\title{Bottom-up meets top-down: the crossroads of multiscale chromatin modeling}

\runningtitle{Bottom-up meets top-down} %% For page header

\author[1]{Joshua Moller}

\author[1,2,*]{Juan J de Pablo}

\runningauthor{Joshua Moller and Juan J de Pablo} %% For page header

\affil[1]{Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, United States}

\affil[2]{Material Science Division, Argonne National Laboratory, Lemont, Illinois 60439, United States}

\corrauthor[*]{depablo@uchicago.edu}

% \papertype{Letters}

\papertype{Article}

% \papertype{Computational Tools}

\begin{document}

\begin{frontmatter}

\begin{abstract}

%\quickwordcount{main}

%\quickcharcount{main}

Chromatin can be viewed as a hierarchically structured fiber that regulates gene expression.

It consists of a complex network of DNA and proteins whose characteristic dynamical modes facilitate compaction and rearrangement in the cell nucleus.

These modes stem from chromatin's fundamental unit, the nucleosome, and their effects are propagated across length-scales.

Understanding the effects of nucleosome dynamics on the chromatin fiber is of central importance to epigenetics, primarily through post-translational modifications that occur on the histones.

Within the last decade, imaging and chromosome conformation capture techniques have revealed a number of structural and statistical features of the packaged chromatin fiber at a hitherto unavailable level of resolution.

Such experiments have led to increased efforts to develop polymer models that aim to reproduce, explain, and predict the contact probability scaling and density heterogeneity.

At nanometer scales, available models have focused on the role of the nucleosome and epigenetic marks on local chromatin structure.

At micrometer scales, existing models have sought to explain scaling laws and density heterogeneity. Less work, however, has been done to reconcile these two approaches - bottom-up and top-down models of chromatin.

In this perspective, we highlight the multi-scale simulation models that are driving towards an understanding of chromatin structure and function, from the nanometer to the micron scale, and we highlight areas of opportunity and some of the prospects for new frameworks that bridge these two scales.

Taken together, experimental and modeling advances over the last few years have established a robust platform for study of chromatin fiber structure and dynamics, which will be of considerable use to the chromatin community in developing an understanding of the interplay between epigenomic regulation and molecular structure.

\end{abstract}

\begin{sigstatement}

The hierarchical structure of chromatin has captured the attention of biologists, chemists, physicists, and engineers for several decades.

While progress has been made in understanding the human genome, the epigenome, and its influence on gene regulation, \textcolor{blue}{the structure of the genome in the nucleus} remain\textcolor{blue}{s} poorly understood.

That state of affairs is starting to change, largely as a result of advances in biology and imaging.

The resolution of emerging measurements has reached a level that perhaps necessitates interpretation through molecular modeling.

Here, we describe a multiscale modeling framework that allows one to bridge chromosomal length scales and could help augment emerging technologies.

In principle, one could quantify the relation between epigenetic modifications that take place on single nucleosomes and the resulting structure and function of the chromatin fiber.

\end{sigstatement}

\end{frontmatter}

% Introduce the hierarchical nature of the chromatin fiber

Genome packaging poses intriguing questions that are relevant not only to biology, but also to polymer physics and chemistry.

The human genome consists of billions of base pairs of DNA that are densely packed in the cell nucleus, well below the theoretical packing limit dictated by the persistence length of double-stranded DNA ($\sim \mathrm{50nm}$).

Such DNA packaging occurs over multiple length scales, eventually leading to a dense nuclear environment.

Despite the dense packaging, DNA must be rendered accessible at the gene-scale ($\sim\mathrm{kbp}$) for necessary processes such as replication \cite{Groth2007}, transcription \cite{Li2007} and DNA repair \cite{Adkins2013}.

The chromatin fiber provides an avenue for such functionality through its hierarchical and dynamic structure, overcoming strict DNA packaging constraints to facilitate organization and epigenomic regulation across length scales. \cite{Bowman2015,Kurdistani2003}

In particular, epigenomic regulation influences chromatin structure through covalent modifications \textcolor{blue}{of the fiber such as post-translational modifications (PTMs) and DNA methylation, and substitutions of structural proteins.} \sout{to the fiber, known as post-translational modifications (PTMs), and substitutions of structural proteins.}

% Introduce modeling philosophies here

There are well documented links between chromatin structure, gene regulation, and epigenetics, but less is known about how these phenomena directly influence one another.

Moreover, these features or processes are manifested at all relevant chromosomal length-scales, thereby requiring multiscale methodologies for their study.

As the structure of chromatin is inherently dynamic and falls within the ``dark region''\cite{Zhang2016} between conventional microscopy and x-ray crystallography, multiscale modeling is particularly useful in ``filling the gaps''.

Thus far, modeling efforts to describe chromatin have adopted two perspectives. Bottom-up models, including detailed descriptions of the nucleosomes, have sought to understand their influence on chromatin structure.

\sout{Top-down models, aim to explain chromatin conformation and deep sequencing experiments at coarse levels of resolution.}

Bottom-up approaches have been primarily focused on the role of the nucleosome in chromatin; how does the nucleosome facilitate regulation of the underlying DNA, and how does the epigenome, in the form of PTMs and histone variants, affect nucleosome structure and dynamics.

Bottom-up methods make use of atomistic and coarse-grained molecular models.

Top-down approaches rely on chromosome conformation capture and immunoprecipitation techniques, such as Hi-C \cite{Lieberman-Aiden2009}, ChIA-PET \cite{Li2010}, ChIP-seq \cite{Johnson2007}, \textblue{and Micro-C} \cite{Hsieh2015,Hsieh2016} to understand the 3D-genome - the physical organization of the genome in the nucleus.

Such methods rely on polymer physics and statistical approaches to develop models that replicate specific behavior.

% The summary of the perspective

Bottom-up models of increasing sophistication have been able to capture nucleosome physics up to 10-100 kbp, and top-down methods parameterized with vast amounts of experimental data are broaching the 3D-genome at a <1 kbp resolution.

At this point in time, chromatin models are reaching a ``convergence'' of sorts, where top-down and bottom-up methodologies can both resolve the intermediate length scales that hold many of the secrets that encode epigenetic information. Despite the wide disparity between these modeling philosophies, there is now a crossover and tremendous potential for cross-fertilization. At intermediate lengths scales, these two types of models are stretched to their limits; by combining the two, it should be possible to arrive at more complete and content-rich representations of chromatin. (Figure \ref{fig:modeling})

A complete tool-set is needed to address the following questions: What physics from the nucleosome scale influence the 3D genome? what physics do not? and how can they be manipulated?

In this perspective, we discuss recent modeling advances, the potential to bridge length and time scales, and, ultimately, the impact that forming such bridges will have in chromatin research.

% Intro to bottom-up modeling of chromatin

\section*{Bottom-up Design Philosophy}

Bottom-up models \sout{offer a low potential for} \textcolor{blue}{require significant effort to achieve} representability, but can achieve a high level of transferability.

For example, if a nucleosome is the system of choice, a bottom-up model that is able to reproduce the effects of histone modification on DNA affinity will be capable of reproducing how nucleosomes interact and the thermodynamics that emerge from those interactions in a wide variety of contexts.

\sout{The opposite is true in top-down models, where high representability is accompanied by low transferability. }

\sout{A model that is capable of representing Hi-C data for one particular genome is unlikely to describe Hi-C data for a different genome or organism.}

\sout{Bottom-up approaches start from detailed molecular principles, and systematic coarse-graining provides the means to scale them up.}

Here, we briefly highlight various modeling and experimental efforts that, when coupled, can take chromatin starting from the atomistic to the mesoscale, and we discuss several relevant questions at each length scale.

\subsection*{Atomistic simulations: nucleosome structure and the epigenome}

% X-ray crystallography drove atomistic modeling of the nucleosome

Experiments that rely on x-ray crystallography have provided an atomistically detailed view of the nucleosome, in which a stable complex is formed between an octamer of histones wrapped ~1.7 times by a 147 bp left-handed DNA superhelix.\cite{Luger1997,Davey2002}

That view led directly to the first fully atomistic simulations of the nucleosome. \cite{Bishop2005}

Initial work at this scale focused on nucleosome stability and dynamics. \cite{Eslami-Mossallam2015}

More recent, ambitious efforts have considered nucleosome structure and function influenced by nucleosome-binding proteins \cite{Ozturk2018}, histone variants \cite{Bowerman2016, Melters}, or PTMs \cite{Li2016,Zhang2017a}.

These efforts are starting to elucidate major pieces of the role of the nucleosome in the epigenetic landscape.

For example, Bowerman et al. revealed that H2A histone variants alter nucleosome dynamics, while Melters et al. explained that the enhanced elasticity of Cenp-A nuclesomes is related to enhanced levels of transcription. \cite{Bowerman2016, Melters}

% The prospect of atomistic simulations

Unfortunately, encompassing every permutation of histone modifications, histone variants, and nucleosome-binding protein with atomistic resolution is beyond currently available computational resources.

\textcolor{blue}{And while bottom-up methods have a higher probability of transferability, it is not necessarily guaranteed.

Atomistic methods and force-fields tend to not be as transferable as theory would often dictate.

The onset of machine learning for force-field development could increase the transferability of atomistic simulations, encompassing histone modifications with higher fidelity.}

More importantly, the question of how histone interactions translate into chromatin changes at longer scales is not accessible at this level of resolution,\cite{Jung2019} and careful coarse-graining is therefore necessary to address it.

% PTMs influence nucleosome structure and dynamics

%The vast number of potential PTMs and their combinations comprise what is referred to as the ``histone code.''\cite{Jenuwein2001}

%Of the possible PTMs, the most prevalent ones include acetylation, methylation, ubiquitination, and phosphorylation.\cite{Felsenfeld2003}

%Atomistic simulations of nucleosomes are chipping at the histone code by linking PTMs to changes in nucleosome function.\cite{Li2016,Bowerman2016}

%Each mark induces a change in the nucleosome structure, dynamics, or binding affinity of proteins to locally tune functionality. \cite{Bowman2015,Brownell1996}

%For example, acetylation of the H4 tail at lysine 16 (H4K16ac) is associated with active transcription, while trimethylation of the H3 tails at lysine 9 (H3K9me3) is associated with repressed chromatin, or constitutive heterochromatin at the supranucleosomal scale. \cite{VanHolde1989,Strahl2000}

%While it is simply impossible to consider every PTM at this scale, it is reassuring that a large number of them affect inter-nucleosome interactions moreso than intranucleosome interactions, which reduce this number substantially. \cite{Eslami-Mossallam2015}

\begin{figure}[hbt!]

\centering

\includegraphics[width=0.95\linewidth]{figs/fig_modeling.eps}

\caption{Current modeling approaches to chromatin.

From the left, bottom-up modeling approaches begin with all-atom resolution with each bead around the scale of 1-2 \angstrom.

The coarse-grained scale representation here is that of the three site per nucleotide (3SPN) DNA model and the atomic interaction-based coarse-grained (AICG) protein model. \cite{Freeman2014a,Li2011}

Particles at this coarse-grained scale are in the range of 3-6 \angstrom.

The colors of the proteins here represent the net charge of the residue - white for no charge, blue for +1 charge, and red for -1 charge.

The mesoscopic model is the 1 cylinder per nucleosome (1CPN) model, which represents each nucleosome as a cylinder and DNA as a sphere on the scale of 1 nm. \cite{Lequieu2019}

Here, the 1CPN configuration is an oligomer with linker histones bound to the nucleosome.

From the right, top down models use coarse polymer representations of chromatin.

The chromosomal model is the minimal chromatin model (MICHROM) at a 50 kbp per bead resolution. \cite{DiPierro2016}.

The supranucleosomal model is the mean-field model representing chromatin at a nucleosome per bead resolution. \cite{MacPherson2018}

This image has been published with approval of PNAS.

These models converge at intermediate, mesoscopic length scales, and offer nucleosome-scale resolution and capture the relevant, underlying physics.

}

\label{fig:modeling}

\end{figure}

\subsection*{Coarse-grained models: nucleosome thermodynamics}

% Why coarse-grain and what is it good for?

Coarse-grained nucleosome models feasibly resolve longer time- and lengthscale nucleosome dynamics, while drawing information from the atomistic scale.

Through coarse-graining, nucleosome models can be better compared to experimental data, but extensive validation is needed to establish their limits of applicability.

As a result, representability at this length scale often becomes a time-consuming endeavor.

Nevertheless, significant progress in chromatin research has been performed at the coarse-grained nucleosome scale.

Recent efforts have resolved several key questions raised by experiments. \cite{Freeman2014a,Eslami-Mossallam2015,Zhang2016} Freeman et al., for example, explained that the affinity of particular sequences of DNA to the histone core is encoded in the curvature of such sequences \cite{Freeman2014a}.

Such models have thermodynamically assessed experimental force-induced unwrapping \cite{Lequieu2016,Zhang2016}, sequence-dependent nucleosome stability \cite{Freeman2014a,Eslami-Mossallam2015}, and nucleosome sliding \cite{Lequieu2017,Brandani2018}.

% Internucleosome energetics was the next step, what is now?

While single-nucleosome studies have been the primary focus of atomistic and coarse-grained models, multi-nucleosome \textcolor{blue}{are the subject of coarse-grained models.}

With multiple nucleosomes, additional information is necessary.

Inter-nucleosome interactions, and the length of linker DNA, must also be considered - both are known to greatly influence the structure of chromatin. \cite{Routh2008,Song2014}

The inter-nucleosome interaction has been difficult to quantify.

Experiments pulling isolated chromatin fibers have attempted to extract the energy of association between nucleosomes within that context.\cite{Meng2015a}

Three significant studies have reported values of nucleosome interaction energies, with variable results.\cite{Kruithof2009,Cui2000,Funke2016}

Support for these results in the form of atomistic simulations cite the histone tails as the primary contributor for these interactions.\cite{Zhang2017a}

In particular, the positively-charged residues of the H4 tail are reported to attract nearby nucleosomes through interactions with an acidic patch located on the H2A histone of the nucleosome core.\cite{Zhang2017a}

Acetylation of these residues disrupts these interactions and, to some degree, lowers the attractive interactions between nucleosomes.\cite{Collepardo-Guevara2015,Potoyan2012,Winogradoff2015}

This phenomenon serves to underscore that epigenetic modifications directly influence structural changes in chromatin.

One could conclude that the ubiquitous H4K16ac modification is associated with transcription because the chromatin becomes unraveled after this change in attractions between nucleosomes. \cite{Vettese-Dadey2018,Brownell1996}

The picture outlined above has now been put on firmer grounds through extensive simulations with coarse-grained models, where inter-nucleosome interactions where derived from more detailed representations, leading to excellent agreement with experiments with and without H4K16ac.\cite{Moller}

Importantly, such models provide a generalized description of nucleosome interactions as a function of both relative orientation and separation distance. Such a description is invaluable from a chromatin modeling perspective, as interactions within the chromatin fiber have thus far been poorly characterized.

% Maybe this is too much

% Recent work at this scale

With such detailed characterization of nucleosome physics now available, current modeling efforts have taken different directions.

Drawing inspiration from recent atomistic studies, one such direction has sought to investigate the influence of histone variants, nucleosome-binding proteins, etc. on nucleosome thermodynamics and structure.

Another direction has sought to use the information now available at this scale to build models capable of describing longer molecules and scales.

Examples are provided by the work of Watanabe et al., who have investigated small complexes with these models, such as dinucleosomes with the HP1 protein and trinucleosomes, to represent their effect on chromatin structure. \cite{Chang2016, Watanabe2018}

Due to limitations in sampling, however, such efforts can currently provide reliable descriptions of one or a few nucleosomes; scaling up to even a few kbps requires that a different, more coarse-grained class of models be developed.

% 1CPN model

\subsection*{Nucleosome-based coarse-graining: mesoscale modeling}

% Experiments at this scale

Within the constellation of nucleosome-based coarse-grained models, the ``mesoscale'' currently represents the largest scale.

At this scale, current models can describe oligomers of nucleosomes connected by linker DNA.

A prominent example of the questions that arise at such length scales is provided by the structure of nanoscopic chromatin fibers or aggregates. Early experiments with isolated chromatin oligomers \textit{in vitro} originally proposed the now controversial concept of a 30-nm fiber. \cite{Olins1974,Robinson2006,Song2014,Collepardo-Guevara2015}

While recent \textit{in vivo} imaging has contradicted the notion of a static 30-nm chromatin fiber, \textit{in vitro} evidence is strong. \cite{Maeshima2010,Ou2017,Tremethick2007}

To ultimately understand how epigenetic phenomena influence the structure of chromatin at this level, the structure of unmodified chromatin and what drives its formation must be assessed.

Put simply, the basis for chromatin fiber structure derives from DNA deformation penalties and the mechanical modes that limit them.

Generally, these are balanced by energetic processes that are favorable.

These include inter-nucleosome contacts, nucleosome positioning energy, and DNA-protein interactions.

%Such favorable interactions.

This balance is modulated by the amount of linker DNA connecting each nucleosome.

From experiments and simulation, it is found that linker DNA lengths that are in integer amounts of DNA pitch ($\sim$10 bp) lead to compact structures, whereas deviations frustrate the fiber, leading to larger structures.

This property influences the eukaryotic genome, and high-throughput sequencing of linker DNA has revealed peaks of intensity at integer lengths of DNA pitch.\cite{Brogaard2012,Widom1992,Chereji2018}

% The push towards larger structures

In general, models at this scale coarse-grain DNA as a few bp per bead, and the nucleosome is represented as a single entity.

The DNA is treated as a wormlike-chain (WLC) model, \cite{Marko1995}

and different variants treat the nucleosome and inter-nucleosome interactions in their own way.

Pioneering work by Arya et al. introduced a coarse-grained scheme that reproduces the electronic field of the nucleosome with a reduced representation, and it has been successful in predicting chromatin fiber structure. \cite{Arya2006,Arya}

This method, known as the discrete surface charge optimization (DISCO), reduces the nucleosome to ~300 \textcolor{blue}{pseudo-charges distributed on the nucleosome surface}\sout{charged particles}.

Building on that work, in recent years, additional modeling approaches at this length scale have emerged.

A different approach to nucleosome coarse-graining draws from experimental values of nucleosome-interaction strengths. \cite{Norouzi2015}

\textcolor{blue}{Additional efforts have developed a nucleosome-resolution model with specified topological interactions to inform the assembly of nucleosomes in the chromatin fiber. \cite{Brackley2017}}

Recently, we have introduced the 1-cylinder-per-nucleosome (1CPN) model, built off of previous nucleosome-centric results and other, well-founded models at this length scale.\cite{Lequieu2019,Wedemann2002}

\textcolor{blue}{Another model at this scale incorporates non-histone proteins which bend DNA, resulting in heterogeneous packing in oligonucleosome fibers}\cite{Bajpai2019}

\sout{Validation for these models has come in the form of comparison to experimental sedimentation coefficients and comparison to fiber-pulling experiments. }\cite{Hansen1989,Meng2015a,Grigoryev}

% The features of these models

In addition to unaltered chromatin, these models incorporate some of the more well-characterized epigenetic phenomena.

For example, the work started by \sout{Luque et al}\textcolor{blue}{Arya et al}, incorporated the H1\sout{.4 rat-variant} linker histone, which was further extended to incorporate \textcolor{blue}{better resolution and} more variants in recent work\sout{ by Perisic et al}.\cite{Arya,Luque2014,Perisic2019}.

These efforts influenced the linker histone model described in the 1CPN model of Lequieu et al.\cite{Lequieu2019}

\sout{Note that the H4K16ac mark has been incorporated as well.}

Recent work by Bascom et al. has pushed modeling efforts at this scale to understand how packaging of the HOXC gene hub is influenced by the amount of linker histones, H4 acetylation, and variable linker lengths. \cite{Bascom2019}

% The most recent feature

\sout{Although the 1CPN model is relatively new, it has already been used to provide a thorough thermodynamic understanding of the dinucleosome complex. The dinucleosomes in that work display structural features similar to those seen in the 30-nm fiber, and it is therefore of interest to note that the 1CPN model does not predict the 30-nm fiber to be a stable structure.

In fact, larger fibers display motifs of fibers that are fluid in radius, much like the experimental ChromEMT results.\cite{Ou2017}

These dinucleosome results, as well as other efforts, indicate that the fiber structure is highly sensitive to linker DNA length.\cite{Lequieu2019,Koslover2010,Dobrovolskaia2010,Nam2016,Routh2008}}

% The current push

Models at the mesoscale are \textcolor{blue}{extending their capabilities to larger length-scales as a part of the larger effort to} \sout{pushing the limits understanding for} \textcolor{blue}{understand} how nucleosomes influence the supranucleosomal scale.

This work is only beginning to knock at the door of higher order structures, such as chromatin topologically associating domains (TADs), compartments, and territories.

While the recent \textit{in silico} advancements are exciting, larger phenomena that occur with 100 kbp - 1 Mbp structures are still out of reach.

The multi-scale modeling approach highlighted above demonstrates the importance of nucleosome physics on the overall structure of chromatin and, moving forward, it will be important to further characterize the effects of individual linker lengths and linker histones on chromatin structure and their impact on chromosome packaging.

Note that several experimental studies have provided information about the distribution of linker lengths \textit{in vivo}\cite{Brogaard2012,Voong2016,Chereji2018}, \sout{and it should now be possible to incorporate that information into ongoing \textit{in silico} efforts.} \textcolor{blue}{which have only recently been integrated into \textit{in silico} efforts.} \cite{Bascom2017,Bascom2019}

\textcolor{blue}{An additional key study for nucleosome positioning is Micro-C, which has been integrated in the model introduced by Wiese et al to predict the structure of chromatin with experimental heterogeneous linker spacing.}\cite{Wiese2019}

To capture larger scale chromatin packaging, top-down modeling approaches are still necessary.

\section*{Top-down Models}

% Hi-C runs the gambit

\textcolor{blue}{Top-down models have aimed to explain chromatin conformation and deep sequencing experiments at coarse levels of resolution.

In general, top-down models offer an easy avenue for representability which is accompanied by low chance of transferability.}

As mentioned earlier, a vast amount of information pertaining to the 3D genome has emerged from high-throughput sequencing and chromosome conformation capture methods (3C, 5C, Hi-C).\cite{Dixon2012,Rao2014,Rao2017,Sanborn2015}

In particular, Hi-C uses high-throughput sequencing to measure the contact probability of genomic segments as a function of genomic distance. \cite{Lieberman-Aiden2009}

These methods are helping elucidate some of the higher length-scale organization of the genome.

Processes such as chromatin looping, topologically associating domains (TADs), and chromatin compartmentalization have been gradually elucidated by relying on Hi-C studies.\cite{Dixon2012,Schwarzer2017,Rao2017}

The advent of these methods provides a direct link between the epigenome and the structural organization of chromatin.

This is further underscored by the fact that chromatin compartments strongly correlate with associated epigenetic marks.\cite{Rao2017}

Recently, single cell Hi-C and novel sequential fluorescence \textit{in situ} hybridization have revealed that population average features, such as TADs, are in fact present in single nuclei,\cite{Bintu2018} thereby providing an additional incentive to rely on modeling approaches that resolve the 3D genome.

% Higher order problems

These chromatin structures play vital roles in transcription and repair as well as cell development.

Disruption of TAD boundaries is linked to genetic diseases, including cancer. \cite{Hnisz2016}

Arising from these higher order structures is a heterogeneous DNA packaging density.

Current label-free imaging technologies, like partial-wave spectroscopy (PWS), have quantified the density heterogeneity of nuclear chromatin.\cite{Dong2016,Almassalha2017}

By capturing fluctuations in DNA density with up to 20-nm resolution, these methods also reveal that the degree of heterogeneity differentiates cancerous from healthy cells, and help connect physical changes in chromatin to genetic diseases at small, molecular scales.\cite{Almassalha2016}

\begin{figure}[hbt!]

\centering

\includegraphics[width=0.9\linewidth]{figs/fig_future_modeling_1.eps}

\caption{ This figure illustrates some of the prospects for chromatin models.

Ideally, modeling could incorporate single epigenetic marks, shown here as a red and blue mark to a histone tail.

Such marks influence the chromatin fiber structure and can be predicted by modeling efforts.

Here, the blue mark results in a condensed fiber and the red in an extended fiber.

The subsequent contact maps are provided for these schematic fibers to highlight the comparison to top-down methods, and the potential to be compared to Hi-C data.

}

\label{fig:future}

\end{figure}

\subsection*{Polymer modeling of chromosomes}

% Polymer models at this scale

The resolution of Hi-C data is relatively coarse, and polymer models of chromatin, informed by these experiments, have been useful in interpreting such data. \cite{Rosa2008}

Initial work attempted to use established polymer physics to describe chromatin packaging.

However, this framework was abandoned, as no ideal polymer model is able to predict the anomalous contact scaling probability of chromatin as a function of genomic separation. \cite{Mirny2011}

From Hi-C, this corresponding exponent is in the range of of 1.1 - 1.33, and falls in between the scaling of an ideal chain (1.5) and that of a poor-solvent chain (1.0). \cite{Lieberman-Aiden2009,Rao2014,Huang2018}

Accounting for this anomaly, polymer models incorporate chromatin looping in the form of \textcolor{blue}{strings and binders switch model} \sout{switches-and-binders}\cite{Barbieri2012}, slip-springs\cite{Brackley2017}, or active extrusion.\cite{Fiorillo2019}

The work of Fudenberg et al. has been important in describing the active extrusion process and its influence on the genome. \cite{Fudenberg2016,Nuebler2017}

% Color-based models (Vaillant, Micheletti, Orlandini etc.)

\textcolor{blue}{As chromatin compartments depending on epigenetic mark have been uncovered through Hi-C, they have been the focus of some modeling approaches.

One such approach is a static definition of epigenetic state, where monomers of a polymer are ``colored'' according to their state, similar to block co-polymer models.

The model of Jost et al. incorporates four such colors, representing different epigenetic states to predict the structure of \textit{Drosophila} chromsomes from relevant Hi-C. \cite{Jost2014}

However, epigenetic state is capable of ``spreading'' through a feedback loop mechanism.

The block co-polymer type model has also been extended to incorporate a change of epigenetic state through a ``two-state'' kinetic model between active and inactive states of chromatin.

This model of Michieletto et al. is able to recapitulate the stability of epigenetic domains subject to perturbations, thereby providing some insight into the presence of TADs. \cite{Michieletto2016}

These modeling approaches are seminal in providing insight into the dynamic interplay between epigenomic regulation and chromatin structure at a coarse scale.}

% Maximum entropy approaches

A different approach to unraveling the structure of chromatin has come from \sout{the} maximum-entropy \textcolor{blue}{minimization between models and experimental data} \sout{model of di Pierro et al.} \cite{DiPierro2016},

which can reproduce \sout{Hi-C} \textcolor{blue}{chromosome conformation} contact maps and \sout{is} \textcolor{blue}{are} able to describe the underlying structures of chromatin.

\textcolor{blue}{One such approach of Georgetti et al. informed the structure of the mouse X-inactivation center region with a simple polymer model from 5C data with high fidelity.\cite{Giorgetti2014}

The maximum entropy model was later extended to cover entire chromsomes through Hi-C data in the minimal chromatin (MICHROM) model of di Pierro et al.\cite{DiPierro2016}}

Expanding upon this work, di Pierro et al drew information from ChIP-seq data, and through machine-learning methods, discovered that the same Hi-C contact maps could be attained.

This leads to the surprising conclusion that the structure of chromatin is encoded in the epigenetic marks. \cite{DiPierro2017}

This model has also been physically characterized and demonstrates varying chromatin diffusion rates of chromosome compartments. \cite{Pierro2018}

% Bystricky lab and Amitai review

\textcolor{blue}{In addition to incorporating Hi-C data, the prevalence of single-particle tracking through optical microscopy leads to easy quantification of chromatin diffusion and the tendency to form loops between distal elements.

Rouse polymer models of chromatin that incorporate such quantities have been especially helpful in elucidating chromatin packaging and dynamics. \cite{Amitai2017}

Recent work of Socol et al. demonstrates that Rouse models require extension in the form of transient chromatin contacts, known as the RouseTIC model, to predict both single-particle tracking of \textit{in vitro} chromatin and Hi-C of the yeast genome. \cite{Socol2019}

Rouse models and looping implementations have been the subject of multiple reviews that are recommended for further interest. \cite{Amitai2017,Tiana2018,LeDily2017,Fiorillo2019}}

% Arbona yeast simulations using Bayesian inference

\textcolor{blue}{A key setback of current top-down approaches is a lack of transferability.

A recent novel approach to budding yeast nucleus simulations by Arbona et al uses Bayesian inference to learn a fixed class of polymer and microtubule parameters using Hi-C, imaging, simulations and Micro-C-XL data. \cite{Arbona2017}

We find approaches that link experiments and simulations through data-driven methods an exciting avenue to determine fundamental physical parameters.

Approaches that draw from multiple experiments like that of Arbona et al and Socol et al could even reconcile the differences in single nucleus versus population average effects that have been the source of discrepancies in results. \cite{LeDily2017}

}

\section*{Where bottom-up meets top-down}

While prior modeling has laid the groundwork for understanding the physics that underlie each length-scale, we now come back to the initial question of how does the nucleosome influence chromosomal organization.

In order to incorporate higher order effects and preserve a nucleosome-first approach, additional features must be included in available models.

The outlook is optimistic, with advances from both approaches reaching similar scales.

Recently, MacPherson et al. introduced a bottom-up mean-field chromatin model that is also discretized at the nucleosome level.\cite{MacPherson2018,Macpherson2019}

This model incorporates chromatin compartmentalization through a Hamiltonian that accounts for constitutive heterochromatin formation.

We find this model to be a promising candidate for bridging the gap between the two scales.

Additionally, the recent self-returning random walk model, at a similar level of resolution, has been promising at capturing both the packaging heterogeneity predicted by PWS and ChromSTEM with the contact probability scaling of Hi-C. \cite{Huang2018}

With advances at the mesoscale, chromatin can be studied in considerable detail, with resolution ranging from nanometers and the histone level, to the gene or kbp scales of DNA.

With careful consideration, the nucleosome scale physics from mesoscale models could be incorporated in these models.

Similarly, the addition of the maximum entropy methodology of di Pierro et al. of mapping ChIP-seq or Hi-C to such a model provides an exciting prospect. (Figure \ref{fig:future})

% Small fibers and their importance

Another promising avenue for crossover is the potential for mesoscale models to uncover larger building blocks of chromatin.

While the nucleosome is the fundamental unit, recent experimental evidence points towards larger potential candidates.

Evidence of small fibers comprising the genome have already been the subject of recent super-resolution imaging techniques, such as stochastic optical reconstruction super-resolution microscopy (STORM), and recent chromosome conformation capture methods.

One STORM experiment links pluripotency and chromatin structure to the size of these clutch-like motifs, which are on the order of 4-16 nucleosomes.\cite{Ricci2015}

These clutches are also linked to the DNA packaging density through acetylations. \cite{Otterstrom2019}

Advances in chromosome conformation capture methods can also map the orientations of all entering and exiting DNA, called Hi-CO.\cite{Ohno2019}

These methods have uncovered two recurrent structural tetranucleosome motif. Perhaps nucleosome resolution models may not be necessary after all, and simpler, less demanding alternatives could be considered.

% Current setbacks and prospects

The current grand challenge for chromatin modeling is to explain epigenetic phenomena in terms of mechanistic, molecular processes.

To effectively understand the structure of chromatin - even at these small scales - one must consider the effect of variable DNA linker length, PTMs, linker histones, histone variants, etc.

Perhaps first on that list is to identify the modifications that have direct impact on the structural properties of chromatin.

In this vein, atomistic simulations have identified the effects of a few such modifications.

\textcolor{blue}{In addition, chromatin is far-from equilibrium in the nucleus due to the effect of ATP-dependent chromatin remodelers and elongation by RNA polymerase II.

Through unique inference about these molecular processes, a few models are beginning to introduce these effects with success, providing great promise for their use in future models.} \cite{Laghmach2019,Brackley2017,Michieletto2019}

This wide parameter space is still in need of significant computational, experimental, and theoretical efforts from the chromatin community; multiscale approaches, such as those outlined here, are likely to play an important role in that endeavor.

Bridging the length scales of chromatin modeling will allow for a hitherto unprecedented fundamental understanding of chromatin.

With exciting new advances in experimental techniques, simulations could provide the critical insights into chromatin structural functionality that are needed to make progress.

%With current ChromSTEM and Hi-C methods unraveling the \textit{in vivo} structure of chromatin, simulations could interpret the functionality of the chromatin structure from first principles.

\section*{Author Contributions}

Joshua Moller and Juan de Pablo wrote the article.

\section*{Acknowledgments}

The authors would like to thank Joshua Lequieu and Andr\'es C\'ordoba for insight in chromatin multiscale modeling.

Additionally, we thank Soren Kyhl for conversations on the potential of future models.

Lastly, we thank Nicholas Jackson, Viviana Palacio-Betancur, and Cody Bezik with editing this work.

% Uncomment if using bibtex (default)

\bibliography{les_houches}

\end{document}

%% Experimental techniques that have advanced understanding of chromatin structure and packing

%To describe the structural complexity of chromatin, researchers have likened packaged chromatin structure to that of protein folding.

%Where proteins can be identified by x-ray crystallography methods, chromatin falls into the ``dark region'' of characterization: it is too large to capture through x-crystallography but too small to be characterized with conventional microscopy.

%Nevertheless, experimental techniques have elucidated pieces of the chromatin puzzle at multiple length-scales.

%At the primary level of organization, the nucleosome is considered to be the fundamental unit of the chromatin fiber.

%Small enough to be characterized by x-ray crystallography, landmark studies revealed the nucleosome structure at atomistic resolution.\cite{Luger1997,Davey2002}

%Chromatin organization and structure beyond this scale has been elusive to understand.

%The structural organization of chromatin in the cell, or the 3D genome has been a substantial goal over the last few decades.