to 8 and a DNA chain of 27 nucleotides (Yin et al. 2017), we created the following mutant complexes: R339Q–DNA, R342C–DNA, S354T–DNA, K365T–DNA, R377H/C–DNA, Q418R–DNA, and R448Q–DNA detected in cancer patients (Walker et al. 2015; Yoshida et al. 2013; Voutsadakis 2018; Le Gallo et al. 2012), and performed all-atom microsecond molecular dynamics (MD) simulations with the AMBER force field (Weiner et al. 1984) (Fig. 1C).

We found that most of the mutations produce less stable complexes compared to the wildtype system. Depending on the specific mutation, this loss of stability is produced by major changes in the electrostatic potential, loss of stabilizing hydrogen bonds between the DNA and CTCF molecules, or destabilization of specific zinc fingers (Fig. 2C). These results provide insights at the molecular level on the effect that single-residue mutations have on CTCF and its function, explaining how small residue changes might translate into the development of cancer or other diseases.

Thus, with relatively small nucleic acid systems, high-resolution models at the atomic level can be used to study the interaction between proteins and DNA, and deduce mechanisms of action. The knowledge obtained at the all-atom level can be later used to introduce wildtype and mutant CTCF binding into our chromatin mesoscale model. For example, with CTCF modeled explicitly, as we did with LHs and antibodies (Luque et al. 2014; Perisic et al. 2019; Myers et al. 2020), the strength of the electrostatic interactions between the CTCF and linker DNA can be modulated by scaling down the Debye-Hückel energy term based on the binding energies obtained at the atomic level. Implicit modeling of CTCF via springs between DNA beads, similar to what we will discuss below for the Myc-Max TF binding, requires the force constant of the spring to be modulated to reflect different binding affinities.

Implicit Myc-Max TF modeling Although all-atom simulations of small DNA–protein complexes can provide insights into mechanisms of genome regulation, simulations at the chromatin level are needed to determine how chromatin architecture is affected by proteins and how this influences gene expression. We recently explored the binding of the Myc-Max pair of TFs to chromatin fibers with an implicit protein model (Fig. 1D) (Portillo-Ledesma et al. 2024). Myc, a leucine zipper (bHLHZip) protein, heterodimerizes with another bHLHZip protein, Max, to form the Myc:Max complex that binds to E-box (5′-CACGTG-3′) regulatory DNA elements throughout the genome to control transcription (Blackwood and Eisenman 1991). These heterodimers can further tetramerize and bridge sequence-distant regions of the genome (Nair and Burley 2003), acting as gene repressors.

Based on the crystal structure of the Myc:Max complex showing how the two dimers bind to sequence-distant DNA segments and bridge them together (Nair and Burley 2003), we model Myc:Max binding implicitly by adding restraints between two genome loci b_i and b_j (Fig. 1D) using a harmonic energy penalty of the form:

$$E_{b_i b_j} = k(l_{b_i b_j} - l_0)^2. \tag{7}$$

Here l_0 is 13 nm based on the distance between the two DNA segments in the crystal structure, and k is set to 20 $kcal/mol\ nm^2$ to produce an energy penalty small compared to the total energy of the system but sufficient to prevent overlapping DNA beads or cores.

Our chromatin model thus designates the DNA beads where TFs bind, defining binding regions, and this is used to specify a TF binding concentration.

We first studied how TF binding location directs chromatin folding architecture and how this depends on the linker DNA length. Second, we determined how the effect of increasing TF binding concentrations changes with chromatin internal parameters like linker DNA length, LH, and histone tail acetylation. Third we showed how TF binding can repress gene expression (Fig. 2D) and how this can be reversed by changing LH density.

We found that, when the length of linker DNA allows (44 to 80 bp), chromatin folding and architecture are dictated by the specific location of TF binding sites. As a result of protein binding, microdomains, or regions of high-frequency contacts in internucleosome contact maps, emerge. Depending on the LH density, extent of tail acetylation, and the length of the linker DNA, the concentration-dependent effect of TF binding is different. While LH impairs the effect of TF binding because it competes for fiber compaction, tail acetylation reduces the repressive effect of TF binding on chromatin architecture due to its fiber-opening tendency. Regarding linker DNA, short linkers produce geometrical restrictions that limit the effect of TF binding, whereas longer linkers provide the fiber with the flexibility required for TF regulation. When fibers have instead a mixture of linker DNA lengths, structural heterogeneity emerges, and TF binding effects are impaired.

The TF binding modeling also helped suggest a mechanism for the Eed gene locus repression (Fig. 2D). By simulating chromatin fibers representative of Eed with Myc:Max bound at two locations based on experimental information (Chip-seq data), we determined that Myc:Max can repress gene expression by occluding the transcription start site (TSS) of the gene. Interestingly, we also found that such effect can be reversed by increasing the LH density; increased fiber stiffness by LH binding reduces TF capability of bending the DNA and occluding the TSS. Thus, we elucidated the Myc:Max and LH-dependent mechanisms of Eed repression and activation.

Overall, our results on TF binding at the all-atom and implicit levels demonstrate that different resolutions can reveal various mechanistic aspects of genome regulation by TF binding. Protein binding to small DNA segments or chromatin fiber arrays shows how DNA architecture and folding can be modulated to increase or reduce its accessibility and compaction, which is directly related to genome expression. Each local variation and trend contributes significantly to the global organization of chromatin fibers and associated chromosomes. Surprisingly, transitions also can occur, as in LH binding to fiber arrays, where a sharp transition around a density of 0.5 occurs. Crowding effects in the cell intensify the competition among various chromatin elements, and all these complex interactions are modulated by post-translational modifications.

Summary and future machine learning prospects

Over the past 20 years, the integration of proteins into genome mesoscale models has created a rich toolkit with wide ranging applications (Figs. 1 and 2). Different models for proteins such as LH, antibodies, and TFs reveal intriguing mechanisms concerning LH-induced transitions (Yusufova et al. 2021; Portillo-Ledesma et al. 2022), LH regulation of metaphase chromatin via hierarchical loops (Grigoryev et al. 2016), CTCF single-residue mutation effects (Mao et al. 2023), creation of microdomains by TF binding (Portillo-Ledesma et al. 2024), or antibody competition and structure (Myers et al. 2020). In general, modeling protein-bound chromatin systems at the atomic, coarse-grained, mesoscale, and polymer levels (Huertas and Cojocaru 2021; Huertas et al. 2022; Portillo-Ledesma et al. 2023) have shed mechanistic insights on single nucleosomes, small fibers, gene systems, and chromosomes.

Although much progress has been made, further efforts are needed to better describe the dynamic nature of protein binding, crowding effects, competition among different proteins, and instantaneous changes in binding affinities.

For example, dynamic binding has been treated implicitly by our group for the LH binding (Collepardo-Guevara and Schlick 2011, 2012; Portillo-Ledesma et al. 2022), and explicitly for general proteins in polymer models (Brackley et al. 2013; Buckle et al. 2018). An approach similar to that we used for the antibodies, in which LHs can diffuse and bind to chromatin fibers and later unbind, combined with competition with other proteins and tunable binding affinities, could provide further insights into the role of LH variants with different affinity or on the role of post-translational modifications in the regulation of chromatin architecture and gene expression.

How the concentration of other proteins or salt concentration affects protein binding could be studied by considering crowding effects with explicit and dynamic protein binding. These crowding effects could be implicitly introduced by boundary conditions setting up constrained chromatin regions or by explicitly introducing chromatin arrays. The packing effect of multiple fibers at the nucleosome level

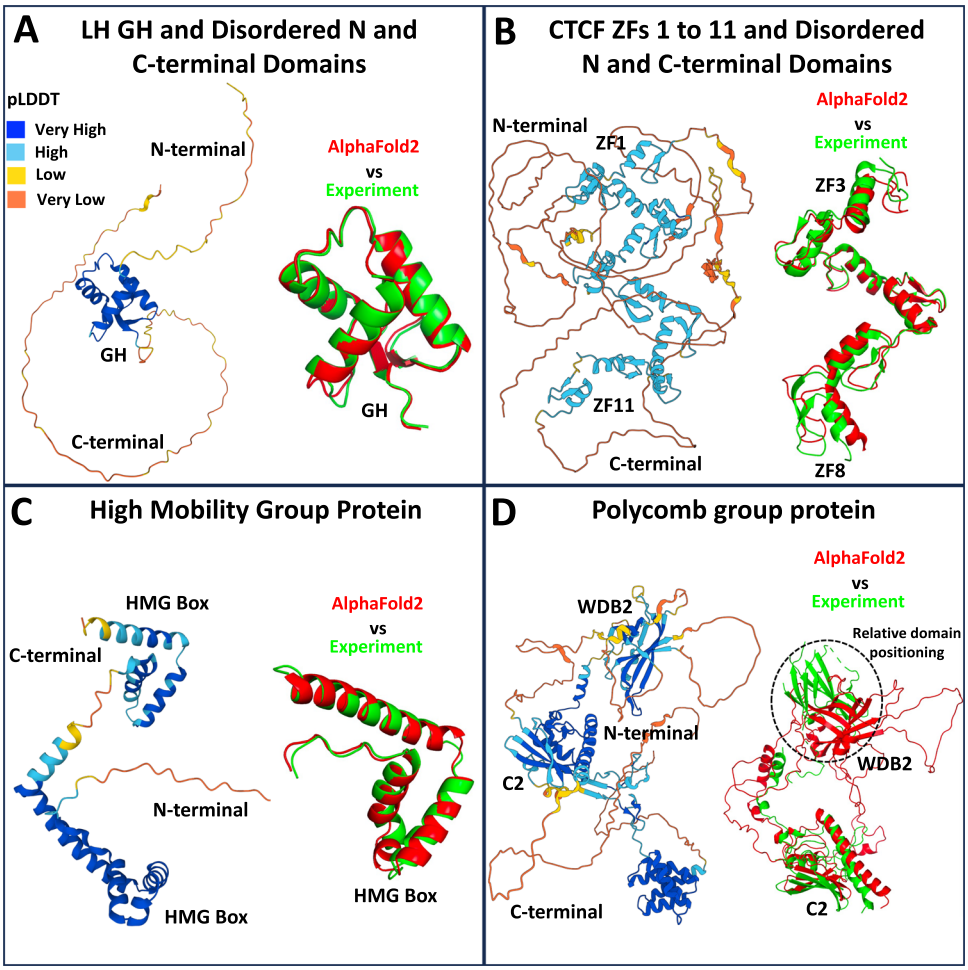

Fig. 3 Chromatin binding proteins predicted with AlphaFold2. Each panel shows the AlphaFold predicted structure colored by residue predicted local distance difference test (pLDDT value, see the end of this caption) (left) and the superposition of the predicted (red) with experimental (green) structures (right), showing for the predicted structure only the region that has been solved experimentally. **A** Human LH 1.2 globular head (GH) is predicted with very high accuracy but not the disordered N- and C-terminal domains. Superposition of AlphaFold structure with the experimental structure of the GH (PDBID: 8H0V) shows excellent alignment. **B** Human CTCF structure containing ZFs 1–11 is predicted with high accuracy. Disordered N- and C-terminal domains are predicted with very low accuracy. AlphaFold structure aligns well with the experimental structure containing ZFs 3–8 (PDBID: 5YEF). **C** Predicted structure of an HMG-domain containing protein (gene: M896-070310) from *Ordospora colligata* whose experimental structure is unknown has high and very high accuracy, except for the N-terminal region. The HMG domain aligns very well with the experimental structure of a human HMG protein (PDBID: 1QRV). **D** Predicted structure of a human polycomb group protein (gene: SUZ12) has high and very high accuracy, except for the disordered N- and C-terminal regions. Superposition of AlphaFold and experimental (PDBID: 6NQ3) structures shows failure in the relative positioning of one of the domains. AlphaFold2 uses as input the sequence of the target protein to generate a multiple sequence alignment (MSA), called the "MSA representation," and a list of structural templates from homologous proteins. In the Evoformer module, pair of residues in proximity are represented as edges of graphs, called the "pair representation." The MSA representation is used to update the graph edges based on geometrical proximity inference and to satisfy constraint conditions. At the same time, the pair representation is used to refine the MSA representation. This iterative process is performed 48 times. Thus, in each iteration, the model uses the current structural hypothesis to improve the assessment of the MSA, which in turns leads to a new structural hypothesis. Both sequence and structure information are exchanged until the network reaches the 48 cycles. Using the final pair and MSA representations, the 3D structure is constructed in the Structure module. In particular, every residue is represented with one free-floating rigid body for the backbone (a triangle) and χ angles for the side chains. Each backbone triangle is allowed to independently rotate and translate without any physical or geometrical restriction. The side chain positions are predicted based on parameterized torsion angles. The accuracy of each residue's prediction is performed by the pLDDT measure that evaluates local distances between all pair of atoms present in a radius R_0 belonging to different residues (Mariani et al. 2013). Thus, pLDDT measures the accuracy of the local environment surrounding each residue in the predicted structure. pLDDT is trained by X-ray data of high resolution (0.1–3 Å)

has been studied (Farr et al. 2021), finding that nucleosome breathing regulates liquid-liquid phase separation and multivalent nucleosome interactions. It remains to be seen how proteins modulate such complexes.

Finally, the system size that can be studied at high resolution remains a limitation. For the binding of LHs or TFs to genome systems in the Mb size range, new strategies are needed to incorporate these important local factors. In the spirit of QM/MM or MM/CG methods, methods that combine coarse-grained models at nucleosome resolution with polymer models at kb resolution could help study protein-bound whole genome systems.

Machine learning tools can assist in various ways by learning the patterns that dictate the folding at various levels and applying these combined effects to fold model systems. In Fig. 3, we analyze structures predicted with AlphaFold2 (Jumper et al. 2021) for illustrating the folding of various chromatin-associated units: small histones, CTCF, HMG proteins, or polycomb repressive proteins. AlphaFold2 performs a multiple sequence alignment (MSA) for the amino acid sequence of the target protein and evolutionary-related proteins and uses atomic coordinates of available homologous structures to predict with neural networks the 3D structure of the target protein (see details in Fig. 3 caption).

Overall, we see that structured domains for these four chosen systems are predicted with high or very high accuracy, but not the disordered domains, which are predicted with very low accuracy. This is evident in the N and C-terminal domains of the LH and CTCF proteins. Further, the relative alignment of domains is also not perfect. For example, in the polycomb group protein, one of the domains does not align with the experimental structure. However, AlphaFold2 can help predict entire structures that were partially solved experimentally. For example, the structure of CTCF containing all ZFs 1 to 11 is predicted with high accuracy, whereas experiments have solved only part of the ZFs.

While highly imperfect, this rapidly improving machine learning technology may be a starting point where no experimental data are available for initial model setups. Combined strategies of models and techniques, such as described here for protein-bound chromatin models at the mesoscale, will undoubtedly help provide an increasingly more complex and complete picture of how chromatin behaves in the natural cellular milieu and evolve us from genomes to cells to organisms.

Acknowledgements Support from the National Institutes of Health, National Institute of General Medical Sciences Award R35-GM122562, National Science Foundation Awards (DMS-215177 and DMS-2330628) from the Division of Mathematical Sciences, and Philip-Morris USA Inc to T.S. is gratefully acknowledged. S. P-L. is grateful for financial support from the Simons Center for Computational Physical Chemistry at NYU.

Author Contributions S. P-L. and T. S. wrote the review.

Data Availability Data sharing is not applicable to this review article as no new data were created.

Declarations

Conflict of interest The authors declare no competing interests.

References

Arya G, Schlick T (2009) A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments. J Phys Chem A 113:4045–4059. https://doi.org/10.1021/jp810375d

Azzaz AM, Vitalini MW, Thomas AS, Price JP, Blacketer MJ, Cryderman DE, Zirbel LN, Woodcock CL, Elcock AH, Wallrath LL, Shogren-Knaak MA (2014) Human heterochromatin protein 1α promotes nucleosome associations that drive chromatin condensation. J Biol Chem 289:6850–6861. https://doi.org/10.1074/jbc.M113.512137

Bajpai G, Jain I, Inamdar MM, Das D, Padinhateeri R (2017) Binding of DNA-bending non-histone proteins destabilizes regular 30-nm chromatin structure. PLOS Comput Biol 13:1–19. https://doi.org/10.1371/journal.pcbi.1005365

Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 410:120–124. https://doi.org/10.1038/35065138

Bascom GD, Schlick T (2017) Mesoscale modeling of chromatin fibers. In: Lavelle C, Victor J-M (eds) Nuclear Architecture and Dynamics, Academic Press, Boston, vol 2, pp 123–147. http://www.sciencedirect.com/science/article/pii/B9780128034804000053

Bascom GD, Sanbonmatsu KY, Schlick T (2016) Mesoscale modeling reveals hierarchical looping of chromatin fibers near gene regulatory elements. J Phys Chem B 120:8642–8653. https://doi.org/10.1021/acs.jpcb.6b03197

Bates DL, Thomas JO (1981) Histories H1 and H5: one or two molecules per nucleosome? Nucleic Acids Res 9:5883–5894. https://doi.org/10.1093/nar/9.22.5883

Bednar J, Garcia-Saez I, Boopathi R, Cutter AR, Papai G, Reymer A, Syed SH, Lone IN, Tonchev O, Crucifix C, Menoni H, Papin C, Skoufias DA, Kurumizaka H, Lavery R, Hamiche A, Hayes JJ, Schultz P, Angelov D, Petosa C, Dimitrov S (2017) Structure and dynamics of a 197 bp nucleosome in complex with linker histone H1. Mol Cell 66:384–397. https://doi.org/10.1016/j.molcel.2017.04.012

Bharath MMS, Chandra NR, Rao MRS (2003) Molecular modeling of the chromatosome particle. Nucleic Acids Res 31:4264–4274. https://doi.org/10.1093/nar/gkg481

Blackwood EM, Eisenman RN (1991) Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with myc. Science 251:1211–1217. https://www.science.org/doi/abs/10.1126/science.2006410

Brackley CA, Johnson J, Kelly S, Cook PR, Marenduzzo D (2016) Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains. Nucleic Acids Res 44:3503–3512. https://doi.org/10.1093/nar/gkw135

Brackley CA, Taylor S, Papantonis A, Cook PR, Marenduzzo D (2013) Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization. Proc Natl Acad Sci 110:E3605–E3611. https://doi.org/10.1073/pnas.1302950110

Brandani GB, Gopi S, Yamauchi M, Takada S (2022) Molecular dynamics simulations for the study of chromatin biology. Curr Opin Struct Biol 77:102485. https://www.sciencedirect.com/science/article/pii/S0959440X22001646

Buckle A, Brackley CA, Boyle S, Marenduzzo D, Gilbert N (2018) Polymer simulations of heteromorphic chromatin predict the 3D folding of complex genomic loci. Mol Cell 72:786-797.e11. https://doi.org/10.1016/j.molcel.2018.09.016

Calero-Rubio C, Saluja A, Roberts CJ (2016) Coarse-grained antibody models for "weak" protein-protein interactions from low to high concentrations. J Phys Chem B 120:6592–6605. https://doi.org/10.1021/acs.jpcb.6b04907

Calero-Rubio C, Ghosh R, Saluja A, Roberts CJ (2018) Predicting protein-protein interactions of concentrated antibody solutions using dilute solution data and coarse-grained molecular models. J Pharm Sci 107:1269–1281. https://www.sciencedirect.com/science/article/pii/S0022354917308857

Collepardo-Guevara R, Schlick T (2012) Crucial role of dynamic linker histone binding and divalent ions for DNA accessibility and gene regulation revealed by mesoscale modeling of oligonucleosomes. Nucleic Acids Res 40:8803–8817. https://doi.org/10.1093/nar/gks600

Collepardo-Guevara R, Schlick T (2011) The effect of linker histones nucleosome binding affinity on chromatin unfolding mechanisms. Biophys J 101:1670–1680. https://www.ncbi.nlm.nih.gov/pubmed/21961593. https://www.ncbi.nlm.nih.gov/pmc/PMC3183807/

Cutter AR, Hayes JJ (2017) Linker histones: novel insights into structure-specific recognition of the nucleosome. Biochem Cell Biol 95:171–178. https://doi.org/10.1139/bcb-2016-0097

Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V (2002) Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell 111:185–196. https://doi.org/10.1016/S0092-8674(02)00975-3

Elton TS, Nissen MS, Reeves R (1987) Specific A·T DNA sequence binding of RP-HPLC purified HMG-I. Biochem Biophys Res Commun 143:260–265. https://doi.org/10.1016/0006-291X(87)90659-0

Farr SE, Woods EJ, Joseph JA, Garaizar A, Collepardo-Guevara R (2021) Nucleosome plasticity is a critical element of chromatin liquid-liquid phase separation and multivalent nucleosome interactions. Nat Commun 12:2883. https://doi.org/10.1038/s41467-021-23090-3

Filippova GN, Qi C-F, Ulmer JE, Moore JM, Ward MD, Hu YJ, Loukinov DI, Pugacheva EM, Klenova EM, Grundy PE, Feinberg AP, Cleton-Jansen A-M, Moerland EW, Cornelisse CJ, Suzuki H, Komiya A, Lindblom A, Dorion-Bonnet F, Neiman PE, Morse HCr, Collins SJ, Lobanenkov VV (2002) Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter tts DNA-binding specificity. Cancer research 62:48–52

Fudenberg G, Abdennur N, Imakaev M, Goloborodko A, Mirny LA (2017) Emerging evidence of chromosome folding by loop extrusion. Cold Spring Harb Symp Quant Biol 82:45–55. https://doi.org/10.1101/sqb.2017.82.034710

Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA (2016) Formation of chromosomal domains by loop extrusion. Cell Rep 15:2038–2049. http://www.sciencedirect.com/science/article/pii/S2211124716305307

Fyodorov DV, Zhou BR, Skoultchi AI, Bai Y (2018) Emerging roles of linker histones in regulating chromatin structure and function. Nat Rev Mol Cell Biol 19:192–206. https://doi.org/10.1038/nrm.2017.94

Grigoryev SA, Arya G, Correll S, Woodcock CL, Schlick T (2009) Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. Proc Natl Acad Sci 106:13317–13322. http://www.pnas.org/content/106/32/13317

Grigoryev SA, Bascom G, Buckwalter JM, Schubert MB, Woodcock CL, Schlick T (2016) Hierarchical looping of zigzag nucleosome chains in metaphase chromosomes. Proc Natl Acad Sci USA 113:1238–1243. https://doi.org/10.1073/pnas.1518280113

Huertas J, MacCarthy CM, Schöler HR, Cojocaru V (2020) Nucleosomal DNA dynamics mediate Oct4 pioneer factor binding. Biophys J 118:2280–2296. https://doi.org/10.1016/j.bpj.2019.12.038

Huertas J, Cojocaru V (2021) Breaths, twists, and turns of atomistic nucleosomes. J Mol Biol 433:166744. https://www.sciencedirect.com/science/article/pii/S0022283620306690

Huertas J, Woods EJ, Collepardo-Guevara R (2022) Multiscale modelling of chromatin organisation: resolving nucleosomes at near-atomistic resolution inside genes. Curr Opin Cell Biol 75:102067. https://www.sciencedirect.com/science/article/pii/S0955067422000151

Izadi S, Anandakrishnan R, Onufriev AV (2016) Implicit solvent model for million-atom atomistic simulations: insights into the organization of 30-nm chromatin fiber. J Chem Theory Comput 12:5946–5959. https://doi.org/10.1021/acs.jctc.6b00712

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2

Kepper N, Foethke D, Stehr R, Wedemann G, Rippe K (2008) Nucleosome geometry and internucleosomal interactions control the chromatin fiber conformation. Biophys J 95:3692–3705. https://doi.org/10.1529/biophysj.107.121079

Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D (2002) Histone methyltransferase activity associated with a human multiprotein complex containing the enhancer of zeste protein. Genes Dev 16:2893–2905. https://doi.org/10.1101/gad.1035902

Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature 410:116–120. https://doi.org/10.1038/35065132

Le Gallo M, O'Hara AJ, Rudd ML, Urick ME, Hansen NF, O'Neil NJ, Price JC, Zhang S, England BM, Godwin AK, Sgroi DC, Program NIHISCNCS, Hieter P, Mullikin JC, Merino MJ, Bell DW (2012) Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. Nat Genet 44:1310–1315. https://pubmed.ncbi.nlm.nih.gov/23104009. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3515204/

Lequieu J, Córdoba A, Moller J, De Pablo JJ (2019) 1CPN: a coarse-grained multi-scale model of chromatin. J Chem Phys 150. https://doi.org/10.1063/1.5092976

Li Z, Portillo-Ledesma S, Schlick T (2023) Brownian dynamics simulations of mesoscale chromatin fibers. Biophys J 122:2884–2897. https://doi.org/10.1016/j.bpj.2022.09.013

Lin X, Zhang B (2024) Explicit ion modeling predicts physicochemical interactions for chromatin organization. eLife 12:RP90073. https://doi.org/10.7554/eLife.90073

Lobbia VR, Trueba Sanchez MC, van Ingen H (2021) Beyond the nucleosome: nucleosome-protein interactions and higher order chromatin structure. J Mol Biol 433:166827. https://www.sciencedirect.com/science/article/pii/S0022283621000218

Luque A, Collepardo-Guevara R, Grigoryev S, Schlick T (2014) Dynamic condensation of linker histone C-terminal domain reg-

ulates chromatin structure. Nucleic Acids Res 42:7553–7560. https://doi.org/10.1093/nar/gku491

Luque A, Ozer G, Schlick T (2016) Correlation among DNA linker length, linker histone concentration, and histone tails in chromatin. Biophys J 110:2309–2319. https://doi.org/10.1016/j.bpj.2016.04.024

MacCarthy CM, Huertas J, Ortmeier C, vom Bruch H, Tan D, Reinke D, Sander A, Bergbrede T, Jauch R, Schöler H, Cojocaru V, (2022) OCT4 interprets and enhances nucleosome flexibility. Nucleic Acids Res 50:10311–10327. https://doi.org/10.1093/nar/gkac755

Mao A, Chen C, Portillo-Ledesma S, Schlick T (2023) Effect of single-residue mutations on CTCF binding to DNA: insights from molecular dynamics simulations. Int J Mol Sci 24:6395. https://doi.org/10.3390/ijms24076395

Mariani V, Biasini M, Barbato A, Schwede T (2013) lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics (Oxford, England) 29:2722–2728. https://doi.org/10.1093/bioinformatics/btt473

Marshall AD, Bailey CG, Rasko JE (2014) CTCF and BORIS in genome regulation and cancer. Curr Opin Genet Dev 24:8–15. https://www.sciencedirect.com/science/article/pii/S0959437X13001482

Müller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA (2002) Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell 111:197–208. https://doi.org/10.1016/S0092-8674(02)00976-5

Myers CG, Olins DE, Olins AL, Schlick T (2020) Mesoscale modeling of nucleosome-binding antibody PL2-6: mono-versus bivalent chromatin complexes. Biophys J 118:2066–2076. http://www.sciencedirect.com/science/article/pii/S0006349519307167

Nair SK, Burley SK (2003) X-ray structures of myc-max and mad-max recognizing DNA: molecular bases of regulation by proto-oncogenic transcription factors. Cell 112:193–205. https://www.sciencedirect.com/science/article/pii/S0092867402012849

Nicodemi M, Prisco A (2009) Thermodynamic pathways to genome spatial organization in the cell nucleus. Biophys J 96:2168–2177. https://www.sciencedirect.com/science/article/pii/S0006349509003750

Olins AL, Langhans M, Monestier M, Schlotterer A, Robinson DG, Viotti C, Zentgraf H, Zwerger M, Olins DE (2011) An epichromatin epitope: persistence in the cell cycle and conservation in evolution. Nucleus (Austin, Tex.) 2:47–60. https://doi.org/10.4161/nucl.2.1.13271

Öztürk MA, Cojocaru V, Wade RC (2018) Toward an ensemble view of chromatosome structure: a paradigm shift from one to many. Structure 26:1050–1057. http://www.sciencedirect.com/science/article/pii/S0969212618301734

Öztürk MA, Pachov GV, Wade RC, Cojocaru V (2016) Conformational selection and dynamic adaptation upon linker histone binding to the nucleosome. Nucleic Acids Res 44:6599–6613. https://doi.org/10.1093/nar/gkw514

Öztürk MA, De M, Cojocaru V, Wade RC (2020) Chromatosome structure and dynamics from molecular simulations. Annu Rev Phys Chem 71:101–119. https://doi.org/10.1146/annurev-physchem-071119-040043

Perisic O, Portillo-Ledesma S, Schlick T (2019) Sensitive effect of linker histone binding mode and subtype on chromatin condensation. Nucleic Acids Res 47:4948–4957. https://doi.org/10.1093/nar/gkz234

Perišić O, Collepardo-Guevara R, Schlick T (2010) Modeling studies of chromatin fiber structure as a function of DNA linker length. J Mol Biol 403:777–802. http://www.sciencedirect.com/science/article/pii/S0022283610008338

Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, De la Cruz CC, Otte AP, Panning B, Zhang Y (2003) Role of histone H3 lysine 27 methylation in X inactivation. Science 300:131–135. https://doi.org/10.1126/science.1084274

Portillo-Ledesma S, Wagley M, Schlick T (2022) Chromatin transitions triggered by LH density as epigenetic regulators of the genome. Nucleic Acids Res 50:10328–10342. https://doi.org/10.1093/nar/gkac757

Portillo-Ledesma S, Chung S, Hoffman J, Schlick T (2024) Regulation of chromatin architecture by transcription factor binding. eLife 12:RP91320. https://doi.org/10.7554/eLife.91320

Portillo-Ledesma S, Li Z, Schlick T (2023) Genome modeling: from chromatin fibers to genes. Curr Opin Struct Biol 78:102506. https://www.sciencedirect.com/science/article/pii/S0959440X22001853

Portillo-Ledesma S, Schlick T (2020) Bridging chromatin structure and function over a range of experimental spatial and temporal scales by molecular modeling. Wiley Interdisciplinary Reviews: Computational Molecular Science 10:wcms.1434. https://doi.org/10.1002/wcms.1434

Reeves R, Nissen MS (1990) The A·T-DNA-binding domain of mammalian high mobility group I chromosomal proteins: a novel peptide motif for recognizing DNA structure. J Biol Chem 265:8573–8582. https://doi.org/10.1016/S0021-9258(19)38926-4

Seredkina N, Van Der Vlag J, Berden J, Mortensen E, Rekvig OP (2013) Lupus nephritis: enigmas, conflicting models and an emerging concept. Mol Med (Cambridge, Mass.) 19:161–169. https://doi.org/10.2119/molmed.2013.00010

Song F, Chen P, Sun D, Wang M, Dong L, Liang D, Xu RM, Zhu P, Li G (2014) Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. Science 344:376–380. https://doi.org/10.1126/science9.312251413

Sridhar A, Orozco M, Collepardo-Guevara R (2020) Protein disorder-to-order transition enhances the nucleosome-binding affinity of H1. Nucleic Acids Res 48:5318–5331. https://doi.org/10.1093/nar/gkaa285

Sridhar A, Farr SE, Portella G, Schlick T, Orozco M, Collepardo-Guevara R (2020) Emergence of chromatin hierarchical loops from protein disorder and nucleosome asymmetry. Proc Natl Acad Sci U S A 117:7216–7224. https://www.pnas.org/content/early/2020/03/11/1910044117

Stehr R, Kepper N, Rippe K, Wedemann G (2008) The effect of internucleosomal interaction on folding of the chromatin fiber. Biophys J 95:3677–3691. http://www.sciencedirect.com/science/article/pii/S0006349508785114

Tan C, Takada S (2020) Nucleosome allostery in pioneer transcription factor binding. Proc Natl Acad Sci 117:20586–20596. https://doi.org/10.1073/pnas.2005500117

Voutsadakis IA (2018) Molecular lesions of insulator CTCF and its paralogue CTCFL (BORIS) in cancer: an analysis from published genomic studies. High-Throughput 7. https://www.mdpi.com/2571-5135/7/4/30

Walker CJ, Miranda MA, O'Hern MJ, McElroy JP, Coombes KR, Bundschuh R, Cohn DE, Mutch DG, Goodfellow PJ (2015) Patterns of CTCF and ZFHX3 mutation and associated outcomes in endometrial cancer. JNCI: Journal of the National Cancer Institute 107. https://doi.org/10.1093/jnci/djv249, djv249

Watanabe S, Mishima Y, Shimizu M, Suetake I, Takada S (2018) Interactions of HP1 bound to H3K9me3 dinucleosome by molecular simulations and biochemical assays. Biophys J 114:2336–2351. https://www.sciencedirect.com/science/article/pii/S0006349518303941

Weiner SJ, Kollman PA, Singh UC, Case DA, Ghio C, Alagona G, Profeta S, Weiner P (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 106:765–784. https://doi.org/10.1021/ja00315a051

White AE, Hieb AR, Luger K (2016) A quantitative investigation of linker histone interactions with nucleosomes and chromatin. Sci

Rep 6:19122. https://doi.org/10.1038/srep19122. https://www.nature.com/articles/srep19122#supplementary-information

Wong H, Victor J-M, Mozziconacci J (2007) An all-atom model of the chromatin fiber containing linker histones reveals a versatile structure tuned by the nucleosomal repeat length. PLOS ONE 2:1–8. https://doi.org/10.1371/journal.pone.0000877

Woodcock CL, Skoultchi AI, Fan Y (2006) Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. Chromosom Res 14:17–25. https://doi.org/10.1007/s10577-005-1024-3

Woods DC, Rodríguez-Ropero F, Wereszczynski J (2021) The dynamic influence of linker histone saturation within the poly-nucleosome array. J Mol Biol 433:166902. https://doi.org/10.1016/j.jmb.2021.166902

Woods DC, Wereszczynski J (2020) Elucidating the influence of linker histone variants on chromatosome dynamics and energetics. Nucleic Acids Res 48:3591–3604. https://doi.org/10.1093/nar/gkaa121

Wu H, Dalal Y, Papoian GA (2021) Binding dynamics of disordered linker histone h1 with a nucleosomal particle. J Mol Biol 433:166881. https://www.sciencedirect.com/science/article/pii/S0022283621000759, diving into Chromatin across Space and Time

Yin M, Wang J, Wang M, Li X, Zhang M, Wu Q, Wang Y (2017) Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. Cell Res 27:1365–1377. https://doi.org/10.1038/cr.2017.131

Yoshida K, Toki T, Okuno Y, Kanezaki R, Shiraishi Y, Sato-Otsubo A, Sanada M, Park M-j, Terui K, Suzuki H, Kon A, Nagata Y, Sato Y, Wang R, Shiba N, Chiba K, Tanaka H, Hama A, Muramatsu H, Hasegawa D, Nakamura K, Kanegane H, Tsukamoto K, Adachi S, Kawakami K, Kato K, Nishimura R, Izraeli S, Hayashi Y, Miyano S, Kojima S, Ito E, Ogawa S (2013) The landscape of somatic mutations in down syndrome-related myeloid disorders. Nat Genet 45:1293–1299. https://doi.org/10.1038/ng.2759

Yusufova N, Kloetgen A, Teater M, Osunsade A, Camarillo JM, Chin CR, Doane AS, Venters BJ, Portillo-Ledesma S, Conway J, Phillip JM, Elemento O, Scott DW, Béguelin W, Licht JD, Kelleher NL, Staudt LM, Skoultchi AI, Keogh MC, Apostolou E, Mason CE, Imielinski M, Schlick T, David Y, Tsirigos A, Allis CD, Soshnev AA, Cesarman E, Melnick AM (2021) Histone H1 loss drives lymphoma by disrupting 3D chromatin architecture. Nature 589:299–305. https://doi.org/10.1038/s41586-020-3017-y

Zhou B-R, Feng H, Kale S, Fox T, Khant H, de Val N, Ghirlando R, Panchenko AR, Bai Y (2021) Distinct structures and dynamics of chromatosomes with different human linker histone isoforms. Mol Cell 81:166–182.e6. https://www.sciencedirect.com/science/article/pii/S1097276520307723

Zhou BR, Feng H, Kato H, Dai L, Yang Y, Zhou Y, Bai Y (2013) Structural insights into the histone H1-nucleosome complex. Proc Natl Acad Sci U S A 110:19390–19395. https://doi.org/10.1073/pnas.1314905110

Zhou BR, Jiang J, Feng H, Ghirlando R, Xiao TS, Bai Y (2015) Structural mechanisms of nucleosome recognition by linker histones. Mol Cell 59:628–638. https://doi.org/10.1016/j.molcel.2015.06.025

Zhou BR, Feng H, Ghirlando R, Li S, Schwieters CD, Bai Y (2016) A small number of residues can determine if linker histones are bound on or off dyad in the chromatosome. J Mol Biol 428:3948–3959. https://doi.org/10.1016/j.jmb.2016.08.016