



# Bottom-up modeling of chromatin segregation due to epigenetic modifications

Quinn MacPherson<sup>a</sup>, Bruno Beltran<sup>b</sup>, and Andrew J. Spakowitz<sup>b,c,d,e,1</sup>

<sup>a</sup>Department of Physics, Stanford University, Stanford, CA 94305; <sup>b</sup>Biophysics Program, Stanford University, Stanford, CA 94305; <sup>c</sup>Department of Chemical Engineering, Stanford University, Stanford, CA 94305; <sup>d</sup>Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305; and <sup>e</sup>Department of Applied Physics, Stanford University, Stanford, CA 94305

Edited by José N. Onuchic, Rice University, Houston, TX, and approved October 24, 2018 (received for review July 17, 2018)

**We use a chromosome-scale simulation to show that the preferential binding of heterochromatin protein 1 (HP1) to regions high in histone methylation (specifically H3K9me3) results in phase segregation and reproduces features of the observed Hi-C contact map. Specifically, we perform Monte Carlo simulations with one computational bead per nucleosome and an H3K9me3 pattern based on published ChIP-seq signals. We implement a binding model in which HP1 preferentially binds to trimethylated histone tails and then oligomerizes to bridge together nucleosomes. We observe a phase reminiscent of heterochromatin—dense and high in H3K9me3—and another reminiscent of euchromatin—less dense and lacking H3K9me3. This segregation results in a plaid contact probability map that matches the general shape and position of published Hi-C data. Analysis suggests that a roughly 20-kb segment of H3K9me3 enrichment is required to drive segregation into the heterochromatic phase.**

epigenetic regulation | genomic architecture | Hi-C | chromosomal organization | polymer simulation

**C**hemical modifications to a chromosome correlate with its 3D organization in the nucleus (1), and this 3D organization is integral to controlling gene expression (2). Thus, identifying the physical mechanisms by which epigenetic modifications control chromosomal organization would provide a predictive framework for bridging epigenetics and expression.

The fundamental organizational unit of chromosomal DNA is the nucleosome, which consists of a 147-bp segment of DNA wrapped around eight histone proteins. Among the host of epigenetic modifications, methylation of the ninth lysine of the N-terminal tail of histone protein 3 (i.e., H3K9) is a prototypical example where enrichment of such a modification leads to large-scale changes in physical state and biological function. Enrichment in H3K9 methylation correlates with the DNA and associated proteins composing chromatin existing in a heterochromatic state that is densely packed and generally less expressed than more loosely packed euchromatin (2). Roughly micrometer-size domains of euchromatin and heterochromatin are observed in cell nuclei by electron (3) and fluorescence microscopy (4). Predicting how epigenetic markers affect the physical segregation of the chromosome is critical to interpreting the biological consequences of the epigenetic code (5).

The determination of the 3D organization of DNA from nucleosomes to the full chromosome remains a challenge. Interphase chromatin organization appears quite random (6), with cell-to-cell variability in relative loci positions (7). The most prominent nonrandom feature of chromatin organization is the physical connectivity of the DNA backbone itself, and physical models of chromatin must begin with this connectivity (8–21).

Experimental advances in chromosome-capture techniques (e.g., Hi-C) have revealed many additional features of chromatin organization (1, 22). These techniques use chemical cross-linking of loci that are in spatial proximity, followed by sequencing. The most prominent feature of the resulting contact maps is a plaid-

like pattern of high contact probability indicative of two or more types of chromatin that self-associate.

The observed Hi-C patterns have inspired a number of block copolymer models of chromatin where the blocks represent genomic regions with various epigenetic marks (17–21). Simulations of alternating block copolymers produce plaid contact patterns reminiscent of the Hi-C data, but do not incorporate genomic data (17, 19, 23). Jost et al. (19) demonstrate that diagonal Hi-C features are predicted when the blocks of the polymer correspond to epigenetic domains and that a few off-diagonal features are retained when the simulations are seeded with initial configurations containing them.

Further efforts aim to reproduce the Hi-C data with in silico block copolymers, using a top-down approach where model parameters are trained for agreement with Hi-C contact maps (17, 20, 21). DiPierro et al. (20, 21) classify the chromatin into six block types, using 11 epigenetic marks. After training the interaction constants between these block types as well as nonspecific binding energy as a function of genomic distance between loci, the model is able to predict Hi-C contact patterns, conclusively showing that the information in the epigenetic marks is sufficient to predict much of the observed Hi-C pattern. However, the DiPierro et al. (20, 21) model does not explain the physical mechanisms behind the fitted interactions, and the coarse-grained nature of the model makes it difficult to predict biological mechanisms for how chromosomal segregation affects accessibility and how epigenetic marks are established and maintained.

In this paper, we use a bottom-up approach to modeling chromosomal DNA. We develop a physical polymer model of chromatin and explore the consequences of introducing a binding model where heterochromatin protein 1 (HP1) is more likely to bind to trimethylated nucleosomes, which we identify from

## Significance

**Predicting how epigenetic marks control the 3D organization of the genome is key to understanding how these marks regulate gene expression. We show that a physical model of a chromosome with experimentally measured local interactions segregates into euchromatin- and heterochromatin-like phases. The model reproduces many of the features of the large-scale organization of the chromosome as measured by Hi-C. Our work provides an estimate of the amount of epigenetic marking needed to segregate a gene into heterochromatin.**

Author contributions: Q.M., B.B., and A.J.S. designed research; Q.M. and B.B. performed research; Q.M. and B.B. analyzed data; and Q.M., B.B., and A.J.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup> To whom correspondence should be addressed. Email: [ajspakow@stanford.edu](mailto:ajspakow@stanford.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1812268115/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1812268115/-DCSupplemental).

Published online November 26, 2018.

H3K9me3-ChIP. In contrast to the top-down approach of fitting Hi-C, we use only experimentally measured or physically motivated parameters in our model and observe to what extent this untrained model reproduces the Hi-C data. Our untrained model is not susceptible to reproducing experimental biases in the Hi-C data and is advantageous for establishing the causal effect of the introduced interactions on the organization. While we do not completely reproduce the Hi-C contact maps, we observe that this binding model alone explains much of the plaid pattern. Visualizing our modeled microstructure, we observe a phase segregation reminiscent of heterochromatin and euchromatin. Additionally, the model aids in the interpretation of the ChIP-seq data by providing physically motivated guidelines for predicting whether a chromosomal segment is segregated.

## The Model

**Chromosomal DNA.** We simulate each of the  $\sim 400,000$  nucleosomes of human chromosome 16 as an individual bead. We use the shearable, stretchable wormlike chain model described in *SI Appendix* (24–26), which captures the energetics of a semiflexible polymer chain while allowing for the greater bead spacing needed for large-scale simulations. For simplicity, we model chromatin as having the same stiffness as bare DNA with nucleosomes spaced a backbone path length 16.5 nm apart, corresponding to a 50-bp linker between nucleosomes. See *SI Appendix* for the effect of reducing this stiffness. A more detailed model of chromatin would include the effects of geometrical constraints from bound proteins, chain topology, linker histones, the entry and exit angles of the DNA from each nucleosome, geometric effects that tend to align adjacent fibers, and variation in linker lengths. Here, we present a bottom-up model on which detailed effects can be built.

**Confinement.** The DNA is highly confined by the nucleus and, in the absence of a confining boundary, it would expand to occupy an unconfined radius of gyration much larger than the radius of the nucleus. We simulate only a single chromosome (of 46), so we choose a confinement 1.8  $\mu\text{m}$  in diameter, roughly the size of a chromosome territory (27). We argue that since each chromosome is primarily found within its chromosomal territory in the nucleus (28), this smaller single-chromosome confinement is physically reasonable.

**Nonspecific Interaction.** Interactions between chromosomal segments include electrostatic interactions, van der Waals forces, steric interactions, and interactions with a complex solvent containing a zoo of interacting proteins. We take a coarse-grained approach and define the interaction free energy from the sum of these effects to be

$$F_{\text{int}}(\phi_c) = \begin{cases} \chi \Delta^3 \phi_c^2 & \text{if } \phi_c < 0.5 \\ \infty & \text{otherwise,} \end{cases}$$

where  $\phi_c$  is the volume fraction of chromatin calculated inside discrete bins of width  $\Delta$ . To calculate the total interaction free energy, we divide the system into cells of volume  $\Delta^3$  and use linear density interpolation to calculate  $\phi_c$  for each cell as in ref. 29. We choose  $\Delta = 28.7$  nm to capture density fluctuations on the size scales of interest. Each bead is given a volume of  $520$  nm<sup>3</sup>, roughly corresponding to the volume of a nucleosome. We choose the  $\chi$  parameter to distribute chromatin without leaving large chromatin-less regions, but not so large to prevent significant variation in chromatin density. See *SI Appendix* for details on the choice of  $F_{\text{int}}$ . While this binned density approach does not guarantee individual nucleosomes do not overlap, it is appropriate for capturing equilibrium density fluctuations of disordered chromatin on length scales greater than  $\Delta$ .

**Methylation Profile.** We focus on how chromatin organization depends on H3K9me3, which is associated with pericentric heterochromatin (30). We assume a fixed H3K9me3 profile based on ChIP-seq data from GM12878 cells (31, 32). We divide the ChIP-seq signal into 200-bp bins, corresponding to our nucleosome spacing. As there are two H3 tails per nucleosome, we apply cutoffs to the binned signal to classify each nucleosome into both tails trimethylated, one tail trimethylated, or neither tail trimethylated. For simplicity, we used evenly spaced cutoffs chosen so that roughly half of all tails are classified as trimethylated. See *SI Appendix* for details. Our trimethylated state is meant to capture the effects of in vivo trimethylation [20–30% of histone tails (33, 34)] on HP1 binding, and to a lesser extent, the effects of dimethylation [30–40% of tails (33, 34)].

**H3K9me3-Dependent HP1 Binding.** The chromatin regulatory protein HP1 binds histone tails through its chromodomain. Its chromo-shadow domain oligomerizes with other HP1s from different histone tails, which may be from different nucleosomes in close spatial proximity (35). HP1 preferentially binds to trimethylated tails, condensing regions of the chromatin marked with H3K9me3. In this work, we do not have the detail to differentiate the paralogs HP1- $\alpha$  and HP1- $\beta$ , so we refer only to the combined effect of the two.

We follow Mulligan et al. (30), using data from Canzio et al. (36) to find methylated or unmethylated histone-HP1 binding affinities  $\epsilon_M = -0.01$  k<sub>B</sub>T and  $\epsilon_U = 1.52$  k<sub>B</sub>T. When nucleosomes with bound HP1 come within an interaction radius, they experience a  $J = -4.0$  k<sub>B</sub>T free-energy benefit from oligomerization (30). In this work, we choose the interaction distance as  $r_{\text{int}} = 3$  nm based on an estimate of the histone tail length (37). We use a two-state model (bound by an HP1 or not bound) for each H3 tail, biased by the chemical potential  $\mu = k_B T \log([\text{HP1}]_{\text{free}})$  induced by the concentration of free HP1. For computational purposes, we use a coarse-grained approach for calculating the interaction between separate nucleosomes based on  $\rho$ , the local number density of bound HP1 interpolated from a linear grid with spacing  $\Delta$  as in ref. 29. In summary, we include binding energies (*i*) between HP1 and histone tails, (*ii*) for the oligomerization of HP1, and (*iii*) for explicit intranucleosomal HP1 interaction,

$$E_{\text{bind}} = \sum_{i=1}^N \left[ \underbrace{J' \sigma_i^1 \sigma_i^2}_{(iii)} + \sum_{j=1}^2 \underbrace{(\epsilon_i^j - \mu) \sigma_i^j}_{(i)} \right] + \sum_{\text{bins}} \underbrace{\frac{1}{2} J v_{\text{int}} \Delta^3 \rho^2}_{(ii)}$$

where  $v_{\text{int}} = (4/3)\pi r_{\text{int}}^3$ ,  $N$  is the number of nucleosomes,  $\sigma_i^j$  is 1 if the  $j$ th H3 tail of the  $i$ th nucleosome is bound by HP1 and 0 otherwise, and  $\epsilon_i^j$  is  $\epsilon_M$  or  $\epsilon_U$ , depending on the methylation state of the corresponding histone tail. Note that  $J' = J(1 - v_{\text{int}}/\Delta^3)$  to prevent double counting of the interaction between two tails of the same nucleosome.

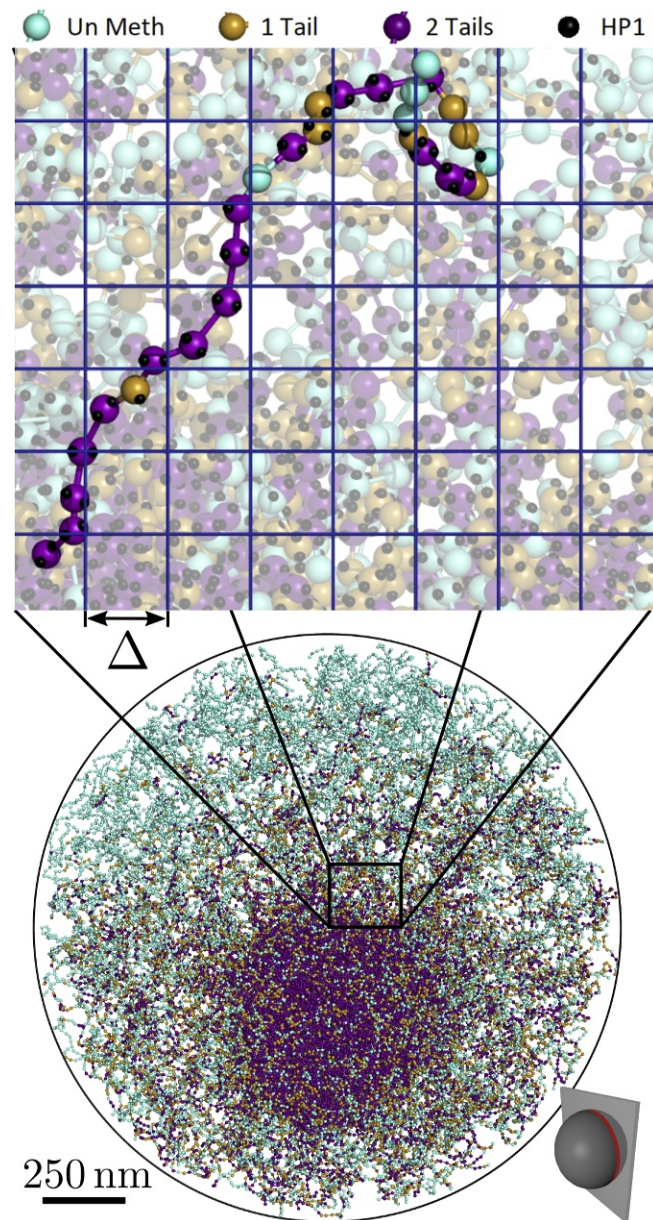
**Monte Carlo Algorithm.** We use Metropolis Monte Carlo sampling to draw configurations from the equilibrium ensemble of the above system, which has total energy from polymer-chain deformation energy, nonspecific repulsion, and HP1 binding:  $E = E_{\text{poly}} + \sum_{\text{bins}} F_{\text{int}}(\phi_c) + E_{\text{bind}}$ . Our simulations use moves which rotate a segment of polymer about the axis which runs through its ends, change the binding state of the histone tails, rotate a single bead, translate beads, and pivot the end of the chain. Moves were performed repeatedly to bring the system to equilibrium (*SI Appendix*). Our model does not include any topological constraints on the DNA. Therefore, our individual configurations are representative of thermodynamic equilibrium in the case where topoisomerase activity is



sufficiently high to relax topological constraints at equilibrium. Our Fortran source code was developed based on previously published code (38); the version used in this paper corresponds to the branch MBS2018.PNAS of the GitHub repository ([https://github.com/SpakowitzLab/wlcsim/tree/MBS2018\\_PNAS](https://github.com/SpakowitzLab/wlcsim/tree/MBS2018_PNAS)).

## Results

**H3K9me3–HP1 Binding Leads to Phase Segregation.** Fig. 1 shows a 115-nm-thick slice of a snapshot of our simulation. Phase segregation is observed between a dense phase enriched in



**Fig. 1.** (Top) Schematic representation of our model of chromosomal DNA. Each bead represents a single nucleosome and is colored cyan, tan, or purple if neither, one, or both of its histone-3 tails are trimethylated at lysine 9, respectively. Small black beads represent the number of HP1 proteins bound to a nucleosome. The chromatin density  $\phi_c$  and the number density of bound HP1  $\rho$  are calculated based on the displayed  $\Delta = 28.7$  nm grid. (Bottom) A 115-nm-thick slice through the center of a simulation snapshot with an HP1 concentration of 285  $\mu\text{M}$ .

H3K9me3 (purple), reminiscent of heterochromatin, and a more loosely packed region enriched in nucleosomes with neither tail trimethylated (cyan). A similar phase segregation driven by HP1 has been observed in vivo (39).

The nucleosome density difference between the two phases is controlled by the nonspecific interaction function  $F_{\text{int}}(\phi)$ . The choice of  $\chi$  shown in Fig. 1 is sufficiently small that the density of the heterochromatic phase is limited by the volume fraction cutoff of  $\phi_c \sim 0.5$ . See *SI Appendix* for determination of  $\chi$ .

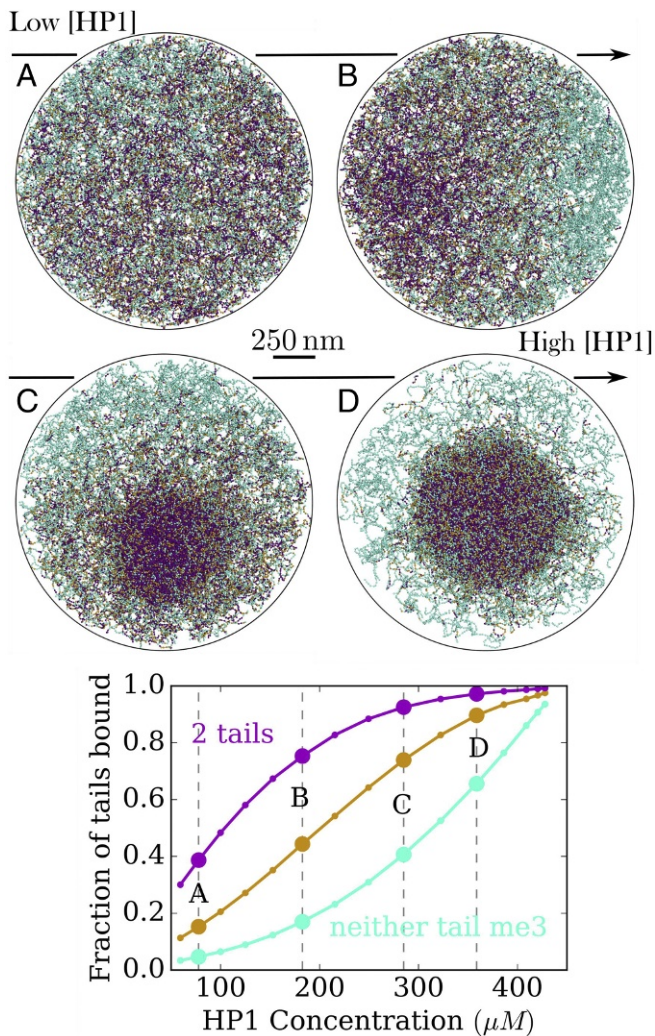
Heterochromatin is typically found in the periphery of the nucleus, although examples exist of inverted nuclei such as the rod photoreceptors in nocturnal mammals (40). Fig. 1 exhibits a structure with heterochromatin in the interior of the chromatin territory because no interaction between the chromatin and the confinement is included other than keeping the beads within the confinement. An interaction between the chromatin and nuclear lamina could be added to cause heterochromatic regions to move to the periphery, such as that used by ref. 41. In this paper, we focus on the thermodynamics of the binding model with as few free parameters as possible. We note that despite their striking visual difference, regular and inverted nuclei have been shown to have quite similar Hi-C profiles (41).

**HP1 Concentration Dependence.** We probe the effects of HP1 on segregation by varying its total concentration. The fraction of histone-3 tails bound by HP1 is plotted in Fig. 2 for nucleosomes with neither, one, or both tails methylated. At low HP1 concentration the probability that two bound HP1 molecules from different nucleosomes encounter each other is small. Without the oligomerization of HP1, the chromatin is a free polymer which distributes itself randomly throughout the nucleus as shown in Fig. 24. With increasing HP1 concentration the average fraction of tails bound by HP1 increases first for nucleosomes with fully trimethylated tails and later for nucleosomes with only one tail trimethylated. At intermediate concentrations (e.g., 285  $\mu\text{M}$ , as displayed in Figs. 1, 2C, and 3) there are enough bound HP1s in regions high in H3K9me3 to cause them to condense but not enough in regions low in the mark. This leads to the segregation of the regions. At sufficiently high HP1 concentrations nearly all nucleosomes are bound and condensed.

We constructed a specific binding model with experimentally derived parameters. However, any model with an attractive potential (in our model HP1 coupling  $J = -4$   $\text{k}_\text{B}\text{T}$ ) along with a specificity for heterochromatin (in our model differential binding energy  $\epsilon_U - \epsilon_M \sim 1.5$   $\text{k}_\text{B}\text{T}$ ) would show similar behavior. For example, we model the variability in ChIP-seq signal from H3K9me3 as coming from differing methylation of histones. At least some of this variability is caused by nucleosome density, which has a similar specificity in that it would also increase the bound HP1 content and H3K9me3 signal. We present HP1 concentration as modulating HP1 binding and, hence, the degree of compaction. In this regard, the HP1 level in the cell acts as a global regulator of chromatin state. Alternatively, phosphorylation of serine 10, by reducing the binding affinity of HP1 to H3K9 (42), could regulate chromatin architecture in the same way as we presented for HP1 concentration.

**Predicting Chromatin State from H3K9me3.** Our simulations provide a means of predicting whether each nucleosome is segregated into the densely packed heterochromatic phase or the more lightly packed euchromatic phase based on the H3K9me3 profile from ChIP-seq. We aim to extract general rules from our simulations for interpreting arbitrary ChIP-seq profiles for different chromosomes, cell types, or species without the need to perform simulations for each case. While it is true that the methylation state of a nucleosome is correlated with its chromatin state, the snapshot in Fig. 3, *Top Left* indicates that there are a number of unmethylated nucleosomes (cyan) in the heterochromatic region (dense,





**Fig. 2.** (A–D) Simulation snapshots at progressively higher HP1 concentrations. The plot shows, for each type of nucleosome, the fraction of tails bound by HP1 at different total HP1 concentrations. Cyan corresponds to nucleosomes with neither tail trimethylated, tan to one tail trimethylated, and purple to both tails trimethylated. The larger points correspond to the snapshots displayed in A–D with C being at 285  $\mu\text{M}$ , the value displayed in Figs. 1 and 3.

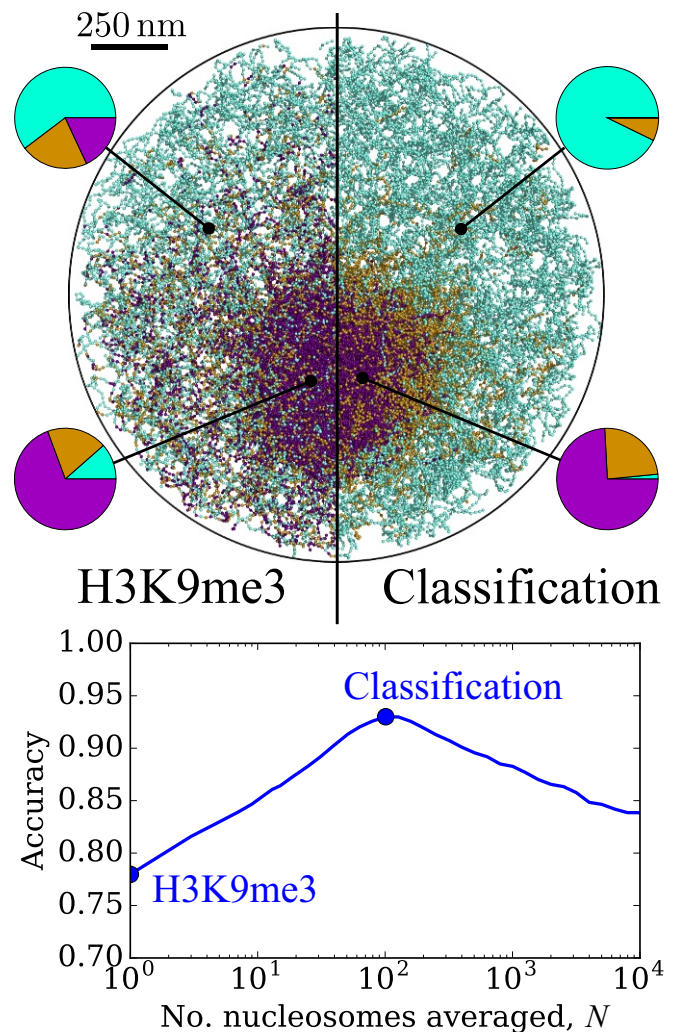
mostly purple phase) and vice versa. The phase in which a nucleosome resides is not determined just by the methylation state of that nucleosome, but also by the methylation state of its neighbors on the chromatin fiber. As in other copolymer systems (43), the strength of segregation is determined both by the strength of the interactions (defined by  $J$  in our model) and by the length of the segments with common chemical identity (the size of methylated regions in our model). Thus, when predicting in which phase a particular nucleosome is found, one must assess the average methylation of its genomic region.

It is not a priori clear how many adjacent nucleosomes  $N$  over which to average the degree of methylation when predicting the chromatin state of a nucleosome. Longer methylated regions are more likely to be segregated into the heterochromatic phase due to the additive binding energy of more HP1 molecules. The chain connectivity ensures that proximal genomic segments are in close spatial proximity, leading to more prevalent contacts and higher cooperativity in HP1 binding. However, genomic loci that are sufficiently separated do not cooperatively affect each other's chromatin state, since the chain connectivity does not

ensure that they are in close spatial proximity. Therefore, the thermodynamic advantage of cooperative HP1 binding starts to break down once the region size  $N$  becomes larger than the characteristic length of chromatin in the heterochromatic phase.

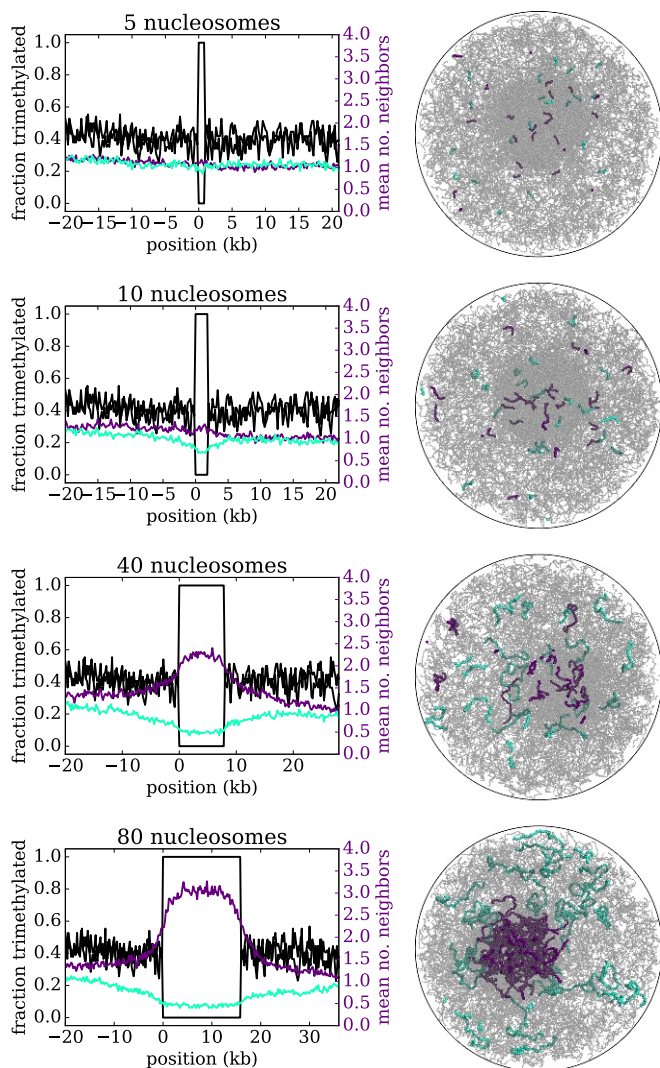
We define a simulated nucleosome as being in the heterochromatic phase if it is surrounded by a high density of chromatin (SI Appendix). We then predict which phase each nucleosome will be in, based on the average methylation state  $(N-1)/2$  nucleosomes to either side. Fig. 3 displays the accuracy of this prediction for a range of  $N$ . Classification based on the optimal window,  $N = 101$ , is displayed in Fig. 3, Top Right. For simplicity, we present a sliding-window average; similar results are obtained by replacing the window with a Gaussian or power-law filter (SI Appendix).

The above observations provide guidance for classifying the genome into regions of heterochromatin and euchromatin based on ChIP-seq data for H3K9me2/3. For this classification, our model recommends an average of the H3K9me3 signal over a



**Fig. 3.** (Top Left) Simulation slice with colors corresponding to the number of tails trimethylated. Pie charts show proportions of each color in the interior of each phase. (Top Right) Same as Top Left except the beads are recolored by the average methylation of their genomic neighbors, using a sliding window of 101 beads. Cyan corresponds to an average fraction trimethylated  $< 0.53$ , followed by tan, and finally purple for average trimethylation  $> 0.705$ . (Bottom) Accuracy with which the sliding-window average correctly classifies bead regions of high and low density as measured by number of neighboring beads. Optimal window size is  $\sim 100$  nucleosomes.





**Fig. 4.** Simulations with blocks of 5, 10, 40, and 80 nucleosomes in the input methylation sequence set either fully methylated (purple) or fully unmethylated (cyan). The two black curves show the average fraction of histone tails trimethylated in the vicinity of the inserted blocks. The average number of neighbors—a measure of chromatin compaction—shows that 80 nucleosome blocks segregate into their respective phases (as seen in the snapshots) and drag roughly 10-kb adjacent regions with them. In the snapshots the blocks are shown in color and surrounding chromatin with wild-type methylation is shown in gray.

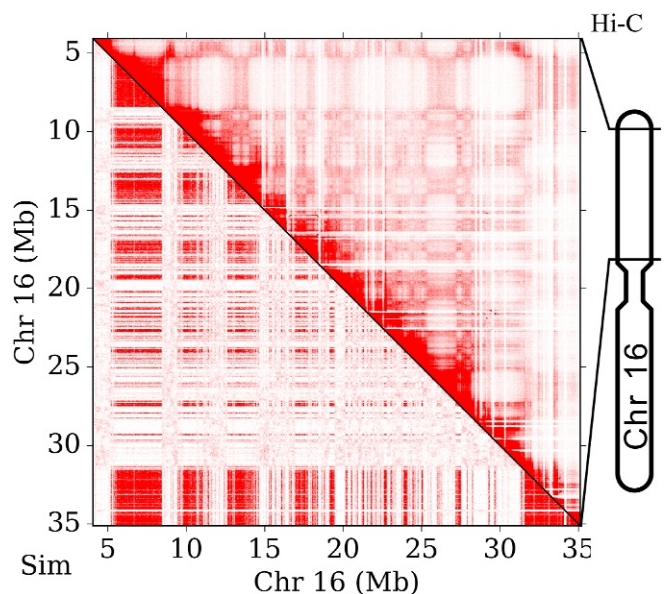
window  $\sim 20$  kb wide. Regions with low averaged signal will be euchromatic, regions with high ChIP-seq signal will be heterochromatic, and intermediate levels will be at the boundary or stochastically incorporated into either phase.

Beyond its usefulness for interpreting ChIP-seq data, the optimal window size implies that modulating the methylation level of a roughly 20-kb-long strand of chromatin would cause it to move between heterochromatic and euchromatic phases in the nucleus. To emphasize this, we perform simulations with a modified input methylation sequence containing blocks of fully methylated (purple in Fig. 4) or fully unmethylated (cyan) nucleosomes interspersed among the ChIP-seq-based methylation profile. As shown in Fig. 4, the blocks that contain only five (un)methylated nucleosomes are scattered throughout both phases. Their positions are determined by their surrounding methylation profile. Progressively larger blocks are more likely to be incorporated into their respective phases, with blocks of at least 80 nucleosomes

being well segregated. We note that this analysis differs from that presented in Fig. 3 because the latter incorporates the autocorrelation in methylation state present in the original ChIP-seq. The decaying number of neighbors to either side of the inserted block in Fig. 4 suggests that segregating blocks drag roughly 10-kb regions with them into their respective phases. The decay at the edge of unmethylated (cyan) blocks is generally outside the block while the decay around methylated blocks (purple) starts inside the block. This indicates that loci near epigenetic boundaries tend to sway euchromatic as measured by local density.

**Phase Segregation Reflected in Contact Maps.** In Fig. 5, we show a comparison of the simulated contact map (*Lower Left* triangle) with a Hi-C contact map for GM12878 cells (1) (*Upper Right* triangle). Bright red corresponds to a high likelihood that the two loci are in contact while white corresponds to a low probability. The simulated contact map was calculated by counting beads within 3 nm as “in contact” and averaging over 18 snapshots. Fig. 5 was generated at an intermediate HP1 concentration of  $285 \mu\text{M}$  (see *SI Appendix, Fig. S2* for Hi-C maps at other HP1 concentrations). We emphasize that none of the simulation parameters (apart from the color scale) are adjusted to fit the Hi-C data. Our goal is to understand which features of the Hi-C data are predicted by the model as a way of both verifying the model and interpreting the Hi-C data.

A prominent feature of the Hi-C map is the plaid pattern with rectangles of 1–4 Mb. Many of the rectangles found in the simulated contact map agree in both size and genomic location with the experimental data. In the simulation, this pattern is caused by regions high in H3K9me3 condensing together into a dense phase with elevated contact probability. While the corresponding rectangles in the Hi-C data could be caused by something else that is correlated only with H3K9me3, the affinity between HP1 and H3K9me3 provides a plausible mechanism. There are also rectangles in the Hi-C data which are not matched by the simulation. Some of these may be the result of interactions with epigenetic marks other than H3K9me3 that are not included in the model. Biochemical measurement of the



**Fig. 5.** Comparison of simulated contact probabilities (*Lower Left* triangle) and those from Hi-C data from Rao et al. (1), GEO accession no. GSM 733664 (*Upper Right* triangle). The numbering is in megabase pairs along chromosome 16. The region shown is the majority of the arms of the chromosome. The model predicts multiple features of the Hi-C contact map.

interactions selective to other marks would allow for a more complete physics-based model of chromatin. We note that while the Hi-C pattern appears like a checkerboard (i.e., corners touching), the simulation pattern has squares surrounded by continuous white bands. This difference may suggest that the heterochromatic and euchromatic regions are closer in density than in the model. Alternatively, the Hi-C data collection method may not result in a higher read count from a dense area in the way that counting neighbors does.

The Hi-C data also contain a number of white lines with widths on the order of 100 kb. These features are reliably matched by the simulation. However, this good agreement could also be explained by the existence of genomic regions that Hi-C and ChIP-seq are biased against and therefore show up as white in both.

Another prominent feature of the Hi-C contact map is the bright region near the diagonal, representing a higher contact probability for regions which are close genomically. While the connectivity of the chromatin polymer in our model does result in an increased contact probability for regions very close to each other genomically, this effect influences the contact probability only at short length scales not visible in Fig. 5. On the scale of megabases, the simulation shows little preferential contact for genomically proximal loci. This difference in contact probability scaling has been explained in the past by arguing that the chromatin is trapped in a nonequilibrium fractal globule state (15) or by the action of loop extrusion proteins (13). Our simulations confirm that extra physics are required to explain this scaling.

## Summary

The 3D chromatin architecture and its regulation by epigenetics are a complex system involving many interactions, most of which have unknown energetics. By introducing a simple model with as few unknown parameters as possible, we can understand the implications of these few interactions. Our model consists of a self-repulsive polymer, discretized on the single-nucleosome level and placed in a confinement commensurate to a single-chromosome territory. The only distinguishing feature we give the beads is a methylation state based on the ChIP-seq signal for H3K9me3. Introducing an HP1-binding model based on measured energetics (30, 36), we show that these interactions are capable of causing a phase segregation into a dense, H3K9me3-rich phase and a less dense, H3K9me3-lean phase. This segregation is capable of reproducing some of the features of the Hi-C map on the megabase size scale. We leverage this model to find that translocation of a nucleosome into heterochromatin is governed by the methylation state of the 20 kb surrounding it. Our approach is amenable to further exploration of the physical phenomena that underlie the organization and function of chromatin in a living cell.

**ACKNOWLEDGMENTS.** The authors thank Bradley Bernstein and Suhas Rao, respectively, for ChIP-seq and Hi-C data. The authors thank Johannes Nübeler for his valuable feedback. Financial support for this work is provided by the National Science Foundation (NSF), Physics of Living Systems Program (PHY-1707751). Q.M. and B.B. acknowledge funding support from the NSF Graduate Fellowship program (DGE-1656518). B.B. acknowledges support from NIH Training Grant T32GM008294.

- Rao SS, et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159:1665–1680.
- Grewal SIS, Moazed D (2003) Heterochromatin and epigenetic control of gene expression. *Science* 299:798–803.
- Fey EG (1986) The nonchromatin substructures of the nucleus: The ribonucleoprotein (RNP)-containing and RNP-depleted matrices analyzed by sequential fractionation and resinless section electron microscopy. *J Cell Biol* 102:1654–1665.
- Towbin BD, et al. (2012) Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* 150:934–947.
- Grewal SI, Jia S (2007) Heterochromatin revisited. *Nat Rev Genet* 8:35–46.
- Ou HD, et al. (2017) ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* 357:eaag0025.
- Dekker J (2012) Three-dimensional architecture of genomes. *Genome Organization and Function in the Cell Nucleus*, ed Karsten R (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany), pp 211–234.
- Rosa A, Everaers R (2008) Structure and dynamics of interphase chromosomes. *PLoS Comput Biol* 4:e1000153.
- Strickfaden H, Cremer T, Rippe K (2012) Higher order chromatin organization and dynamics. *Genome Organization and Function in the Cell Nucleus*, ed Rippe K (Wiley-VCH, Heidelberg, Germany), pp 417–447.
- Rao SS, et al. (2017) Cohesin loss eliminates all loop domains. *Cell* 171:305–320.e24.
- Benedetti F, Dorier J, Burnier Y, Stasiak A (2013) Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes. *Nucleic Acids Res* 42:2848–2855.
- Paulsen J, et al. (2017) Chrom3D: Three-dimensional genome modeling from Hi-C and nuclear lamin-genome contacts. *Genome Biol* 18:1–15.
- Nuebler J, Fudenberg G, Imakaev M, Abdennur N, Mirny L (2018) Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc Natl Acad Sci USA* 115:E6697–E6706.
- Wong H, et al. (2012) A predictive computational model of dynamic 3D interphase yeast nucleus. *Curr Biol* 22:1881–1890.
- Mirny LA (2011) The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res* 19:37–51.
- Goloborodko A, Imakaev MV, Marko JF, Mirny L (2016) Compaction and segregation of sister chromatids via active loop extrusion. *eLife* 5:1–16.
- Chiariello AM, Annunziatella C, Bianco S, Esposito A, Nicodemi M (2016) Polymer physics of chromosome large-scale 3D organization. *Sci Rep* 6:1–8.
- Michieletto D, Orlandini E, Marenduzzo D (2016) Polymer model with epigenetic recoloring reveals a pathway for the de novo establishment and 3D organization of chromatin domains. *Phys Rev X* 6:041047.
- Jost D, Carrivain P, Cavalli G, Vaillant C (2014) Modeling epigenome folding: Formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res* 42:9553–9561.
- Di Pietro M, Zhang B, Aiden EL, Wolynes PG, Onuchic JN (2016) Transferable model for chromosome architecture. *Proc Natl Acad Sci USA* 113:12168–12173.
- Di Pietro M, et al. (2017) De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture. *Proc Natl Acad Sci USA* 114:12126–12131.
- Lieberman-Aiden E, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326:289–293.
- Barbieri M, et al. (2012) Complexity of chromatin folding is captured by the strings and binders switch model. *Proc Natl Acad Sci USA* 109:16173–16178.
- Koslover EF, Spakowitz AJ (2013) Discretizing elastic chains for coarse-grained polymer models. *Soft Matter* 9:7016.
- Koslover EF, Spakowitz AJ (2013) Systematic coarse-graining of microscale polymer models as effective elastic chains. *Macromolecules* 46:2003–2014.
- Koslover EF, Spakowitz AJ (2014) Multiscale dynamics of semiflexible polymers from a universal coarse-graining procedure. *Phys Rev E* 90:013304.
- Cremer T, Cremer C (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2:292–301.
- Dekker J (2008) Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction. *J Biol Chem* 283:34532–34540.
- Pike DQ, Detcheverry FA, Müller M, de Pablo JJ (2009) Theoretically informed coarse grain simulations of polymeric systems. *J Chem Phys* 131:084903.
- Mulligan PJ, Koslover EF, Spakowitz AJ (2015) Thermodynamic model of heterochromatin formation through epigenetic regulation. *J Phys Condens Matter* 27:64109.
- ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74.
- Bernstein BE (2010) GM12878 H3K9me3 histone mods by ChIP-seq signal from ENCODE/Broad. Available at hgdownload.soe.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHistone/wgEncodeBroadHistoneGm12878H3k9me3StdSig.bigWig. Accessed October 28, 2017.
- Zee BM, Britton LMP, Wolle D, Haberman DM, Garcia BA (2012) Origins and formation of histone methylation across the human cell cycle. *Mol Cell Biol* 32:2503–2514.
- Huang H, Lin S, Garcia BA, Zhao Y (2015) Quantitative proteomic analysis of histone modifications. *Chem Rev* 115:2376–2418.
- Hiragami-Hamada K, et al. (2016) Dynamic and flexible H3K9me3 bridging via HP1 $\beta$  dimerization establishes a plastic state of condensed chromatin. *Nat Commun* 7:11310.
- Canzio D, et al. (2011) Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Mol Cell* 41:67–81.
- Arya G, Schlick T (2006) Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proc Natl Acad Sci USA* 103:16236–16241.
- Mao S, MacPherson Q, Qin J, Spakowitz A (2017) Field-theoretic simulations of random copolymers with structural rigidity. *Soft Matter* 13:2760–2772.
- Strom AR, et al. (2017) Phase separation drives heterochromatin domain formation. *Nature* 547:241–245.
- Solovei I, et al. (2009) Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* 137:356–368.
- Falk M, et al. (2018) Heterochromatin drives organization of conventional and inverted nuclei. bioRxiv:244038. Preprint, posted January 9, 2018.
- Fischle W, et al. (2005) Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* 438:1116–1122.
- Leibler L (1980) Theory of microphase separation in block copolymers. *Macromolecules* 13:1602–1617.