Thermodynamic model of heterochromatin formation through epigenetic regulation

Peter J Mulligan1, Elena F Koslover2 and Andrew J Spakowitz1,2,3,4

1 Department of Chemical Engineering, Stanford University, Stanford, CA 94305, USA
2 Biophysics Program, Stanford University, Stanford, CA 94305, USA
3 Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, USA
4 Department of Applied Physics, Stanford University, Stanford, CA 94305, USA

E-mail: ajspakow@stanford.edu

Received 30 May 2014, revised 2 July 2014
Accepted for publication 8 July 2014
Published 7 January 2015

Abstract
Gene regulation in eukaryotes requires the segregation of silenced genomic regions into densely packed heterochromatin, leaving the active genes in euchromatin regions more accessible. We introduce a model that connects the presence of epigenetically inherited histone marks, methylation at histone 3 lysine-9, to the physical compaction of chromatin fibers via the binding of heterochromatin protein 1 (HP1). Our model demonstrates some of the key physical features that are necessary to explain experimental observations. In particular, we demonstrate that strong cooperative interactions among the HP1 proteins are necessary to see the phase segregation of heterochromatin and euchromatin regions. We also explore how the cell can use the concentration of HP1 to control condensation and under what circumstances there is a threshold of methylation over which the fibers will compact. Finally, we consider how different potential in vivo fiber structures as well as the flexibility of the histone 3 tail can affect the bridging of HP1. Many of the observations that we make about the HP1 system are guided by general thermodynamics principles and thus could play a role in other DNA organizational processes such as the binding of linker histones.

Keywords: chromatin, epigenetic regulation, thermodynamic model

(Some figures may appear in colour only in the online journal)

1. Introduction

Inside the cell nucleus, DNA wraps around histone proteins to form nucleosomes that assemble into higher-order chromatin structures, resulting in a tightly packed nuclear environment. Within this environment, enzymes, transcription factors, and signaling molecules must find the appropriate genes to regulate and carry out the essential functions of the cell. The cell must organize its genome to allow transcription machinery access to active genes while restricting access to genes that are silenced. Many of these silenced genes, along with non-coding segments, are densely packed into heterochromatin regions within the nucleus, leaving loosely packed euchromatin regions more accessible [1]. Epigenetic marks, inherited from previous generations, are believed to be critical in establishing which regions condense [2]. However, the physical process by which such marks lead to compaction of the chromatin and segregation into heterochromatin domains is unknown.

The histone code hypothesis proposes that epigenetic information is passed on through heritable modifications to the histone proteins [3]. Several studies show an enrichment of methylation at lysine-9 of histone 3 (H3K9) within pericentric heterochromatin, with di- and tri-methylation most common within these regions [2]. Furthermore, the methyltransferase responsible for methylation at H3K9 predominantly resides within heterochromatin regions, and methyltransferase loss is shown to decrease the H3K9 methylation levels [4, 5]. A key molecule for heterochromatin formation and maintenance, heterochromatin protein 1 (HP1), binds methylated H3K9 with its chromodomain [6]. The chromodomain has a hydrophobic cage made up of aromatic residues that wraps around the methylation mark, with greatest affinity for tri-methylation marks at H3K9 [7]. Knockouts of HP1 are shown to completely
disrupt repression [8], while targeting HP1 to genomic regions can induce those regions to become heterochromatin [9, 10]. HP1 targeting results in a quantifiable degree of chromatin condensation [9], and local HP1 enhancement results in colocalization of distant chromosomal sites and repression of gene activity within the targeted regions [10]. HP1 and its analogues have been implicated in heterochromatin formation for species ranging from fission yeast (where it is called Swi6) [11], to Drosophila [Su(var)3-7] [12], to mammals (HP1) [6].

In this paper, we present a model that connects epigenetic regulation with a physical model for chromatin condensation via the binding of HP1. We model HP1 binding to H3K9 sites within chromatin fibers based on the geometric organization of the nucleosomes. The chromatin structure is determined using our previous work on minimum-energy nucleosome fibers [13], and we now consider the packing of these fibers into a dense array. HP1 dimerizes through its chromoshadow domain [14], and these interactions are critical for maintaining HP1 condensation within heterochromatin [15]. It is proposed that these bridging interactions between HP1 help condense the heterochromatin fibers [16]. We create a thermodynamic model for this hypothesis by introducing a coupling energy between bound HP1 that are proximally located within packed fibers, including both intra-fiber and inter-fiber HP1 bridging. We demonstrate that this interaction is critical to enhancing the concentration of HP1 in heterochromatin regions.

The presence of cooperative interactions leads to phase segregation of HP1 above some critical level of interaction. Extensive oligomerization is needed for HP1 to accumulate specifically in highly methylated regions. This accumulation can drive the segregation of heterochromatin and euchromatin into defined domains, as seen in vivo [17]. There exists a well-defined range of HP1 concentrations where this segregation occurs. HP1 binds to regions that have a degree of methylation above some threshold value, promoting localized condensation. Such a threshold leads to a buffering effect where small changes in methylation do not cause precipitous decondensation, as has been suggested in the case of methylation profile recovery in dividing cells [18].

Stability and control of heterochromatin condensation are critical in globally regulating the activation and repression of different genes [2]. Our model establishes a thermodynamic foundation for the role of cooperative binding of HP1 in the formation of heterochromatin regions in response to epigenetic signals. This approach provides a physical basis for predicting the stability of condensed regions and the threshold of conditions (i.e. HP1 concentration and degree of methylation) where precipitous decondensation occurs.

2. Theoretical model

Our theoretical model of heterochromatin contains two sequential steps of development. First, the nucleosome-array fiber structure is found using our physical model of chromatin compaction [13], defining the spatial arrangement of nucleosomes within a single fiber and relative to nucleosomes in neighboring fibers. Second, we model the binding of HP1 to H3K9 sites within these fibers. Based on the spatial layout of nucleosomes, we then identify the nucleosomes that are capable of forming bridges between bound HP1 and define an HP1-binding model that incorporates nucleosome methylation states and cooperative binding. The thermodynamic behavior of the resulting model is analyzed using several simulation and analytical approaches that are outlined below.

2.1. Chromatin fiber model

Our nucleosome-array model [13] captures the impact of nanoscale physical and geometric properties on the spatial arrangement of nucleosomes within a regular array. We generate compact structures of nucleosome arrays by energy optimization, considering the elasticity of the linker DNA (i.e. resistance to bending, twisting, and stretching deformation) as well as steric exclusion among the nucleosomes and linker DNA. This framework defines the layout of the nucleosomes within the array fibers, and the spatial position \( r_i \) of the \( i \)th H3K9 site is found using the position and orientation of its nucleosome and the crystal structure of the nucleosome core particle [19]. There are a total of \( N \) H3K9 sites in the model from \( N/2 \) nucleosomes, i.e. two H3K9 sites per nucleosome. In this work, we analyze several nucleosome-array structures based on the family of optimized configurations reported in [13].

The spatial distance \( r_{ij} = |\hat{r}_i - \hat{r}_j| \) between the \( i \)th and \( j \)th H3K9 sites defines the connectivity. If sites \( i \) and \( j \) are within a connectivity cutoff distance \( \alpha \) (i.e. \( r_{ij} < \alpha \)), they are said to be connected. We define a connectivity matrix \( C^i \), where connected sites \( i \) and \( j \) have \( C^i = 1 \); otherwise, the matrix element \( C^i \) is 0. The connectivity matrix excludes self connections; thus, \( C^i = 0 \). These connections define potential HP1 interactions.

We consider a hexagonal packing of chromatin fibers, as indicated in figure 1, with an inter-fiber spacing \( D \). All fibers within a given hexagonal array have the same nucleosome structure as defined from our nucleosome-array model [13], with a linker length of 35 bp and increasing height per nucleosome of 1.5 nm. The connectivity matrix \( C^i \) defines the connectivity between all \( i \) and \( j \) H3K9 sites, including those residing on the same fiber (i.e. intra-fiber connections) and on different fibers (i.e. inter-fiber connections). To identify a fiber within the regular hexagonal packing of fibers, we define the \( x \)-y lattice vectors \( \vec{a}_1 = D\hat{x} \) and \( \vec{a}_2 = D\hat{x}/2 + \sqrt{3}D\hat{y}/2 \) and the \( z \)-axis displacements \( b_1 \) and \( b_2 \). We then define the lattice indices \( m \) and \( n \) to identify the fiber within the hexagonal lattice, resulting in the \( m \)-\( n \) fiber having position \( \vec{R}_{mn} = m\vec{a}_1 + n\vec{a}_2 + (mb_1 + nb_2)\hat{z} \).

We find \( D \), \( b_1 \), and \( b_2 \) by maximizing the number of inter-fiber connections between neighboring fibers in the lattice while avoiding steric overlap between neighboring fibers. Since connections lead to cooperative binding, maximizing connections is equivalent to minimizing the binding free energy (see sections 2.2 and 2.3) for fixed fiber positions. In this work, we neglect fluctuations in the positions of the fibers within the lattice.

5 These structures were generated from our publicly available code, which can be found at http://web.stanford.edu/~ajsapkow/index.html.
Figure 1. Schematic of our theoretical model for HP1 binding to chromatin. Regular chromatin fibers are packed together in an infinite hexagonal array (a). The fiber array contains repeats of the single fiber (b), with each fiber in the array rotated and placed in the z-axis to maximize the number of connections (blue lines in (a)) between H3K9 sites on different fibers. HP1 is denoted in orange, and the HP1 free in solution acts as a source for HP1 to bind (c). HP1 can bind to tri-methylated H3K9 (red binding sites) or unmethylated H3K9 sites (blue binding sites). HP1 bound at connected sites interact (orange lines).

The site index i is then decomposed into a single-fiber index l and lattice indices m and n, such that i → l, m, n. Similarly, the site index j is decomposed into indices l′, m′, n′. Thus, the connectivity matrix is written as $C_{lmn}$. This decomposition will be useful in subsequent analyses of the model; however, derivation of the governing thermodynamic equations in the following sections is more convenient using the more compact i and j indices.

2.2. Thermodynamic model for HP1 binding

We develop a thermodynamic model for HP1 binding to packed arrays of nucleosome fibers. Figure 1 shows a schematic of our model, with each red or blue tail representing an H3K9 site and lines between the H3K9 sites identify the connectivity of the sites.

The binding of HP1 to the ith H3K9 site is represented by the occupancy $\hat{\sigma}_i$. If an HP1 is bound to the ith H3K9 site, $\hat{\sigma}_i = 1$; alternatively, if the site is unoccupied, $\hat{\sigma}_i = 0$. Each H3K9 site has an HP1 binding energy that is related to the degree of methylation. In this work, we consider two states of methylation for each H3K9 site. The methylation state of the ith H3K9 site is defined by the methylation index $\eta_i$. If the ith site is unmethylated, $\eta_i = 0$; whereas, if the ith site is tri-methylated, $\eta_i = 1$. Our model could be easily extended to include mono- or di-methyl marks; however, this simplified methylation model is sufficient to demonstrate the key physical processes governing HP1-induced condensation. We define $-\epsilon_M$ as the tri-methylated binding energy, and $-\epsilon_U$ as the weaker unmethylated binding energy. The red arms in figure 1 represent tri-methylated sites, and the blue arms are unmethylated sites. Finally, if the ith and jth H3K9 sites are connected (i.e. $C_{ij} = 1$) and both sites are occupied by HP1 (i.e. $\hat{\sigma}_i = \hat{\sigma}_j = 1$), there is a favorable coupling interaction, captured by the interaction energy $-J$.

The thermodynamics of HP1 binding are analyzed using a grand canonical ensemble that couples fluctuations in HP1 binding to an external reservoir with HP1 chemical potential $\mu$. This reservoir represents the unbound HP1 in solution that are available for binding, where higher HP1 concentration gives a larger HP1 chemical potential $\mu$ and more prevalent HP1 binding. The governing partition function for this ensemble is the grand canonical partition function $\Xi$, which is found by summing over all possible HP1-binding configurations for a fixed methylation profile $\{\eta\} = \{\eta_1, \eta_2, \ldots, \eta_N\}$. The probability of each binding configuration is weighted by

$$\exp\left(-\beta E + \beta \mu \sum_{i=1}^{N} \hat{\sigma}_i\right).$$

where the total energy satisfies

$$-\beta E = \sum_{i=1}^{N} [\eta_i \epsilon_M + (1 - \eta_i) \epsilon_U] \hat{\sigma}_i + \frac{J}{2} \sum_{i,j=1}^{N} \hat{\sigma}_i C_{ij} \hat{\sigma}_j,$$

and $\beta = 1/(k_B T)$. From this development, we write the grand canonical partition function as

$$\Xi(\{\eta\}) = \sum_{\hat{\sigma}} \sum_{\ldots} \ldots \exp\left(-\beta E + \beta \mu \sum_{i=1}^{N} \hat{\sigma}_i\right).$$

For a fixed methylation profile $\{\eta\}$, the grand canonical partition function $\Xi$ provides all of the statistical information necessary to determine the average HP1 binding to each site and the stability of the binding profile. The thermodynamic behavior for a fixed methylation profile is governed by the Landau potential $\Phi(\{\eta\}) = -k_B T \log \Xi(\{\eta\})$.

2.3. Thermodynamic behavior averaged over methylation profiles

Our current thermodynamic treatment captures the binding of HP1 to nucleosomes with a fixed methylation profile. Since we are interested in large-scale condensation of a genomic region, the region may be best characterized by the average methylation rather than the specific profile. Furthermore, methylation fluctuations are likely present in the nucleus due to nucleosome turnover and methyltransferase activity, so averaging over the methylation profile is a reasonable starting point for our study of heterochromatin condensation.

Our model assumes that HP1 binding and unbinding is relatively rapid compared to the timescales for changing the methylation rather than the specific profile. Furthermore, methylation fluctuations are likely present in the nucleus due to nucleosome turnover and methyltransferase activity, so averaging over the methylation profile is a reasonable starting point for our study of heterochromatin condensation.

We define the average methylation level $\bar{\eta} = \langle \eta_i \rangle_\eta$, where the $\langle \ldots \rangle_\eta$ indicates an average over all methylation profiles. In this treatment, we assume there is no correlation between the methylation of different sites, thus $\langle \eta_i \eta_j \rangle_\eta = \langle \eta_i \rangle_\eta \langle \eta_j \rangle_\eta = \eta^2$. The methylation-averaged thermodynamic
behavior is found by averaging $\Phi$ over the methylation profiles. This methylation-averaged $\Phi$ is defined as

$$
\langle \Phi(\eta) \rangle_\eta = -k_B T \log \Xi(\eta) - k_B T \prod_{i=1}^N \sum_{\eta_i=0,1} [ (1-\eta) (1-\eta_i) + \eta \eta_i \log \Xi(\eta_i) ].
$$

(4)

Evaluation of equation (4) cannot be analytically performed in its current state, so to make further progress, we apply the replica trick that has been used for a range of thermodynamic problems involving quenched disorder [22–32]. The replica trick allows us to calculate the average of a logarithm according to the identity

$$
\log \Xi(\eta) = \lim_{m \to 0} \frac{\Xi^m(\eta) - 1}{m},
$$

(5)

which leads to a more tractable average $\langle \Xi^m \rangle_\eta$. This transformation results in $m$ replicas of the thermodynamic system with a coupling that determines the impact of frustration arising from the quenched disorder. Using equations (3) and (4), we find

$$
\langle \Xi^m \rangle_\eta = \prod_{i=1}^N \sum_{\sigma_i=0,1} \sum_{\alpha=1}^m \sum_{\beta=1}^m \exp \left[ \sum_{j=1}^N \sum_{\beta=1}^m \sum_{\beta=1}^m \sigma_i^\beta C_j^\beta \sigma_j^\beta \right]
$$

$$
\times \sum_{j=1}^N \log \left[ (1-\eta) \exp \left( \sum_{\beta=1}^m \sigma_j^\beta \right) + \eta \exp \left( \sum_{\beta=1}^m \sigma_j^\beta \right) \right],
$$

(6)

where the sum over $\eta_i$ is readily performed. The product of $m$ partition functions $\Xi$ results in $m$ replicas that are indexed by the superscripts $\alpha, \beta$ to distinguish the replica identity from the subscript binding-site identity. At this stage of the theoretical development, there are no approximations to the original binding model.

We now perform several field-theoretic transformations, altering the intractable summations over the occupancies into integrals over field variables [33]. This approach is an exact transformation of the original theory; however, the resulting treatment is amenable to systematic approximation. We replace the occupancy $\sigma_i^{\alpha}$ with the field-variable occupancy $\sigma_i^\alpha$ by introducing a delta function $\delta(\sigma_i^\alpha - \hat{\sigma}_i^\alpha)$ and integrating over $\sigma_i^\alpha$. The Fourier representation of the delta function

$$
\delta(\sigma_i^\alpha - \hat{\sigma}_i^\alpha) = \int du_i^\alpha \exp[ i u_i^\alpha (\sigma_i^\alpha - \hat{\sigma}_i^\alpha) ]
$$

(7)

is introduced into the treatment to decouple $\sigma_i^\alpha$ and $\hat{\sigma}_i^\alpha$ at the expense of introducing the auxiliary field variable $u_i^\alpha$. We note that equation (7) neglects a factor of $1/(2\pi)$, since this factor merely resets the reference free energy and does not alter the thermodynamic model. The summations over $\hat{\sigma}_i^\alpha$ are $0, 1$ can now be solved exactly, leading to the expression

$$
\langle \Xi^m \rangle_\eta = \int \prod_{\alpha=1}^m d\sigma_i^\alpha d\sigma_i^\alpha \exp \left[ \sum_{j=1}^N \sum_{\beta=1}^m (\beta \mu + i u_j^\beta) \sigma_j^\beta + \frac{J}{2} \sum_{j,k=1}^N \sum_{\beta=1}^m \sum_{\beta=1}^m \sigma_j^\beta C_j^{\beta\gamma} \sigma_k^\gamma \right] + \sum_{j=1}^N \log \left[ (1-\eta) \exp \left( \epsilon_U \sum_{\beta=1}^m \sigma_j^\beta \right) + \eta \exp \left( \epsilon_M \sum_{\beta=1}^m \sigma_j^\beta \right) \right].
$$

(8)

At this stage, we have not made any approximations to the theory.

The current treatment cannot be solved exactly, so we turn to a systematic approximation scheme. The lowest-order treatment approximates the integrals over $\sigma_i^\alpha$ and $u_i^\alpha$ as the maximum value of the integrand, resulting in the mean-field approximation of the thermodynamic model. Higher-order approximations are found by expanding the argument of the exponential within the integrand in powers of the fluctuations away from the mean-field solution and performing integrals over the mean-field fluctuations. The resulting expansion results in the methylation-averaged Landau potential

$$
\langle \Phi(\eta) \rangle_\eta = \hat{\Phi}_\eta + \Delta \Phi_\eta,
$$

(9)

where $\hat{\Phi}_\eta$ indicates the mean-field solution, and fluctuation corrections to the mean field are $\Delta \Phi_\eta$. In equation (9), the subscript $\eta$ implies that the quantities are average over methylation profiles. This approach is analytically tractable up to quadratic order (i.e. Gaussian approximation), and the results are accurate for conditions with modest mean-field fluctuations (i.e. far from a critical point). Although critical fluctuations would render the solution inaccurate near a critical point, the current approach is sufficient to determine the general trend of behavior. Furthermore, strong segregation between heterochromatin and euchromatin within the nucleus indicates that under some physiological conditions the system is far from a critical point.

The mean-field approximation emerges by finding the maximum of the integrand of equation (8), i.e. performing a saddle-point calculation [33]. Maximizing the argument of the exponential in equation (8) with respect to $\hat{\sigma}_j^\beta$ gives the first saddle-point equation

$$
\hat{\sigma}_j^\beta = \frac{\exp(-i u_j^\beta)}{1 + \exp(-i u_j^\beta)},
$$

(10)

and the second saddle-point equation from maximizing with respect to $\hat{\sigma}_j^\beta$ is given by

$$
0 = \beta \mu + J \sum_{k=1}^N C_j^{\beta\gamma} \hat{\sigma}_k^\gamma + i u_j^\beta \epsilon_U (1-\eta) + \epsilon_M \eta \exp \left[ \epsilon_M - \epsilon_U \right] \sum_{j'=1}^m \hat{\sigma}_j^{\gamma'},
$$

(11)

where the barred quantities $\bar{\sigma}_j^\beta$ and $\bar{u}_j^\beta$ are the values at the saddle point. The auxiliary field $\bar{u}_j^\beta$ is purely imaginary [33], so we define $\bar{W}_j^\beta = i \bar{u}_j^\beta$ as a purely real auxiliary field. The
saddle point exhibits symmetry across the replicas; therefore, the saddle point solutions are independent of the replica index (i.e. $\bar{\sigma}^\alpha_j = \bar{\sigma}_j$ and $\bar{W}^\alpha_j = W_j$). Noting that the replica trick involves taking the $m \to 0$ limit, the saddle-point equations are solved for the self-consistent mean-field solution for the occupancy

$$J \sum_{k=1}^{N} c_k^j \bar{\sigma}_k + \log \left( \frac{1-\bar{\sigma}_j}{\bar{\sigma}_j} \right) + \beta \mu + \epsilon_{\mu}(1-\eta) + \epsilon_{M} \eta = 0, \quad (12)$$

and the mean-field Landau potential is

$$\bar{\Phi}_\eta = -[\mu + k_B T \epsilon_{\mu}(1-\eta) + k_B T \epsilon_{M} \eta] \sum_{i=1}^{N} \bar{\sigma}_i - \frac{k_B T}{2} \sum_{i,j=1}^{N} \bar{\sigma}_i \bar{C}_j^i \bar{\sigma}_j + k_B T \sum_{i=1}^{N} \bar{\sigma}_i \log \bar{\sigma}_i + k_B T \sum_{i=1}^{N} (1-\bar{\sigma}_i) \log(1-\bar{\sigma}_i). \quad (13)$$

The mean-field occupancy equation, given by equation (12), will reduce to a modified Langmuir isotherm with cooperative interactions when the connectivity is uniform for all binding sites and the sites are either all methylated or all unmethylated [34].

Corrections to the mean-field approximation incorporate binding fluctuations around the mean-field occupancy $\bar{\sigma}_i$. Here, we find corrections to the mean-field approximation up to quadratic order in the mean-field fluctuations. This is done by setting $\sigma_j^\alpha = \bar{\sigma}_j + \Delta \sigma_j^\alpha$ and $w_j^\alpha = \bar{w}_j + \Delta w_j^\alpha$ in the argument of the exponential in equation (8) and expanding the exponential argument up to quadratic order in $\Delta \sigma_j^\alpha$ and $\Delta w_j^\alpha$.

The linear-order perturbation in $\Delta \sigma_j^\alpha$ and $\Delta w_j^\alpha$ is zero, since fluctuations are about the mean-field solution that is a saddle point of the free energy. The resulting quadratic-order integrals are Gaussian in the field variables $\Delta \sigma_j^\alpha$ and $\Delta w_j^\alpha$ and can be performed analytically.

Performing this expansion and integrating over the auxiliary variable fluctuations $\Delta w_j^\alpha$, we arrive at the quadratic-order expression

$$\langle \Xi^m \rangle_\eta = e^{-m \beta \bar{\Phi}_\eta} \int \prod_{i=1}^{N} \prod_{a=1}^{m} d(\Delta \bar{\sigma}_i^a) \times \exp \left\{ -\frac{1}{2} \sum_{j=1}^{N} \sum_{l=1}^{m} \frac{1}{\bar{\sigma}_j^a(1-\bar{\sigma}_j^a)} (\Delta \bar{\sigma}_j^a) \right\} \times$$

$$\times \frac{1}{2} \left[ \sum_{j,k=1}^{N} \sum_{l,m=1}^{m} \bar{C}_j^l \Delta \sigma_j^l \Delta \sigma_k^m \right] + \frac{1}{2} \eta(1-\eta)(\epsilon_{M} - \epsilon_{\mu}) \sum_{j=1}^{N} \sum_{l,m=1}^{m} \Delta \sigma_j^l \Delta \sigma_j^m \right\}. \quad (14)$$

The third term in equation (14) represents the first instance of replica coupling in the theory, reflecting the impact of quenched disorder on the thermodynamic behavior. To perform the integrals over $\Delta \sigma_j^\alpha$, we diagonalize the non-replica-coupled terms. We define the matrix $M_{ij} = -J C_i^j + \delta_{ij} [\bar{\sigma}_j(1-\bar{\sigma}_j)]$, which has eigenvalues $\lambda_{\alpha}$ and corresponding eigenvectors $v_{\alpha}^i$ (i.e. $M_{ij} v_{\alpha}^i = \lambda_{\alpha} v_{\alpha}^i$). We make the substitution $\Delta \sigma_j^\alpha = \sum_{\alpha=1}^{N} v_{\alpha}^i \xi_{\alpha}^i$ in equation (14), perform the integrals over the amplitudes $\xi_{\alpha}^i$, and take the limit as $m \to 0$. These manipulations arrive at the result

$$\Delta \Phi_\eta = \frac{k_B T}{2} \sum_{l=1}^{N} \log[\bar{\sigma}_j(1-\bar{\sigma}_j)] + \frac{k_B T}{2} \sum_{a=1}^{N} \log \lambda_{\alpha} - \frac{k_B T}{2} \eta(1-\eta)(\epsilon_{M} - \epsilon_{\mu}) \sum_{a=1}^{N} \frac{1}{\lambda_{\alpha}}. \quad (15)$$

Our quadratic-order theory is used to find the average total occupancy $M = \sum_{i=1}^{N} \sigma_i$ by setting $\eta = \frac{\delta \Phi_\eta}{\delta \mu}$, given by

$$M = \sum_{i=1}^{N} \bar{\sigma}_i - \frac{1}{2} \sum_{i,j=1}^{N} \frac{1}{\bar{\sigma}_j(1-\bar{\sigma}_j)} \frac{\partial \bar{\sigma}_j}{\partial \mu} - \frac{1}{2} \sum_{a=1}^{N} \frac{1}{\lambda_{\alpha}} \frac{\partial \lambda_{\alpha}}{\partial \mu} - \frac{1}{2} \eta(1-\eta)(\epsilon_{M} - \epsilon_{\mu}) \sum_{a=1}^{N} \frac{1}{\lambda_{\alpha}} \frac{\partial \lambda_{\alpha}}{\partial \mu}. \quad (16)$$

This solution provides the basis for predicting the occupancy of HP1 for a given degree of methylation $\eta$ and HP1 chemical potential. The chemical potential $\mu$ is dictated by the concentration of unbound HP1 in the surrounding solution $[\text{HP1}]_{\text{free}}$ using the equation $\mu = \mu_0 + k_B T \log([\text{HP1}]_{\text{free}})$, where $\mu_0$ is a reference chemical potential that is determined experimentally. This relationship assumes the free HP1 do not oligomerize in solution (i.e. when unbound in the nucleoplasm), which will be discussed further.

### 2.4. Implementation of thermodynamic binding model

The model presented in the preceding sections is implemented using several computational approaches. First, we describe how we implement the replica field theory model on a hexagonal lattice of fibers as described in figure 1. The length of the fibers is selected such that the length matches an integer number of turns of the fiber super helix. This results in a connectivity $C_i^j$ that is amenable to continuous boundary conditions from the bottom to the top of each fiber. The number of integer turns and the number of fibers in the lattice is sufficiently large to render the results insensitive to the system size. With this lattice of fibers, the methylation state $\eta_i$ of the $i$th site is selected randomly with probability $\eta$ of the site being methylated (i.e. $\eta_i = 1$).

The implementation of the methylation-averaged model for a hexagonal lattice of fibers involves finding the mean-field occupancy $\bar{\sigma}_{\text{lna}}$ of the $i$th H3K9 site on the $m-n$ fiber in the lattice, exploiting the three site indices defined in section 2.1. Due to the symmetry of the methylation averaging, the $i$th site on each fiber is identical in connectivity and thus exhibit identical thermodynamic behavior. Thus, we write $\bar{\sigma}_{\text{lna}} = \bar{\sigma}_{000} = \bar{\sigma}_i$. Noting this symmetry, the mean-field
calculation becomes an effective single-fiber calculation with an effective connectivity between H3K9 sites $l$ and $l'$ given by

$$c_{l}^{r} = \sum_{m',n'=-\infty}^{\infty} C_{l}^{m'n'}.$$  

(17)

This greatly reduces the complexity of the self-consistent mean-field (equation (12)). Notably, the quadratic-order corrections are found using the full $C_{l}^{m'n'}$, which has a repeating block structure.

A simple extension to the self-consistent mean-field (equation (12)) governs the binding to a fixed methylation profile [$\eta$]. The governing equation for this system is identical to equation (12) with $\eta$ (the average methylation) replaced by $\eta$. Similarly, the quadratic-order fluctuation theory for a fixed methylation profile results in a similar correction to the Landau potential as presented in equation (15). For a fixed methylation profile [$\eta$], the correction free energy $\Delta \Phi$ contains the first two contributions from equation (15); however, the last contribution from the quenched disorder is not present.

The second approach we take is to use the replica equations developed above through quadratic order, but to instead apply them to a square or cubic lattice with repeating boundary conditions. The advantage of this approach is to capture the overall thermodynamic behavior of HP1 binding which is largely insensitive to the exact connectivity structure, as we discuss below. Furthermore, using lattice representations allows us to compare our results to previous work on similar systems like the random field Ising model [25].

The final approach discussed within this manuscript is to use Monte Carlo simulations [35] of binding and unbinding fluctuations on a hexagonal lattice of fibers as a test of the mean-field treatment and the replica trick for quenched disorder. We also perform simulations on square and cubic lattices for comparison. Furthermore, a single fiber is essentially a 1D lattice at sufficiently long length scales; thus, it is not appropriate to use the mean-field treatment of a single fiber. For simulations of single fibers, it is necessary to utilize Monte Carlo simulation as a treatment that properly accounts for correlated fluctuations.

We utilize these three approaches in the following section. In our discussion, we provide physical insight into the limits of validity of these methods of implementation and guidelines for their use.

3. Results

Our first goal is to determine the values of the thermodynamic parameters in our model. To determine the strength of the binding and interaction energies, we fit our model to experimental results for in vitro binding of HP1 to histone proteins. Although measurements of HP1 binding to histone peptides [37] provide valuable thermodynamic information, we fit our model to recent work by Canzio et al on HP1 binding to mononucleosomes in solution [36]. Binding to mononucleosomes is closer to the in vivo conditions, and comparison with the mononucleosome binding assays allows us to calculate the binding and interaction energies within our model.

We model mononucleosomes as two connected binding sites and assume the mononucleosomes are sufficiently distant from each other for their use. Figure 2(a) provides a visual representation of the binding states considered in our mononucleosome model. We set the reference state for the chemical potential to zero, and then use our model to calculate the unmethylated binding energy $\epsilon_U = -1.52 \pm 0.82 k_B T$, the tri-methylated binding energy $\epsilon_M = 0.01 \pm 0.83 k_B T$, and the interaction energy $J = 3.92 \pm 1.92 k_B T$. The fit can be seen in figure 2(b). The confidence intervals are generated using local sensitivity analysis and the MATLAB Statistics Toolbox function, nparci. With predictions for the thermodynamic parameters, we now proceed to analyze HP1 binding to nucleosome fibers. The key parameters for our analyses are the physiological binding energy difference $\Delta \epsilon = \epsilon_M - \epsilon_U = 5 k_B T$ and the interaction energy $J = 4 k_B T$. However, we also analyze the unphysiological binding energy difference $\Delta \epsilon = \epsilon_M - \epsilon_U = 5 k_B T$ in order to draw several conclusions regarding the behavior.
The only undetermined parameter for binding to a given nucleosome array structure is the connectivity distance $\alpha$, which will control the degree of connectivity. In figure 3(a), we plot the binding curves for the fiber shown in the inset. The fiber structure has a linker length of 35 bp and height per nucleosome of 1.5 bp (same structure as in figure 1). We vary the value of $\alpha$ from 6 nm (red) to 7 nm (blue). For larger values of $\alpha$, the transition from unbound to bound occurs at lower [HP1].

The dashed line indicates a first-order phase transition from the bound to the unbound state at that value of HP1 concentration. The location of the phase transition was determined by finding the intersection point for the bound and unbound phases in the chemical potential $\mu$ and pressure plane. The discontinuity at the phase transition does not arise in the experimental and theoretical binding curves for mononucleosomes shown in figure 2(b). We characterize this jump as being consistent with heterochromatin condensation of the array of nucleosome fibers. The binding of HP1 to H3K9 sites on chromatin fibers is analogous to the adsorption process of gas molecules to a surface that can be described by a Langmuir isotherm with interactions between bound molecules that also leads to similar binding curves [34].

Figure 3(b) shows the binding curves for a different fiber structure. This fiber, shown in the inset, has its nucleosomes oriented at 90 degrees relative to the fiber in figure 3(a), due to the change in linker length to 40 bp. Increasing $\alpha$ from 6 nm (red) to 7 nm (blue) shows similar trends for the location of the phase transition. For the smallest value of $\alpha$, there is no phase transition and instead we have a smooth binding curve, similar to what we saw in the mononucleosome case.

We test HP1 binding to a wide range of fiber structures with varying linker length, rise per nucleosome, connectivity cutoff distance $\alpha$, and H3 tail position. These analyses demonstrate that the coordination number $z$ is a major determinant of the behavior. For an array of nucleosomes, the coordination number $z = \frac{1}{\alpha} \sum_{i,j} C_{ij}$ represents the average number of neighboring sites and depends on the structure of the fibers. The coordination number for the fiber arrays in figure 3(a) are 4 (red) and 5 (blue), whereas those in figure 3(b) are 1 (red) and 4 (blue). Looking at these binding curves, as well as for other fibers and other values of $J$, we see that the concentration at which the phase transition occurs is largely a function of $J$, i.e. the product of the interaction energy and the coordination number. Furthermore, whether or not a phase transition occurs is also dependent upon $J$, as we discuss below. For this reason, we focus subsequent analyses in this manuscript on a connectivity defined as a cubic lattice. Although each individual fiber array will have a different exact connectivity, the major features can be captured by this simpler connectivity model. At the end of this section, we discuss some important ways that the specific fiber structure affects the behavior.

Figure 4 shows binding curves for binding to a 3D cubic lattice with coordination number $z = 6$. We show three different binding curves versus HP1 concentration [HP1]: an
unmethylated array with $\eta = 0$ and $\Delta \epsilon = 1.5 k_B T$ (blue), a tri-methylated array with $\eta = 1$ and physiological binding affinity $\Delta \epsilon = 1.5 k_B T$ (red), and a tri-methylated array with $\eta = 1$ and elevated binding affinity $\Delta \epsilon = 5 k_B T$ (green). In all three cases, the coordination number $\kappa$ and interaction energy $J$ results in the value $J z = 8 k_B T$. HP1 binding in each case increases monotonically as the concentration of free HP1 increases, and this interaction strength results in the fraction bound $\sigma$ exhibiting a discontinuous jump at some threshold value of $[\text{HP1}]$ that we associate with heterochromatin condensation. The greater the degree of methylation, the lower the concentration at which heterochromatin condensation occurs.

The binding curves in figure 4 also give an indication of the preferential binding of HP1 to tri-methylated versus unmethylated chromatins. The horizontal arrows in figure 4 show the range of $[\text{HP1}]$ values where HP1 is preferentially bound to tri-methylated regions over unmethylated regions for physiological $\Delta \epsilon = 1.5 k_B T$ (red) and elevated binding affinity $\Delta \epsilon = 5 k_B T$ (green). Since HP1 binds highly methylated fibers at lower concentrations, this creates a window of concentration in which HP1 will only be bound to highly methylated regions.

The preferential window for HP1 binding to tri-methylated regions is shown in figure 5. The lines plotted in figure 5 show the coexistence boundaries for 0% and 100% tri-methylated arrays, as a function of the interaction energy $J z$. The coexistence boundaries shift with $J z$ according to $[\text{HP1}] = \exp[-J z/2 + \eta \epsilon M + (1 - \eta) \epsilon_U]$, resulting in a strong sensitivity of these boundaries to the degree of methylation $\eta$.

We label points $X$, $Y$, and $Z$ for $J z = 8 k_B T$ in figures 4 and 5 to illustrate this window. For physiological $\Delta \epsilon = 1.5 k_B T$ (figure 5(a)), the low value of $[\text{HP1}]$ ($X$) results in both unmethylated regions (blue curve) and tri-methylated regions (red curve) being unbound by HP1. An intermediate value of $[\text{HP1}]$ ($Y$) results in preferential binding to the tri-methylated regions, resulting in segregation between tri-methylated regions (heterochromatin) and unmethylated regions (euchromatin). For large values of $[\text{HP1}]$ ($Z$), both the unmethylated and tri-methylated regions are bound by HP1. Since the in vivo concentration (labeled in yellow in figure 5) of HP1 has been measured between 0.1–1 $\mu M$ [38], the $J z$ value should fall within $2 - 9 k_B T$ to see the enhanced binding of HP1 to heterochromatin. Given a $J$ value of $2 - 5.8 k_B T$, this suggests that a value for the coordination number $\kappa$ of approximately 2–5 should be expected.

As can be seen in figure 5, no regions of the genome are bound by HP1 at low concentrations. At high concentrations of HP1, all the regions are bound. Within this window, HP1 is enhanced only in heterochromatin regions, facilitating chromatin compaction via bridging interactions between HP1 bound at different H3K9 sites.

The size of the enhancement window depends upon the difference in binding energies between tri-methylated and unmethylated H3K9. For $\Delta \epsilon = 5 k_B T$, the concentration window spans orders of magnitude, in comparison to the experimental $\Delta \epsilon$ value of $1.5 k_B T$ that only spans one order of magnitude. The smaller concentration window would appear at first glance to make heterochromatin less stable, since smaller changes in concentration could lead to decondensation. However, this would also allow the cell to effect condensation and decondensation through moderate changes in HP1 concentration.

The phase diagrams in figure 5 demonstrate how the cell can control which regions will condense by changing the concentration of HP1. When there is a high enough concentration of HP1, it will bind to all regions of the genome regardless of the degree of methylation. Similarly, a very low concentration of HP1 prevents any region from being bound. It is only within the enhancement window that more highly methylated areas can be selectively bound.

In figure 6, we show two methylation levels: 10% trimethylation (A), which corresponds to euchromatin regions,
modulate condensation by altering HP1 concentration. The coexistence curves (solid) for demarcate the enhancement window for 10% versus 90% tri-methylated arrays.

and 90% tri-methylation (B), which corresponds to heterochromatin regions. The arrows from B to B’ show how much the concentration of HP1 needs to decrease to release HP1 from the 90% tri-methylated regions. The starting concentration of HP1 was chosen such that the 10% region was just within the unbound phase to illustrate how large this change in concentration could potentially need to be.

Thus, HP1 could give the cell control over the condensation of fibers, which may be important during different stages of the cell cycle. As shown in figure 6, a cell may need to decondense heterochromatin regions to get access to them during replication. A high binding energy difference between the tri-methylated and unmethylated states, as in figure 6(b), means that the cell must change the concentration several orders of magnitude in order to release HP1 from the most highly methylated fibers. Since changing the concentration over so large a range may not be practical, a smaller binding energy difference affords the cell more control over HP1 binding.

As discussed above, the experimental results for the differences in binding energy between the tri-methylated state and the unmethylated state are only 1.5\( k_BT \) to 25-fold[36]. Based on percentages of free and bound HP1 from FRAP experiments and FCS experiments calculating the in vivo concentrations of HP1, we calculated an enhanced in vivo binding in heterochromatin of eight-fold. This can be explored with our model by taking the ratio of the occupancy of the 100% tri-methylated arrays over the occupancy of the unmethylated arrays. With no cooperativity (\( J_z = 0 \)), the binding enhancement is only about four-fold. In order to achieve the expected enhancement, the \( K_f \) for HP1 binding to H3K9unmod and H3K9me3 would need to have a much greater difference than is found experimentally.

Such enhancements can be achieved by strong cooperative interactions. The distinct heterochromatin and euchromatin regions we see within the cell suggests that these regions are phase segregated. To achieve this phase segregation, the value of \( J_z \) must be above the critical point for the system. We next use our model to determine what the critical value of \( J_z \) is to ensure a phase transition. Figure 7 shows the coexistence phase diagrams for fraction bound \( \sigma \) versus \( J_z \). For conditions below the coexistence curves, the fibers do not exhibit an abrupt cooperative transition in HP1 binding. Above the coexistence curves, HP1 binding exhibits an abrupt transition. Figures 7(a) and (c) show coexistence curves for \( \Delta \epsilon = 1.5 k_BT \) and \( \Delta \epsilon = 5 k_BT \), respectively. The plots on the right of figure 7 (\( \Delta \epsilon = 1.5 k_BT \) for figure 7(b)) and \( \Delta \epsilon = 5 k_BT \) for figure 7(d)) show binding curves versus [HP1] for \( J_z = 5 k_BT \) for 0%, 50%, and 100% methylation.

The mean-field theory (black curves in figures 7(a) and (c)) predicts that the critical value will occur at \( J_z = 4 k_BT \). This clearly underestimates the appropriate value, since the mean field theory (from equation (10)) neglects all fluctuations. To incorporate these fluctuations, we solved for the quadratic level perturbation (in equation (16)) on the system and found, for 0% methylation, a value of 4.4\( k_BT \) (blue lines). However, the true critical point for a three-dimensional cubic lattice occurs at \( J_z = 5.31 k_BT \) [39], so we are still underestimating the correct value. However, the phase diagrams away from the critical point are consistent with Monte Carlo simulations using the Metropolis algorithm to simulate HP1 binding and unbinding on a cubic lattice. The same qualitative behavior was shown, with the exact value of the critical point in line with previous observations [39].

Unlike 0% or 100% tri-methylation, which can have only one profile, there are many different ways of having 50% of the nucleosome sites tri-methylated. Some profiles may have clusters of methylation and some may be a more even distribution of methylation marks. In solving the replica field theory, we found the result for the quenched average of the binding of HP1 over all these methylation profiles. This added disorder in the system causes the critical point to rise. This can be seen in figure 7, where the critical point at the base of
the phase envelope increases from 0% (blue) to 25% (green) to 50% tri-methylation.

The critical point is highest for 50% methylation, due to the combinatorial nature of the methylation profile. The critical point also rises when $\Delta \epsilon$ rises, as we see in the two phase diagrams. The binding energy difference $\Delta \epsilon$ acts to enhance the impact of quenched disorder, resulting in the observed increase in the critical point. Both of these effects are captured by the last term of equation (16), which contains $\eta(1-\eta)\Delta \epsilon^2$. This term will be largest when the methylation level is at 50%, and increases with greater difference between the tri-methylated and unmethylated binding energies. Since the weight of this term is the same for 25% and 75% methylation, the binodal coexistence curve are the same for these methylation levels, as indicated by the green curve in figure 7.

Thus, one of the results of the smaller binding energy difference $\Delta \epsilon$ is a more robust HP1 binding. When $J_z = 5k_B T$, the binding curves maintain a clear phase transition across all methylation levels from the unbound to the bound phase for the smaller $\Delta \epsilon$ (figure 7(b)), but not for the larger one, as shown in figure 7(d). At 50% methylation (red), $J_z = 5k_B T$ is below the critical point for $\Delta \epsilon = 5k_B T$ and thus the HP1 binding can access more intermediately bound states. The loss of a phase transition in this case disrupts the switch-like feature that enabled the cell to change from bound to unbound (i.e. from condensed to uncondensed).

Another important consequence of a phase transition is that there is a threshold of methylation above which HP1 binds and condenses the heterochromatin regions. This is shown in figure 8, where a $J_z$ value of 5 $k_B T$ is above the critical point but a value of 3 $k_B T$ is not. In the case of the stronger interaction of 5 $k_B T$, HP1 will stay tightly bound to the fiber even if the methylation levels fluctuate but remain above 35%. In this plot, we have chosen an HP1 concentration of 0.4 $\mu$M. The location of the threshold methylation level will depend on the HP1 concentration, and for smaller [HP1], the threshold will occur at a larger methylation percentage. Once the methylation drops below this threshold value, the HP1 will unbind and the fibers decondense. However, when the interaction strength is only 3 $k_B T$, we do not see this
sharp phase transition, and instead the HP1 binding gradually changes with methylation level.

A large enough $J_z$ value ensures that gains or losses in methylation do not affect the binding behavior of HP1 unless they cause it to cross the threshold, creating a thermodynamic switch for the two chromatin states. Thus, the system is buffered and protected against small fluctuations. Furthermore, the cell only needs to recover a methylation level above this threshold to ensure faithful inheritance of the condensed state. In light of recent work suggesting a delay in recovering the full methylation profile [18], our model suggests a thermodynamic basis for heterochromatin recovery at high but reduced levels of methylation.

The actual $J_z$ values within the cell may vary according to the fiber structure and compaction. As discussed above, the cooperative binding nature of HP1 to the chromatin fibers is also determined by the interaction distance $\alpha$, which affects the value of $z$. Since HP1 is approximately 25 kDa, we can estimate HP1 size as a sphere of diameter 2–3 nm [40]. This puts a lower bound on the expected interaction distance of 4 nm (two HP1 molecules). Looking at the crystal structures of the chromodomain and the chromoshadow domain bound to the histone and another HP1 chromoshadow domain, respectively, we find that these domains are about 2 nm in the perpendicular direction [7, 41]. Since they are connected by a flexible linker region, two HP1 molecules could reasonably span 6–8 nm [40]. Since the histone tails are themselves flexible, with mobility of perhaps several nm, an upper range for the interaction distance could be more than 10 nm [42].

Looking at a range of different fiber structures, we find that for $\alpha$ values between 5 nm and 10 nm, the $z$ value for average number of neighbors ranges from 2–4. Since experimental measurements of $J$ are approximately 2–6 $k_B T$, this results in $J_z$ values from $4 k_B T$ to $24 k_B T$. Looking back at our phase diagram for HP1 binding in figure 4(b), the $J_z$ values within this range that also have enhancement windows within the biological concentration are the lower end of this window, suggesting that $J_z$ values may be around $8 k_B T$. This is well above the critical point, meaning that phase transitions would be able to segregate the chromatin regions.

The size of the interaction distance can drastically change the underlying connectivity within these fibers. Looking at the 35 bp linker length fiber with 1.5 nm of rise per nucleosome that is shown in figure 1, the bound HP1 will have at most only one other neighbor for values of $\alpha$ as low as 4 nm. Once $\alpha$ is greater than 5 nm, however, the network becomes percolated, with connections spanning the whole length of the fiber. This creates large-scale binding and condensation, which we would expect for heterochromatin structures to form. We also find that across all these potential structures, the tails face inward and create a large number of intrafiber reactions, as is shown in the leftmost image of figure 9. Only at large values of $\alpha$ would some of the bound HP1 be able to interact with neighboring fibers.

Additionally, we note that the picture presented in figure 1, with interfiber interactions creating a connected array of fibers, is only true for larger values of $\alpha$. In fact, $\alpha = 14$ nm, shown in figure 1, is the smallest value of $\alpha$ at which any of the fibers allows interfiber interactions. In figure 9(b), we plot the interfiber connections in red for one representative fiber structure (shown in the leftmost image of figure 9(c)), which has a 40 bp linker length and 1.5 nm rise per nucleosome. No interfiber connections arise even when varying the maximum interaction distance $\alpha$ to 15 nm. Across all potential fiber structures that we analyzed, the tails face inward and create a large number of intrafiber reactions, as is shown in the leftmost connectivity diagram of figure 9(c). Since this value of $\alpha$ may be larger than expected for HP1 bridging, compacted fibers in vivo may not have interfiber connections if this 30 nm fiber structure is maintained and the tails stay pointed inward.

Instead of attributing the tail flexibility as part of a larger interaction distance, we consider different potential tail locations. Figure 9 shows the connectivity in a compact fiber structure with varying tail locations, which are indicated by their position on the single nucleosome crystal structure with red circles. We chose the three locations based on the simulation work reported by Arya and colleagues [42], which found the H3 tail position for different salt values. These three positions were chosen to be representative of the range of H3 tail locations from the simulations.

We find that when the tails are in the location from the crystal structure, they point toward the inside of the fiber structures. This results in more intrafiber connections, but prevents neighboring fibers from interacting. Overall, the number of connections between H3K9 sites stays roughly the same, as the three lines in figure 9(a) track along with one another for varying values of the maximum interaction distance $\alpha$. When the tails instead point outward, they maintain some intrafiber interactions, but also can connect to neighboring fibers, as shown by the rising number of interfiber connections plotted in figure 9(b). The value of the maximum interaction distance $\alpha$ necessary to create interfiber connections is much smaller ($\approx 7$ nm, from green curve in figure 9(b)) than when the tails are in the crystal structure location (where $\alpha$ needed
Figure 9. Tail locations and resultant connectivity. Plot (a) shows the value of coordination number $z$ for the fibers shown in image (c). The red curves are for the first tail location (left image in (c)), the blue curves are for the second tail location, and the green curves are for the third tail location (right image in (c)). Plot (b) shows the connectivity between two neighboring fibers packed in a hexagonal array. The connectivity diagrams in image (c) show connecting lines between nucleosomes (squares in (c), colored according to the number of connections) for intrafiber (black) and interfiber connections (green) for the three tail locations, with $\alpha = 14$ nm. This smaller value corresponds to reasonable estimates of the distance that two HP1 molecules could span as discussed earlier. Finally, we note that these interfiber connections are critical to bringing fibers together and forming the condensed phase as is seen in vivo.

4. Discussion

Biological phenomena such as epigenetic inheritance are driven by a complex coordination of chemical and physical processes within the cell. We demonstrate a physical mechanism that would connect the binding of Heterochromatin Protein 1 to the formation of the compact chromatin structure of heterochromatin. Recent fluorescence recovery after photobleaching (FRAP) experiments detailed HP1 fluorescence recovery in heterochromatin and euchromatin regions, and while slower in heterochromatin, the recovery is nearly complete within minutes [20, 21]. Given the rapid diffusion of HP1 throughout a cell’s nucleus, HP1 accumulation is likely driven by a thermodynamic preference for heterochromatin regions and not due to kinetic factors such as trapping within the crowded chromatin structure. Our equilibrium binding model fits these observations by assuming that the cell will reach a binding equilibrium for HP1.

The model developed in this paper represents the most crystalline, compact case; it sets an upper bound for the cooperativity that could be found in vivo. Many other possible arrangements of nucleosomes may be possible in vivo, as the structure of the chromatin is not well understood [43]. The chromatin fiber model that forms the basis for this work demonstrates that any number of chromatin fiber structures could exist with low energy, and thus there may not be a unique 30 nm fiber [13]. The structures considered within this work ranged in diameter from 28 nm to 46 nm, in linker lengths from 35 bp to 70 bp, in axial packing density from 0.7 nm to 2.2 nm per nucleosome, and in nucleosome orientations both along and perpendicular to the fiber axis. These different structures lead to different connectivities, which can strengthen or weaken the cooperative behavior. In addition, compact fiber structures may not even exist within the nucleus, or they may only exist in heterochromatin regions [44].
Despite the variety of structures that could exist, our model does predict that the exact packing arrangements of the nucleosomes does not greatly affect the expected HP1 binding. Packed fibers may be rotated differently with respect to one another, or different nucleosomes within these fibers may be tri-methylated or unmethylated. As long as the total number of connections stays roughly constant, the binding will largely be determined by the total amount of tri-methylation. This means that the prevalence of the epigenetic mark will control the HP1 binding process. In this work, we consider the fibers as static and thus the structure does not change with respect to HP1 binding; future work would look at how these types of bridging interactions could influence the repeat length as well as the fiber structure, and other work has suggested these interactions should play a role [45].

Our model is also broadly applicable to other models for chromatin in the nucleus. This paper considers a number of compact structures from a model based on nucleosome steric and linker DNA energetics [13]. However, any model that could output the location of the H3K9 binding sites could be used as an input to our model. Since it is unknown how the chromatin folds in vivo on the length scale of the 30 nm fiber, or if that structure is even present, our model can be adapted to the outputs of other models for regular chromatin packing, as in [46]. On larger length scales, it has been proposed that the chromatin is packed with random loops [47] or as a fractal globule [48]. The impact of chain flexibility is currently not considered in this work. However, future work will address the impact of conformational fluctuations in heterochromatin formation.

Of course, the connectivity picture presented here is a simple and straightforward treatment that neglects other factors that will be present within real fibers. Our connectivity is binary; that is, two sites are either connected or not connected. In reality, the histone tails may fluctuate around and thus at some instances two H3K9 sites may be within the needed distance and at other times they may not. The orientation of the HP1 and the H3K9 site also plays a large role in the ability of inter-nucleosomal interactions forming, and some work has looked at how protein-bridging within chromatin fibers can be affected by orientation [45]. Further work investigates how the underlying chromatin structure and genomic region influences the DNA linker length, orientation of the nucleosomes, and structure of the fibers [49–51].

We have used a single type of interaction energy \( J \) to denote all interactions between HP1 molecules. However, the various domains of HP1 may in fact interact with varying degrees of strength, as recent work has shown that not all interactions are mediated through the chromoshadow domain [52]. Furthermore, many molecules interact with HP1 and its variants [2], and the complexes formed may affect the interaction strength. The HP1 bridging interactions are not the only inter-nucleosomal energies, and other forces will contribute to the interactions of nearby bound HP1 and the compaction process [43, 53]. Finally, many other molecules may get in the way of interacting HP1, not the least of which is the DNA and nucleosomes themselves. We did look at the effect of including steric constraints on possible connections. Our preliminary results show that such steric inhibition can reduce the number of connections. However, the introduction of steric does not seem to favor any particular structure and does not reduce the connectivity appreciably for lower values of \( \alpha \).

In this paper, we have shown how the general thermodynamic behavior of HP1 binding to chromatin fiber arrays can be mapped onto a simple lattice. In this state, our model can be equated to a random field Ising model, a well-studied model in statistical physics [25]. For the Ising model, no phase transition occurs in one-dimension, where each site has two neighbors, but a phase transition is possible on the two-dimensional square lattice where each site has four neighbors [54]. However, the presence of a random field, such as the different binding energies at each site due to the methylation profile, raises the critical dimensionality of the system to above two dimensions \((z > 4)\) [25]. This suggests that a stronger degree of connectivity is needed, and thus, the arrangement of the chromatin plays a large role in this process. We will further explore these structural effects in future work. We note that even with a large \( J \) value, the mononucleosomes in figure 2 would not exhibit a phase transition because \( z \) would only be one.

One of our key findings was the enhancement window for HP1 binding only to heterochromatin regions, or regions enriched in the tri-methylation mark. To equate the chemical potential to a concentration for HP1, we assumed that the HP1 remained as a monomer within solution, but recent work has found dimers and even tetramers in solution [36]. As such, the exact range of the value of \( J_z \) needed to operate effectively within the in vivo conditions may differ from those specified from figure 5, but the major qualitative features would remain the same.

While this paper has primarily discussed HP1 binding, we ultimately wish to connect the epigenetic mark of H3K9 tri-methylation to a physical process for gene regulation. As HP1 is known to drive chromatin compaction [9, 10], finding the mechanism for enhanced HP1 binding can then explain increased compaction, and hence repression due to the genes becoming less accessible. As our model has shown, it is the presence of strong cooperative interactions that leads to HP1 to concentrate in highly methylated regions and potentially stabilizes a more compact chromatin structure. Recent simulations have demonstrated that interactions between chromatin segments can lead to compaction of these fibers [55]. In addition, other work has shown that compaction leads to decreased accessibility of the underlying DNA [56].

Our calculations also revealed that strong cooperative interactions lead to a threshold of methylation above which a region would be highly bound by HP1 at nearly all H3K9 sites. This crossover point in the level of local methylation is important biologically because it means that the cell only needs to maintain a high degree of methylation in heterochromatin regions and that there are no critical H3K9 sites that must be methylated. Thus, even if the parent nucleosomes are randomly split between mother and daughter cells, as long as the cell maintains a mechanism to get a region to the appropriate methylation levels, it should become heterochromatin again.

We assume in our averaging over different methylation profiles that there would be no correlations in methylation state between neighboring sites. Much work has suggested
that correlations would exist [57] given the interactions known to be present between the methylation marks, the methyltransferases, and HP1 [2]. These correlations can help define boundaries between heterochromatin and euchromatin regions, and may assist in recovering the methylation profile upon cell division [58].

Finally, while our model is focused on HP1 binding in pericentric heterochromatin, it could easily be extended to other cases which may have similar modes of action. Linker histone H1 is also believed to interact in a similar way to HP1, bridging between neighboring nucleosomes and linker DNA to compact the chromatin [59]. Other proteins such as Polycomb Repressive Group 1 [60], SSN6/Tup1 [61], and Xist in X-chromosome inactivation [62] also could bind to different nucleosomes to aid compaction. Although the parameters such as the interaction distances, the interaction energy, binding energies, and binding locations would change, the phase diagram developed for a different molecule would look similar to results presented in this manuscript.

5. Conclusion

Proper gene expression requires an organized nucleus that segregates the silenced and noncoding regions of the genome. Epigenetic marks like methylation at histone 3 lysine-9 have been identified as critical pieces of inherited information that regulate this process. We have here examined the possibility that heterochromatin protein 1 (HP1), coordinates the segregation and condensation of heterochromatin regions through bridging interactions between nearby bound HP1.

Our model explains how methylation differences between regions can drive differential HP1 accumulation even though the binding energy differences alone cannot explain the degree of enhancement in heterochromatin regions. Cooperative interactions among the HP1 allow for this enhanced accumulation. Binding of HP1, with interactions, can also explain how the compaction process occurs, and how distinct phases are formed.

We find that this binding can be a robust and repeatable process for heterochromatin assembly in successive generations. The small binding energy difference between the unmethylated and tri-methylated states promote stability, and the presence of a phase transition for HP1 binding creates a methylation threshold over which the region will become heterochromatin. Our analyses of intrafiber and interfiber connectivity suggest that H1 tail flexibility may be essential for HP1 to interact across compacted fiber structures. Ultimately, our model connects epigenetic marks to a physical process for their action, the compaction of chromatin fibers into heterochromatin.

Acknowledgments

We thank Aaron Straith and Colin Fuller for helpful discussions with regard to experimental results. This project was supported by an NSF Graduate Research Fellowship, a Hertz Foundation Graduation Fellowship Award, and Award 1305516 from NSF Physics of Living Systems.

References

[34] Ibach H 2006 Physics of Surfaces and Interfaces vol 10 (New York: Springer)