Conditions for positioning of nucleosomes on DNA

Michael Sheinman and Ho-Ryun Chung

Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany

(Received 1 May 2015; published 11 August 2015)

Positioning of nucleosomes along a eukaryotic genome plays an important role in its organization and regulation. There are many different factors affecting the location of nucleosomes. Some can be viewed as preferential binding of a single nucleosome to different locations along the DNA and some as interactions between neighboring nucleosomes. In this study, we analyze positioning of nucleosomes and derive conditions for their good positioning. Using analytic and numerical approaches we find that, if the binding preferences are very weak, an interplay between the interactions and the binding preferences is essential for a good positioning of nucleosomes, especially on correlated energy landscapes. Analyzing the empirical energy landscape, we conclude that good positioning of nucleosomes *in vivo* is possible only if they strongly interact. In this case, our model, predicting long-length-scale fluctuations of nucleosomes' occupancy along the DNA, accounts well for the empirical observations.

I. INTRODUCTION

Our genome is packed and organized by nucleosomes: histone octamers wrapped around by 147 bp of DNA [1,2]. Nucleosomes are in some cases very well positioned, while in others they are rather "smeared" along the DNA molecule [3,4]. Their positioning properties are known to be important in the regulation of gene expression [5–7]. Most transcription factors are not able to bind to nucleosomal DNA, such that sites bound by transcription factors *in vivo* are nucleosome free [8]. This argues that transcription factor binding is dependent on nucleosome occupancy level. Moreover, the location of transcription factor binding sites with respect to nucleosomes is important to fine tune the response of genes to activating signals, as has been revealed for the response to phosphate starvation in budding yeast [9].

There are many factors which determine positioning of nucleosomes along the DNA, and their relative influence is a matter of active debate in the field (see Ref. [10] for a review). The most discussed positioning mechanism is the DNA sequence heterogeneity. It is well known that to wrap DNA around a nucleosome, one needs different energies for different DNA sequences [11,12]. In this case, the debate is only about the importance of DNA sequence preferences, relative to other factors.

An obvious competitor of sequence preferences for nucleosomes' positioning is thermal fluctuations. All measured binding energy differences between different sequences do not exceed a few $k_B T$ even for specially designed strongest binders, which do not exist in known genomes [13–17]. This indicates that, at least in equilibrium, entropic forces are expected to play an important role. It is not entirely clear whether the nucleosomes reach (quasi)equilibrium and how they do it. It was suggested that some active chromatin remodeling enzymes facilitate the equilibration of the nucleosomes by increasing the off rate of the nucleosomes from the DNA [18]. These and others active chromatin remodeling enzymes [19,20] and DNA-binding proteins [21] also affect nucleosomes positioning by actively moving the nucleosomes and by DNA-binding competition. Moreover, ATP-dependent chromatin remodelers and DNA-bound factors, such as RNA polymerase, are believed to be capable of organizing nucleosomes. In particular, they may play a major role *in vivo* [22]. These external interactions, so far, appear to be the most natural explanation for the major differences between *in vitro* and *in vivo* nucleosome occupancy profiles, observed in Ref. [23] and others.

In addition to external positioning signals, there are arguments for and evidences of interactions between neighboring nucleosomes along the DNA [24–35]. A preferential distance between nucleosomes is evident *in vivo* but not *in vitro* (see, e.g., Refs. [22,36]). Biochemical reconstitution of uniform spacing requires the action of ATP-dependent remodelers, giving rise to a density independent spacing at par with statistical positioning [22]. This argues for the idea that the preferential spacing observed *in vivo* is generated by these ATP-dependent remodelers. In principle, a preferential distance can be generated by a combination of strong attraction and steric, hard-core repulsion. In this case, the preferential linker length between two neighboring nucleosomes is expected to be very small.

The interactions are also expected to affect nucleosome positioning. This positioning factor is different from, say, sequence preferences since it depends not on the absolute position of a nucleosome on the DNA, but on the relative position of two nucleosomes: the distance between two neighboring nucleosomes.

In this study, for simplicity, we group all the positioning factors to two types. The first type includes all the factors which determine the position of a single nucleosome. One can characterize it by an effective binding energy landscape of a nucleosome along the DNA molecule, that depends only on the location of the nucleosome along the DNA. The second type corresponds to interactions between neighboring nucleosomes. We characterize this positioning factor by an effective interaction potential, that depends only on the distance between two neighboring nucleosomes. We assume that nucleosomes cannot invade each others DNA territories (although it is not entirely true [37–39], this is not expected to affect significantly the conclusions of this study). In addition, we analyze only the equilibrium distribution of nucleosomes, ignoring nonequilibrium aspects.
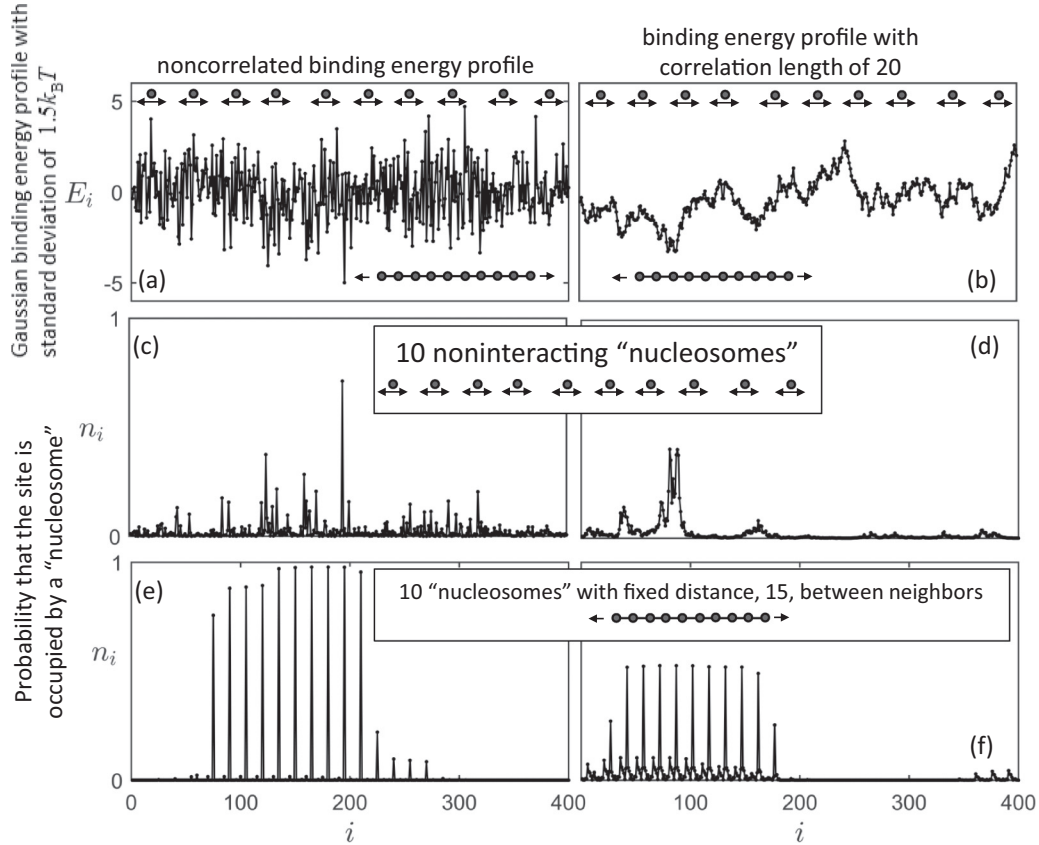
FIG. 1. Illustration of how interactions between nucleosomes improve their positioning along the DNA. In this toy example, "nucleosomes" are merely particles of 1-bp size with disordered binding energy landscape on 400-bp-long, circular "DNA." (a) Binding energies of a single nucleosome are i.i.d random variables with a Gaussian distribution with the standard deviation of $\sigma = 1.5k_BT$. (b) Binding energies of a single nucleosome are normally distributed with the standard deviation of σ and are correlated, such that $\langle E_i E_{i+r} \rangle = \sigma^2 e^{-\frac{r}{r_c}}$, with $r_c = 20$. (c) The probability that the site i is occupied by a nucleosome n_i is plotted vs i for 10 noninteracting nucleosomes located on the binding energy profile from (a). (d) The same as (c) but for the binding energy profile from (b). (e) The same as (c) but for nucleosomes with strong interactions, such that the distance between two neighboring nucleosomes is constrained to 15 bp. (f) The same as (e) but for the binding energy profile from (b).

In general, we address the following question: Within the framework of the above assumptions, what should be the properties of the effective energy landscape and effective interaction potential between neighboring nucleosomes to achieve a good positioning? We analyze energy landscapes with different properties and different interaction potentials and derive conditions leading to good positioning of nucleosomes. Our results suggest that the most crucial parameters for positioning are as follows: standard deviation of the binding energy distribution σ, autocorrelation length of binding energy r_c, and affinity of the interactions κ. We show that strong interactions (large values of κ) between nucleosomes significantly improve their positioning even on an almost flat (small values of σ) and highly correlated (large values of r_c) energy landscape. In this case, if the positioning is good, one expects to observe also large-length-scale fluctuations of nucleosome occupancy along the DNA. We find that such fluctuations can be clearly observed in published experimental data. Moreover, comparing predictions of our model with strong interactions to the experimental study, we find good qualitative (similar long-length-scale fluctuations) and quantitative (similar occupancy profiles on long length scales) agreement. In sum, our results suggest that interactions between neighboring nucleosomes

play an important role in positioning of nucleosomes both locally, on short length scales, and globally, on long length scale. The last type of positioning merely means that there are long regions (more than 10^4 bp) along the genome, which are nucleosomes enriched or depleted.

Before we start with detailed description of the model and its analysis, it might be instructive to illustrate the main message of the paper with a toy example. Consider noncorrelated Gaussian binding energy landscape with standard deviation of $1.5k_BT$, on a circular DNA of length 400 bp [see Fig. 1(a)]. Ten noninteracting "nucleosomes" of size 1 bp cannot be well positioned with such a weak energetic disorder [see Fig. 1(c)]. However, adding very strong interaction between the nucleosomes, such that the distance between them is restricted to 15 bp, one gets good positioning on the same, weak energetic profile [see Fig. 1(e)]. Autocorrelation of an energy [see Fig. 1(b)] makes the positioning even more problematic [see Fig. 1(b)]. However, again, strong interactions between the nucleosomes improve it to a reasonable level [see Fig. 1(d)]. In the paper we will derive, within a quite general set of assumptions, conditions for positioning on uncorrelated and correlated binding energy landscapes. In Figs. 1(e) and 1(f), one can see that, when interactions

between the nucleosomes are exploited for a better positioning, there are long-length-scale fluctuations of the nucleosomes' occupancy along the DNA: there are long enriched and long depleted regions. In the following, we analyze the properties of such regions and demonstrate the relevance of this effect to empirical data.

The structure of the paper is as follows. In Sec. II we formulate the model we use. In Sec. III we define the quantities we use to characterize positioning of nucleosomes on the DNA. In Sec. IV we analyze positioning of a single nucleosome. The purpose of this section is not only didactic, because we use its results in the following. In Sec. V we analyze positioning of many nucleosomes with only hard-core interactions. The purpose of this section is to contrast it to the case with interactions between neighboring nucleosomes and demonstrate the importance of these interactions in Sec. VI. We generalize our results for energy landscapes with an autocorrelation in Sec. VII. In Sec. VIII we discuss relevance of our conclusions to real systems and compare to empirical results. After discussion about tunability and robustness of positioning issues, emerging from our results, in Sec. IX, we summarize in Sec. X. We proceed now with a detailed description of our model.

II. MODEL

We analyze the following lattice based model, which is often used to calculate occupancy of DNA-binding proteins [40]. In a grand-canonical ensemble, on average N nucleosomes are located on a linear DNA of length L, in units of bp, with reflecting boundaries. Note that near the saturation of the DNA by nucleosomes the grand-canonical and canonical ensembles may be different [20,41]. Each nucleosome occupies $W = 147$ bp on the DNA, such that if its leftmost position is bound to a site i, another nucleosome cannot bind with its leftmost position to any of the sites in the interval $[i - W + 1, i + W - 1]$. Due to DNA sequence preferences of nucleosomes, or any other reason, a nucleosome bound with its leftmost position to site i possesses an effective free energy E_i. In addition to this energy there is an interaction energy between two neighboring nucleosomes. Given the distance $r \geqslant 0$ between the leftmost positions of the two nucleosomes, the interaction energy is denoted by $V(r)$. The hard-core interaction is realized by $V(r) = \infty$ for $0 \leqslant r < W$. To obtain the equilibrium properties of the nucleosomes we numerically solve the recursive equation for the partition function [42,43].

Our focus is the following question: What should be the properties of the signal in the one-nucleosome energy profile along the DNA E_i and the interaction between the nucleosomes $V(r)$ to achieve good positioning of nucleosomes on the DNA? In the next section we define this question in more quantitative terms.

III. QUANTITIES OF INTEREST

In this paper, we focus on several quantities which reflect positioning of nucleosomes. Each one of them can be derived from an equilibrium probability of the site i to be covered by the leftmost position of a nucleosome n_i (start site

probabilities). The average number of nucleosomes N is given by

$$N = \sum_{i=1}^{L} n_i. \tag{1}$$

We also define an ordered vector of occupancies n_m^o, such that n_1^o is the occupancy of the most occupied site (site with the highest value of n_i), n_2^o is the occupancy of the second-most occupied site, etc.

For cases when it is *not* important how a base pair along the DNA is covered by a nucleosome (by which part of the nucleosome it is covered) the occupancy function

$$\rho_i = \sum_{j=0}^{W-1} n_{i-j} \tag{2}$$

is of interest. We define the average coverage of the DNA by

$$\rho = \frac{1}{L} \sum_{i=0}^{L} \rho_i = \frac{NW}{L}. \tag{3}$$

There are different ways to define a measure of how well nucleosomes are positioned along the DNA. In this paper, we use a very simple one: given that there are N nucleosomes, we define as $\mathcal{P}$ the fraction of nucleosomes which occupies N (the average number of nucleosomes) most occupied locations along the DNA. Namely,

$$\mathcal{P} = \frac{\sum_{m=1}^{N} n_m^o}{\sum_{m=1}^{L} n_m^o} = \frac{1}{N} \sum_{m=1}^{N} n_m^o. \tag{4}$$

In the case of a single nucleosome $N = 1$ this definition becomes simply the occupation probability of the ground state: the order parameter of, say, the random energy model [44]. For multiple nucleosomes $\mathcal{P}$ can be viewed as a fraction of well positioned nucleosomes. In the case of a perfect positioning $\mathcal{P} = 1$, while in case of no positioning whatsoever, $\mathcal{P} = N/L \ll 1$. The last quantity N/L can be at most $1/W$ for $\rho = 1$.

As we show in the following, for correlated energy landscapes, in some cases nucleosomes are not positioned well on a single bp length scale but are positioned well within a few base pairs. In this case, the value of $\mathcal{P}$ does not characterize fully the positioning goodness. For those cases we exploit the following generalization of $\mathcal{P}$. Denoting the profile around the mth largest values of n_i, as $n_m^o(s)$ we define for odd values of k

$$\mathcal{P}_k = \frac{1}{N} \sum_{m=1}^{N} \sum_{s=-(k-1)/2}^{(k-1)/2} n_m^o(k). \tag{5}$$

The value of $\mathcal{P}_k$ is the measure of positioning given that one does not care about fuzziness on the length scale of k. One can easily see that on the level of one bp resolution $\mathcal{P}_1$ is given by $\mathcal{P}$. However, as we show in the following, on correlated energy landscapes and/or with interaction potential with wide wells the function $\mathcal{P}_k$ can be much more informative than its single-bp resolution value $\mathcal{P}_1 = \mathcal{P}$. We turn now to the consideration of positioning for different scenarios, starting from the simplest one.

IV. POSITIONING OF ONE NUCLEOSOME

It is instructive to consider first the simplest case of a single nucleosome $N = 1$ on a DNA of a certain length L. To position it on the DNA in equilibrium one should generate a nonuniform energy profile along the DNA. Following, we discuss possible energy profiles which can be roughly divided to designed and generic ones.

A. Designed energy landscape

Conceptually, the easiest way to position a nucleosome on a site j is to design an energy landscape such that up to an additive constant $E_{i=j} = -E$ and $E_{i \neq j} = 0$. In this designed DNA case, the positioning measure (4) is given by (we measure all energies in units of $k_B T$)

$$\mathcal{P} = n_1^o = \frac{1}{1 + (L-1)e^{-E}}, \quad (6)$$

such that the nucleosome occupies site i with probability of order one if $E \gtrsim \ln L$. Therefore, having in the arsenal only energies of the order of a few $k_B T$, one can position a single nucleosome only on short sequences of tens to hundreds base pairs, even on the best possible energy landscape.

B. Disordered energy landscape with Gaussian distribution

In the more generic case of a disordered energy landscape, the problem can be mapped to the random energy model [44–47]. In this case, the probability that the leftmost location of the nucleosome is located on site i along the DNA is given by

$$n_i = \frac{e^{-E_i}}{Z}, \quad (7)$$

where the partition function is

$$Z = \sum_{i=1}^{L} e^{-E_i} \quad (8)$$

Therefore, to position a nucleosome on site i one has to fulfill the condition $e^{-E_i} \sim Z$. Consider the case where the energies $\{E_i\}$ are a set of i.i.d random variables with a normal probability distribution with standard deviation σ:

$$\Pr(E_i) = \frac{e^{-\frac{E_i^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}. \quad (9)$$

In this case, the cumulative density distribution for average lowest energy is given by the Lth power of the cumulative distribution for E [48]:

$$\Pr\left(E_1^o < \mathcal{E}\right) = [\Pr(E < \mathcal{E})]^L. \quad (10)$$

Defining a typical value of a distribution as the value where the cumulative distribution drops to e^{-1}, one gets that the lowest energy E_1^o in the large L limit is approximated by

$$\int_{-\infty}^{E_1^o} \Pr(E_i) dE = \frac{1}{L}. \quad (11)$$

One can interpret this equation in the following way. The minimal energy should be such that the number of sites with such an energy, or below it, is 1 on average.

The solution of Eq. (11) is given by

$$E_1^o \simeq \sigma\sqrt{2}\mathrm{erf}^{-1}\left(\frac{2}{L} - 1\right) \simeq -\sigma\sqrt{2\ln L}. \quad (12)$$

In the limit of zero temperature (or, equivalently, an infinite disorder strength σ), the nucleosome will occupy the state with the lowest energy (the ground state) with probability 1. However, for small nonzero temperatures, non-ground states will be partly occupied, such that the occupation probability of the ground state is smaller than one. Consider the m-lowest energy (1-lowest energy means the lowest one, 2-lowest energy is the second lowest energy, etc.) on the DNA E_m^o. Its value can be well approximated using a similar argument to the one above, for the $m = 1$ case: The m-minimal energy should be such that the number of sites with such an energy, or below it, is m on average. This results in

$$\int_{-\infty}^{E_m^o} \Pr(E) dE = \frac{m}{L}. \quad (13)$$

The solution is given by

$$E_m^o \simeq \sigma\sqrt{2}\mathrm{erf}^{-1}\left(\frac{2m}{L} - 1\right) \simeq E_1^o + \frac{\sigma}{\sigma_f} \ln m, \quad (14)$$

where the freezing disorder strength is given by

$$\sigma_f \simeq -\sqrt{2}\mathrm{erf}^{-1}\left(\frac{2m}{L} - 1\right) \simeq \sqrt{2\ln L}. \quad (15)$$

The affinities of the m-lowest state are given by

$$K_m^o = e^{-E_m^o} \simeq e^{-\sqrt{2}\mathrm{erf}^{-1}(\frac{2m}{L}-1)} \simeq \frac{K_1^o}{m^{\sigma/\sigma_f}}. \quad (16)$$

In the low temperature (high disorder) limit $\sigma > \sigma_f$, the occupation of the lowest state is given by

$$\mathcal{P} = n_1^o \simeq \frac{K_1^o}{\sum_m K_m^o} = \frac{1}{\zeta\left(\frac{\sigma}{\sigma_f}\right)} \simeq \left(1 + \frac{2^{1-\frac{\sigma}{\sigma_f}}}{\frac{\sigma}{\sigma_f} - 1}\right)^{-1}, \quad (17)$$

where $\zeta(s) = \sum_{m=1}^{\infty} m^{-s}$ is the Riemann zeta function [49]. Above the freezing point, $\sigma < \sigma_f$, the sum in the equation above diverges. In this case, the annealed approximation of the free energy is valid and the partition function is not widely distributed around its mean value:

$$Z \simeq \langle Z \rangle = Le^{\sigma^2/2}. \quad (18)$$

In this regime, the probability of occupation of any site is given by

$$\mathcal{P} = n_1^o = \frac{e^{-E_m^o}}{Z} \simeq \frac{e^{\sigma\sqrt{2\ln L}}}{Le^{\sigma^2/2}}, \quad (19)$$

such that it vanishes in the thermodynamic limit $L \to \infty$. In genomes of lengths in the range $L = 10^6$–10^9 bp the freezing transition happens in the range $\sigma_f \simeq \sqrt{2\ln L} = 5.3$–$6.4 k_B T$.

In sum, positioning of a single nucleosome is determined by the disorder strength. It is well positioned on DNA of length L (such that $\mathcal{P} \simeq 1$) with energetic Gaussian, uncorrelated disorder with width σ for $\sigma \gg \sqrt{2\ln L}$, and is "smeared" along the DNA (such that $\mathcal{P} \ll 1$) in the opposite limit of weak disorder $\sigma \ll \sqrt{2\ln L}$. In Fig. 2, the above considerations are illustrated. In the Appendix Sec. A 1 we discuss positioning on
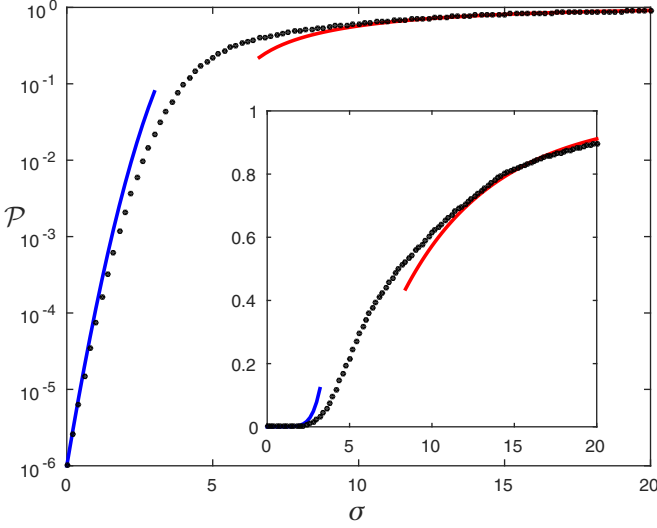
FIG. 2. (Color online) Positioning of a single nucleosome on the DNA with length $L = 10^6$ bp, such that the freezing transition is at $\sigma_f \simeq 5.3 k_B T$. Occupancy of the deepest energy well is plotted vs disorder width. The dots represent numerical simulation: median of 100 realizations of the disordered energy landscape energy. The lines represent the analytic solution: Eq. (17) for $\sigma > \sigma_f$ and Eq. (19) for $\sigma < \sigma_f$. Inset: the same plot in the linear scale.

energy landscape with two non-Gaussian distributions. Now, we turn to discuss positioning of many nucleosomes on the DNA.

V. POSITIONING OF MULTIPLE NUCLEOSOMES WITH ONLY HARD-CORE INTERACTIONS

Here, we analyze positioning of $N \gg 1$ nucleosomes with only hard-core interactions. The study of particles with hard-core repulsion on disordered energy landscapes has a long history and is relevant to many applications (see, e.g., [50]). In the context of protein-DNA binding, such a repulsion between proteins leads to crowding on attractive sites and influences the binding properties [51–54]. Here, we consider positioning of nucleosomes of a finite size $W \geqslant 1$ [55,56] and focus on positioning properties on different energy landscapes.

A. Designed energy landscape

Consider first a designed case when there are N nucleosomes of size W on a DNA of length L and N energy wells of energy $-E$ while the rest of the DNA positions have zero energy. For the best positioning, all the distances have to be larger or equal to W, such that nucleosomes do not have to overlap to occupy all the energy wells. In this case, to position well the nucleosomes one needs $E \gtrsim \ln \frac{L}{N}$. Then, having, say, 10%–70%–90% coverage of the DNA by nucleosomes of length $W = 147$, to position the nucleosomes one needs the energy well to be deeper than 7–4–3 $k_B T$. Moreover, even if the wells are that deep, but the number of nucleosomes differs from the number of energy wells, the positioning is getting worse. To make the positioning more robust, one can make the distance between the wells being random. However, doing this one has to keep the minimal distance between two neighboring wells to
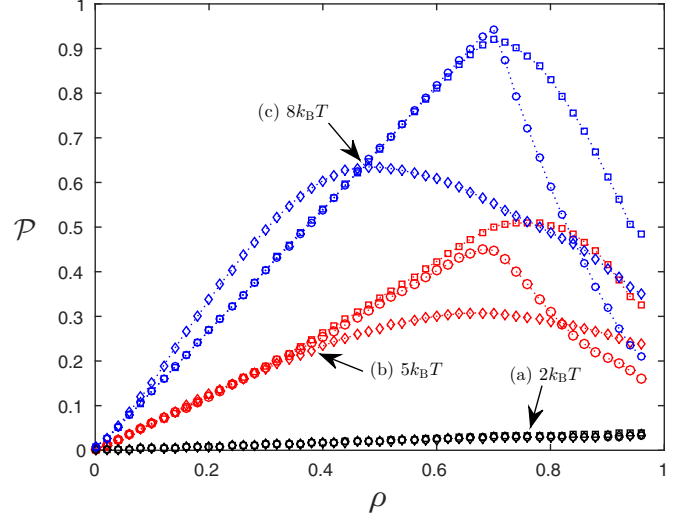

FIG. 3. (Color online) Numerical results for positioning of nucleosomes with only hard-core interactions with $W = 147$ bp on designed energy landscapes of a length of $L = 10^3 \times W$. Positioning parameter is plotted as a function of the coverage fraction $\rho = NW/L$. The energy profile is designed such that neighboring energy wells are separated by distances with the following properties: Each distance is $W + 63$ bp (circles), any distance is a sum of W and a number drawn from a geometric distribution with an average of 63 bp (squares), and each distance is a number drawn from a geometric distribution with an average of $W + 63$ bp (diamonds). The depths of the energy wells relative to the rest positions on the DNA are (a) 2, (b) 5, and (c) 8 $k_B T$ (see arrows in the figure). The lines are to guide the eye.

be W. Otherwise, the nucleosomes, being not able to overlap, spread more on the DNA decreasing the positioning parameter $\mathcal{P}$ (see Fig. 3). In a more generic, disordered case the positioning is more problematic. We turn now to discuss this case.

B. Disordered energy landscape with Gaussian distribution

Consider positioning of N nucleosomes on uncorrelated disordered energy profile normally distributed with standard deviation σ [see Eq. (9)]. In the regime $\sigma \ll \sqrt{2 \ln \frac{L}{N}}$, the nucleosomes are poorly positioned, while in the opposite regime

$$\sigma \gg \sqrt{2 \ln \frac{L}{N}}, \qquad (20)$$

the positioning is good. In sum, having, say, 10%–70%–90% coverage of the DNA by nucleosomes of length $W = 147$, to position the nucleosomes one needs disorder strength σ to be larger than 3.8–2.9–2.4$k_B T$.

The derived requirement for positioning may sound weak. However, in fact it means that, say, for $\rho = 70\%$ and $\sigma = 5 k_B T$ (moderate positioning regime, $\mathcal{P} \simeq 0.6$, as shown in Fig. 4), the typical energy well for a nucleosome is $14 \pm 2 k_B T$ deep [see Eq. (12) with L replaced by L/N], relative to a random DNA sequence. In sum, one can see that without interactions the energy variations required for a good positioning seem to be above the ones measured in experiments [13–17]. In the Appendix Sec. A 2 we discuss positioning of nucleosomes with only hard-core interaction on energy landscape with two non-Gaussian distributions and show that the results in this
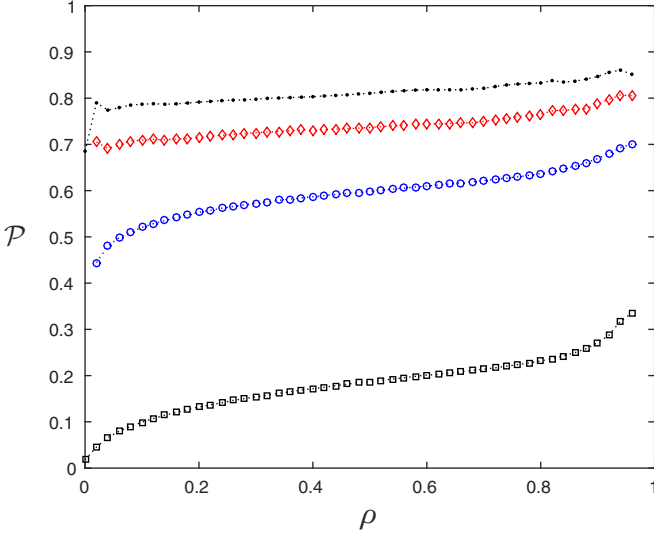
FIG. 4. (Color online) Positioning of nucleosomes with only hard-core interactions. Positioning parameter the DNA with length $L = 10^3 \times W$ and $W = 147$ as a function of the coverage fraction $\rho = NW/L$. The markers represent numerical simulation for $\sigma = 2$ (squares), 5 (circles), 8 (diamonds), and 11 (dots) $k_B T$.

section do not change qualitatively in this case. In the next section we show how interactions between nucleosomes allow to position them with much weaker energetic disorder along the DNA.

VI. POSITIONING OF STRONGLY INTERACTING NUCLEOSOMES

Consider N nucleosomes on a DNA of length L. As we show above, weak sequence specificity cannot position nucleosomes. In this section we analyze positioning of interaction nucleosomes on designed and disordered energy landscapes.

Interaction between neighboring nucleosomes were suggested before as one of the driving forces ordering nucleosomes [32,33,35,57]. Here, we consider for simplicity the minimal model of an interaction with an energy well when two neighboring nucleosomes are at a distance $R - \Delta \leqslant r \leqslant R + \Delta$ from each other. Namely, the interaction potential is of the form

$$e^{-V(r)} = \begin{cases} 0, & r < W \\ e^v = \kappa, & R + \Delta \geqslant r \geqslant R - \Delta \geqslant W \\ 1, & \text{else} \end{cases} \quad (21)$$

and is schematically shown in Fig. 5. For simplicity, we analyze the narrow interaction energy well of only one bp, $\Delta = 0$. Further, we discuss other possible potential functions in general and, in particular, importance of a finite width of the interaction potential well $\Delta > 0$.

For strong enough interactions strength (large values of κ) the nucleosomes gather to clusters, such that in each cluster the distance between the neighboring nucleosomes is R. As we show in the following, this clustering effect can significantly improve positioning of nucleosomes. We demonstrate it first on a designed energy landscape.
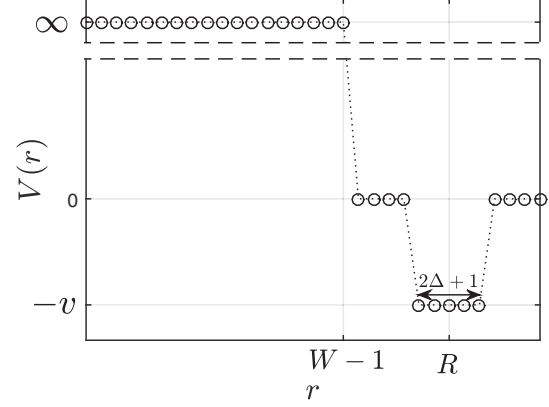

FIG. 5. Illustration of the interaction strength between neighboring nucleosomes used in the paper. $V(r)$ from Eq. (21) is presented. In this particular case, $R = 154$, $W = 147$, and $\Delta = 2$.

A. Designed energy landscape

In order to exploit interactions between nucleosomes and position them on a weak but designed energy landscape, one should have a spatial resonance between the energy wells' distance and the preferable distance between neighboring nucleosomes R. Consider first the case of periodic array of wells with energy $-E$, such that the affinity is $K = e^E$, and set the number of nucleosomes to be the number of wells. Due to interactions, nucleosomes locally crystallize to ordered arrays with nearest-neighbor distance of R. Denoting the average length of a cluster by M, one gets the condition for the positioning of the cluster (and, therefore, all the nucleosomes in the cluster)

$$ME \gg \ln \frac{L}{N/M} \quad (22)$$

or

$$\frac{K^M}{M} \gg \frac{L}{N}. \quad (23)$$

The average number of nucleosomes in a cluster M is given by

$$M = \frac{1}{1 - \Pr(r = R)}, \quad (24)$$

where $\Pr(r = R)$ is the probability that the distance between two neighboring nucleosomes is R. These quantities can be calculated using the self-consistency equation

$$\Pr(r = R) = 1 - \frac{1}{M} = \frac{K\kappa}{K\kappa + \frac{L}{N/M}}. \quad (25)$$

The last equality comes from the fact that a clustered nucleosome contributes energy $-E - v$. In contrast, nonclustered can be in any of $\frac{L}{N/M}$ sites per cluster. The solution of Eq. (25) is given by

$$M = \begin{cases} \sqrt{\kappa K \frac{N}{L}}, & \sqrt{\kappa K \frac{N}{L}} \gg 1 \\ 1, & \sqrt{\kappa K \frac{N}{L}} \ll 1. \end{cases} \quad (26)$$

Therefore, satisfying condition (23), one gets strong improvement of the positioning ($\mathcal{P} \simeq 1$) even with very weak wells

$K \simeq 1$. However, if the wells do not have a good periodicity or their period is different from the preferential distance between the nucleosomes, the interactions do not improve the positioning.

In sum, one can position strongly interacting nucleosomes on a designed energy landscape with shallow energy wells. However, this positioning is not robust to change of nucleosome coverage fraction and requires fine tuning of the distances between energy wells along the DNA. We turn now to discuss positioning on generic energy landscapes where the positioning is not so strong, but is more robust to properties of the energy landscape and the interaction potential.

B. Disordered energy landscape with Gaussian distribution

In the case when the energy along the DNA is a set of independent normally distributed variables with standard deviation σ [see Eq. (9)], local crystallization of interacting nucleosomes also plays a major role in positioning for small values of σ. As shown in Sec. V B, with only hard-core interactions the positioning is possible only when condition (20) holds. Here, we discuss the opposite limit of weak disorder and show that interactions between the nucleosomes can position nucleosomes even in this case. Consider a cluster of crystallized nucleosomes. The effective energy landscape for such a cluster possesses a stronger disorder than for an individual nucleosome. Namely, for a cluster of, say, m nucleosomes the standard deviation of cluster's total energy distribution is $\sqrt{m}\sigma$, where σ, as before, is the energy standard deviation of a single nucleosome. However, for $m \gg 1$, this effective energetic disorder possesses an approximate periodicity of R because by shifting the cluster by this length the total energy of the cluster does not change much. Nevertheless, local crystallization, increasing the effective disorder strength relative to the one for a single nucleosome, causes the positioning of clusters and, therefore, positioning of individual nucleosomes.

Consider a single typical cluster of a size $M \gg 1$. Typical available space for it is given by $\frac{L}{N/M}$, while the standard deviation of its binding energy distribution is given by $\sqrt{M}\sigma$. Thus, the typical minimal energy is given by $\sqrt{M}\sigma\sqrt{2\ln R}$. Therefore, using the same arguments as in Sec. IV B, it will be frozen if

$$\sqrt{M}\sigma \gg \sqrt{2\ln R} \qquad (27)$$

and will be "smeared" on its available space in the opposite limit.

The value of M can be estimated in the following way, using Eq. (24):

$$\Pr(r = R) = 1 - \frac{1}{M} = \frac{\kappa}{\kappa + \frac{L}{N/M}}. \qquad (28)$$

Thus,

$$M = \begin{cases} \sqrt{\kappa \frac{N}{L}}, & \kappa \gg \frac{L}{N} \\ 1, & \kappa \ll \frac{L}{N}. \end{cases} \qquad (29)$$
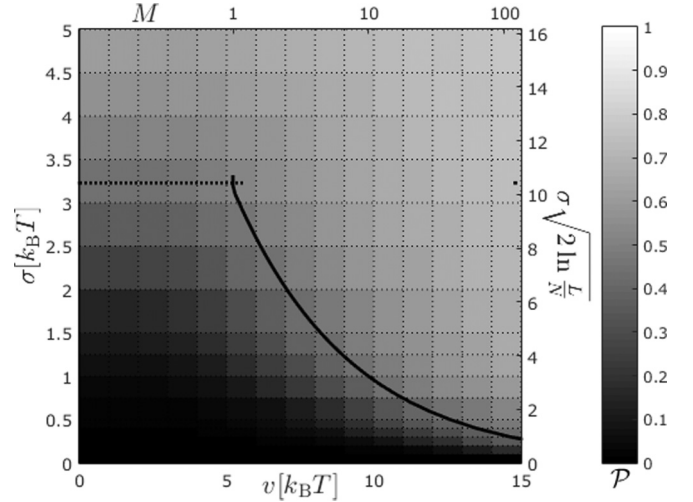

FIG. 6. Positioning of nucleosomes with interactions on Gaussian energy landscape. Positioning parameter $\mathcal{P}$ for the DNA with length $L = 10^4 \times W$ and $W = 147$ for the coverage fraction $\rho = NW/L = 80\%$ is plotted vs disorder strength (left axis) and interaction strength (bottom axis) with preferable distance of $R = 148$ (such a close to W value of R is chosen to make the numerical calculations faster). On the top one can see the average size of the crystallized cluster of nucleosomes, derived from Eq. (29). On the right the typical binding energy of a nucleosome (relative to the average energy) is shown. The lines represent the analytic conditions for a good positioning [Eqs. (20) (dotted line) and (30) (solid line)].

Combining Eqs. (27) and (29), the required strength of interactions to position nucleosomes is given by

$$\kappa \gg \frac{4}{\sigma^4} \frac{W}{\rho} \ln^2 R. \qquad (30)$$

In the case when $\sigma \gg \sqrt{2\ln R}$, condition (30) has to be replaced by $\kappa \gg \frac{L}{N}$. However, in this case condition (20) is satisfied, such that the positioning is possible with only hard-core interactions. Thus, as shown in Fig. 6, if at least one of the conditions (20) and (30) holds the positioning is good, such that $\mathcal{P} \simeq 1$, while otherwise $\mathcal{P} \ll 1$. One can see that strong interactions between nucleosomes are able to improve their positioning. In Appendix A 3 we discuss positioning of strongly interacting nucleosomes on energy landscape with two non-Gaussian distributions and show that the results in this section do not change qualitatively in this case.

The described local clustering of nucleosomes not only improves positioning of nucleosomes but also has other consequences: large-length-scale fluctuations of occupancy. We turn now to discuss this aspect of interactions-assisted positioning of nucleosomes.

C. Large-scale fluctuations of occupancy

Apart from the positioning of nucleosomes on small length scale, there is another feature that is highly influenced by interactions between neighboring nucleosomes: the large-scale fluctuations of occupancy of nucleosomes. One can also interpret this effect as long-length-scale positioning. Without interactions, occupancy ρ_i averaged over thousands of base pairs is not expected to deviate significantly from

its average value ρ. However, if neighboring nucleosomes possess a preferential distance, R which is smaller than the average linker length W/ρ and the nucleosomes are very well positioned, one gets long DNA regions which are enriched by nucleosomes and, therefore, regions depleted with nucleosomes. The correlation length of the occupancy on enriched regions for strongly interacting and very well positioned nucleosomes is given [using Eq. (29)] by

$$RM \sim R\sqrt{\kappa \frac{\rho}{W}}. \tag{31}$$

On this length scale the nucleosomes are enriched and their mean occupancy is given by W/R.

The distance between the clusters of crystallized nucleosomes (with highly depleted occupancy) scales as

$$\frac{L}{N/M} - RM = \left(\frac{W}{\rho} - R\right)\sqrt{\kappa \frac{\rho}{W}}. \tag{32}$$

In sum, strongly interacting and well positioned nucleosomes are expected to exhibit highly fluctuating occupancy on large length scales. In Sec. VIII we discuss how the described above considerations are relevant for more realistic energy landscapes and existing experimental data, but before that we consider effects of another important feature of real systems: an autocorrelation of the energy landscape.

VII. ENERGY LANDSCAPE WITH CORRELATIONS

So far, we discussed random, noncorrelated energy landscapes. However, DNA sequence possesses correlations [58]. Moreover, even on a random DNA sequence an energy landscape is expected to be correlated for distances smaller than 147 because small shifts of nucleosomes along the DNA does not change completely the sequence covered by the nucleosome. This is why, as we discuss in the next section, real energy landscape is expected to possess certain autocorrelation. In this section we discuss how the autocorrelation of the energy landscape affects positioning of nucleosomes.

We analyze the following scenario in this section. The energy landscape is assumed to be Gaussian [see Eq. (9)] with an exponentially decaying autocorrelation, such that

$$\frac{\langle E_i E_{i+r}\rangle}{\sigma^2} = e^{-\frac{r}{r_c}}. \tag{33}$$

Here, $r_c \gg 1$ is the correlation distance, such that for distances much larger than r_c the energies are nor correlated, while for distance much smaller than r_c the variation of energy is much smaller than σ. This model can be mapped to the generalized random energy model [59]. This model possesses multiple phase transitions. The condition for a good positioning on the single bp resolution (we derive in the following) corresponds to the lowest temperature phase transition of that model. In contrast, the condition for a good positioning on a few bp resolution corresponds to high temperature phase transitions of the generalized random energy model.

In this section we assume that for $r_c = 0$ the nucleosomes are well positioned on the DNA. The correlation is additional trouble for positioning and here we derive an additional condition for a good positioning in presence of the autocorrelation, on top of the conditions (20) and (30), for the

noncorrelated energy landscapes. We start from the simplest single-nucleosome case.

Consider a single nucleosome on DNA of length L with a Gaussian energy landscape with standard deviation σ and exponential autocorrelation with correlation length $r_c \gg 1$. Conceptually, we divide the DNA to L/r_c "boxes" of length r_c. In order to position the nucleosomes on the length scale of r_c, in a box with the highest affinity (inside the box we allow the nucleosome to be fuzzy, at this point), one needs to satisfy $\sigma \gg \sqrt{2\ln\frac{L}{r_c}}$. This condition is satisfied because we assume here that without autocorrelation the positioning is good and, therefore, $\sigma \gg \sqrt{2\ln L}$. Thus, the problem is the positioning of the nucleosome within the box.

Within the box all the energies are highly correlated. A way to generate such a correlated energy landscape is to set [60]

$$E_{i+1} = e^{-\frac{1}{r_c}} E_i + \sqrt{1 - e^{-\frac{2}{r_c}}} G_i, \tag{34}$$

where G_i is an uncorrelated set of Gaussian random variables with standard deviation σ. Thus, the standard deviation of energies E_i, if i is in the range much smaller than r_c, is given by $\frac{\sigma}{\sqrt{2r_c}}$. With such a standard deviation, to position a nucleosome in a box of size r_c with a single bp resolution, one needs $\frac{\sigma}{\sqrt{2r_c}} \gg \sqrt{2\ln r_c}$ or

$$\sigma \gg \sqrt{4r_c \ln r_c}. \tag{35}$$

This is a strong constraint on the positioning. Even for $r_c = 5$ bp the positioning is bad unless σ is much larger than $6k_B T$.

Condition (35) remains the same also for the case of noninteracting nucleosomes or nucleosomes with only hard-core interactions. This is because (35) does not depend on the length of DNA per nucleosome but only on the correlation distance of the energy landscape. This makes positioning of noninteraction (or interacting with only hard-core repulsion) extremely problematic. In the next section we show that the value of r_c is, at least, tens of base pairs. For such a correlated energy landscape, the positioning condition (35) is not expected to be satisfied.

For interacting nucleosomes, the standard deviation of the effective energy landscape is given by $\sqrt{M}\frac{\sigma}{\sqrt{2r_c}}$, where M is the average number of nucleosomes in a crystallized cluster and given by Eq. (29). The positioning condition for strongly interacting nucleosomes is given by

$$\sigma \gg \sqrt{\frac{4r_c}{M} \ln r_c} \tag{36}$$

or

$$\kappa \gg \frac{W}{\rho} \frac{16r_c^2}{\sigma^4} \ln^2 r_c. \tag{37}$$

In Figs. 7 and 8 one can see the comparison of condition (36) or (37) to numerical results. The obtained results imply that even for short-range autocorrelation of the energy profile with a realistic value of σ, it is impossible to position nucleosomes on a single bp resolution without strong interactions between them.

However, the positioning on correlated energy landscape is easier if one allows the nucleosome to be fuzzy within a few bp and still consider this as a good positioning. In order to see
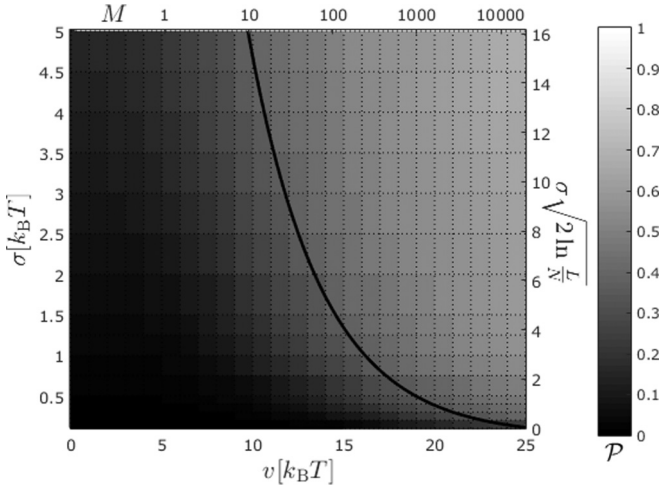
FIG. 7. Positioning of nucleosomes with interactions on Gaussian energy landscape with an exponentially decaying autocorrelation with correlation coefficient $r_c = 20$. Positioning parameter $\mathcal{P}$ for the DNA with length $L = 10^4 \times W$ and $W = 147$ for the coverage fraction $\rho = NW/L = 80\%$ is plotted vs disorder strength (left axis) and interaction strength (bottom axis) with preferable distance of $R = 148$. On the top one can see the average size of the crystallized cluster of nucleosomes, derived from Eq. (29). On the right the typical binding energy of a nucleosome (relative to the average energy) is shown. The line represents the analytic conditions for a good positioning [Eq. (36) or (37)].
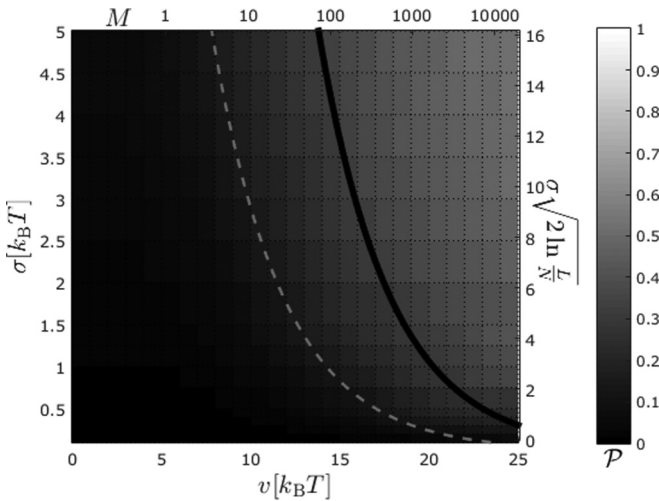

FIG. 8. Positioning of nucleosomes with interactions on Gaussian energy landscape with an exponentially decaying autocorrelation with correlation coefficient $r_c = 100$. Positioning parameter $\mathcal{P}$ for the DNA with length $L = 10^4 \times W$ and $W = 147$ for the coverage fraction $\rho = NW/L = 80\%$ is plotted vs disorder strength (left axis) and interaction strength (bottom axis) with preferable distance of $R = 148$. On the top one can see the average size of the crystallized cluster of nucleosomes, derived from Eq. (29). On the right the typical binding energy of a nucleosome (relative to the average energy) is shown. The solid line represents the analytic conditions for a good positioning on a resolution of $k = 1$ bp [Eq. (36) or (37)]. The dashed line represents the analytic conditions for a good positioning on a resolution of $k = 9$ bp [Eq. (38) or (39)].

it, we exploit the positioning function, defined in Eq. (5). This function $\mathcal{P}_k$ characterizes the positioning within the resolution of k base pairs. In Fig. 9 one can see that in some cases even when the single bp resolution positioning is bad, $\mathcal{P} = \mathcal{P}_1 \ll 1$, the positioning within $k = 3, 5, \ldots$ is significantly better. The condition for $\mathcal{P}_k$ to be of the order one is equivalent to condition (36) or (37) with r_c replaced by r_c/k. Namely, the condition for a good positioning within k bp is given by

$$\sigma \gg \sqrt{\frac{4r_c}{Mk} \ln \frac{r_c}{k}} \tag{38}$$

or, equivalently,

$$\kappa \gg \frac{W}{\rho} \frac{16\left(\frac{r_c}{k}\right)^2}{\sigma^4} \ln^2 \frac{r_c}{k}. \tag{39}$$

One can see in Fig. 8 that condition (38) [or (39)] can be much weaker than (36) [or (37)]. In the next section we study empirical landscape for which σ is roughly $1.5 k_B T$ and r_c is roughly 100 bp. Thus, if the interaction between nucleosomes is strong enough to crystallize them to clusters of size $M = 10$–100 nucleosomes, one would expect to see bad positioning on the level of a single bp resolution with $\mathcal{P}_1 = 0.1$–0.25 [(36) does not hold] but within $k = 9$ bp the nucleosomes are positioned significantly better [(38) does hold] with $\mathcal{P}_9 = 0.5$–0.8 (see Fig. 9). In the next section we discuss positioning properties on empirical energy landscape and, in general, relevance of the above considerations to real systems.

VIII. RELEVANCE TO EMPIRICAL RESULTS

So far, we discussed positioning on artificial energy landscapes and dissected the phase diagram to different regimes. An obvious question to ask now is the following: Where is the real system on the phase diagram? In this section we try to get insight into this problem. To do so, we calculate an energy landscape using a model in Ref. [23]. In fact, there are many different models for a binding energy of a nucleosome to a given sequence (examples include [23,61–63], for a review see, e.g., [64,65]). We exploit only one of them, from Ref. [23], because we are not interested in predicting locations of nucleosomes on some piece of DNA but in general properties of positioning of nucleosomes. In particular, our goal in this section is to validate our results on artificial energy landscapes and show their relevance to more realistic energy landscapes.

We start with contrasting the presented, artificial energy landscapes and the one generated using the model in Ref. [23] on the genome of *S. cerevisiae*. The first thing to note is that the distribution of the binding energies on the *S. cerevisiae* genome is close to, but deviates from, a Gaussian (see Fig. 10). The standard deviation of the energy landscape is given by $1.6 k_B T$. This seems to be an overestimate because an energy landscape with $\sigma = 1.6 k_B T$ on a genome of length $L \simeq 10^7$ bp and correlation length of $r_c \simeq 100$ bp possesses the strongest binder with energy $-7 k_B T$, relative to the energy of a random sequence, while even the strongest artificial binder, 601, has the energy around 2–5$k_B T$ [13]. Interestingly, as shown in Fig. 10, the same model on a randomly shuffled *S. cerevisiae*
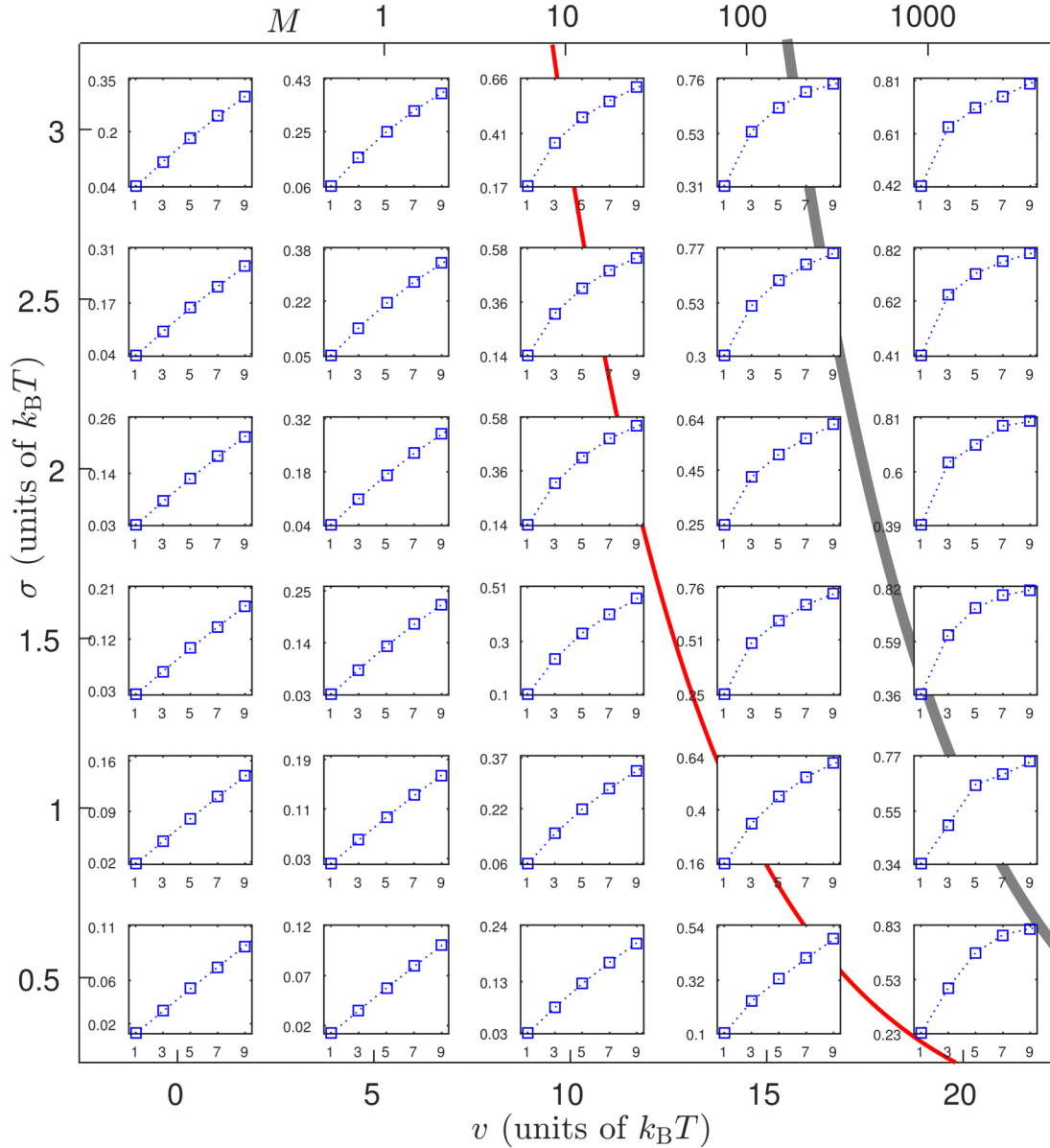
FIG. 9. (Color online) Positioning of nucleosomes with interactions on Gaussian energy landscape with an exponentially decaying autocorrelation with correlation coefficient $r_c = 100$. Positioning parameter $\mathcal{P}_k$ for the DNA with length $L = 10^4 \times W$ and $W = 147$ for the coverage fraction $\rho = NW/L = 80\%$ is plotted vs k for different disorder strengths (left axis) and interaction strengths (bottom axis) with preferable distance of $R = 148$. The line represents the analytic conditions for a good positioning on a single bp resoltion $k = 1$ [Eq. (36) or (37)].

genome yields significantly narrower energy landscape, well fitted by a Gaussian with $\sigma = 1.24k_{\mathrm{B}}T$. The energy landscape, generated by the model in Ref. [62], predicts even narrower energy landscape, with more realistic standard deviation of $0.8k_{\mathrm{B}}T$ (also shown in Fig. 10), such that, if $L = 10^7$ bp and $r_c \simeq 100$ bp, the strongest binder possesses a more realistic energy of $-3.5k_{\mathrm{B}}T$, close to energies of the strongest natural binders, found in Ref. [13]. Following, we use the energy landscape from Ref. [23] to demonstrate that, even with such a clear overestimation of the energetic landscape roughness, one needs interactions in order to get a good positioning. We do not use the energy landscape to predict locations of nucleosomes along the genome but focus on general properties of positioning, such as global positioning parameter and

long-length-scale fluctuations of occupancy level. We expect our main conclusions (in contrast to detailed predictions of locations of nucleosomes along the genome) to be valid for other models for the energy landscapes with a reasonable roughness.

Energy landscape from Ref. [23] with such a narrow distribution even with no autocorrelation is not expected to position well the nucleosomes without strong interactions [see Eq. (20) and Fig. 4]. Interaction between nucleosomes can improve the positioning [see Eq. (30) and Fig. 6].

However, as is expected, the calculated energy landscape possesses certain autocorrelation. Some part of this autocorrelation is because shifting a nucleosome a few bp does not change entirely the bound sequence. This sort of
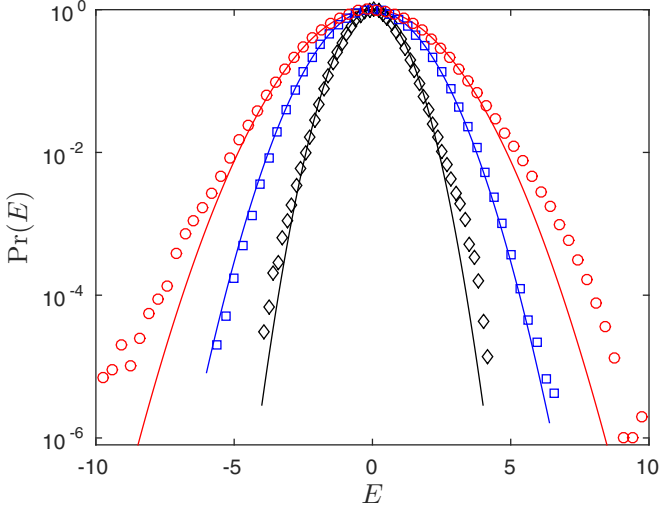
FIG. 10. (Color online) Distribution of the binding energy of a nucleosome on a *S. cerevisiae* (circles) and randomly shuffled *S. cerevisiae* genome (squares), calculated using the model in Ref. [23]. The standard deviation is equal to $1.6k_BT$ for *S. cerevisiae* and $1.24k_BT$ for randomly shuffled *S. cerevisiae* genome. The diamonds represent the distribution of the energy landscape calculated in Ref. [62] for *S. cerevisiae* genome with a standard deviation of $0.8k_BT$. The lines are Gaussian fits with zero mean and respective standard deviations.

autocorrelation, for distances smaller than 147 bp, exists even on a randomly shuffled genome (see Fig. 11). One can clearly see the periodic oscillations with an approximate 10 bp period [13,66].

On top of that, due to some sequence correlation along the real, nonshuffled, *S. cerevisiae* genome, the autocorrelation of binding energy persists even for distances larger than 147 bp (see Fig. 11). One reason for such a long-scale autocorrelation is that the model in Ref. [23] may be biased by GC content the
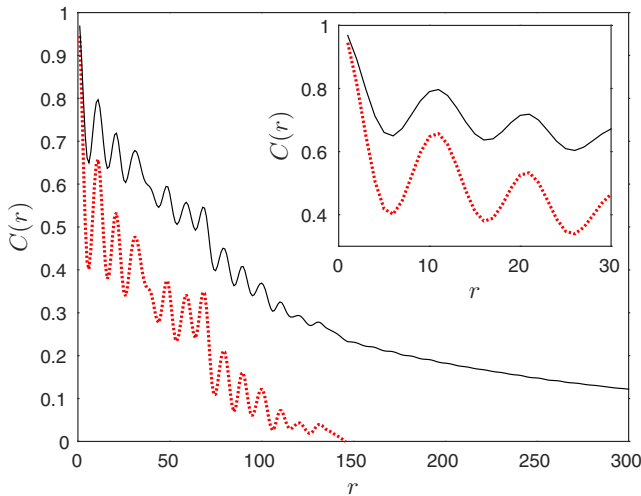

FIG. 11. (Color online) Autocorrelation of the binding energy of a nucleosome on a *S. cerevisiae* (black solid line) and randomly shuffled *S. cerevisiae* genome (dotted, red line), calculated using the model in Ref. [23]. The inset is a zoom-in on the main plot.

percentage of bases on a DNA molecule that are either guanine or cytosine [67] and GC content possesses significant autocorrelation along the *S. cerevisiae* genome [68] (see Fig. 12). The autocorrelation function clearly deviates from the simple exponential decay, which we assumed in our theoretical considerations above. Roughly, the correlation distance is close to $r_c \simeq$ 100 bp, making the positioning much more problematic, relative to uncorrelated energy landscape with the same standard deviation. This is demonstrated in Figs. 13(a) and 13(b). On the left in each panel we plot the average profile of N peaks of n_i. Namely, we take N highest peaks and average n_i around them. The width of such a profile indicates the resolution of the positioning, a width of an average peak. On the right of each panel we plot the positioning parameter $\mathcal{P}_k$ [see Eq. (5)] versus k.

On such an autocorrelated energy landscape, with only hard-core interactions, $v = 0$, nucleosomes are not well positioned on a single bp resolution (see a typical n_i profile in Fig. 12). Namely, the single bp positioning parameter $\mathcal{P} = 0.06$ (calculated on the first chromosome of *S. cerevisiae*) is much smaller than 1. With only hard-core interactions, the peaks in n_i are not only low, but also wide. In fact, the positioning function $\mathcal{P}_k$, shown in Fig. 13(a), demonstrates that the nucleosomes are "fuzzy" and poorly positioned even on the resolution of 10 bp.

The absence of good positioning can be partially attributed to strong autocorrelation of energy because the distribution of uncorrelated Gaussian energy landscape with $\sigma = 1.6k_BT$ results in much better positioning, as is discussed in Sec. V B. Indeed, calculation of positioning of nucleosomes with only hard-core interaction on randomly shuffled energy landscape, calculated using the model in Ref. [23], results in narrow (1 bp) peaks of n_i of an average height of $\mathcal{P}_1 = \mathcal{P} = 0.18$ [see Fig. 13(b)].

In contrast to the case with only hard-core interactions, the nucleosomes can be positioned much better in the presence of interactions between neighboring nucleosomes. For example, for the interaction strength of $v = 9$ and the preferable distance $R = 154$, the peaks of n_i are much higher, such that the positioning parameter is $\mathcal{P}_1 = 0.3$ [see Figs. 12 and 13(c)]. The width of the peaks is 3–5 bp, such that the positioning function is $\mathcal{P}_k \sim 0.45$ for $k \geqslant 3$, as can be seen in Fig. 13(c).

Can one position a nucleosome with a more realistic interaction potential? A reasonable choice seems to be the one used in Ref. [33] to fit qualitatively the $10n + 5$ (or, sometimes, $10.6n + 8$ [69]) periodicity found in many works, starting from Ref. [70]. In order to verify that our results do not depend qualitatively on the precise form of the interaction potential, we used the same form as in Ref. [33], but with higher prefactor, i.e., 12 instead of 5 (otherwise the potential is too weak to position nucleosomes) and with a cutoff of 180 bp for the computational purposes:

$$V(r) = \begin{cases} \infty, & r < W \\ 12k_BT \cos\left(\frac{2\pi(r-W)}{10\,\text{bp}}\right)e^{-\frac{r-W}{50\,\text{bp}}}, & W \leqslant r \leqslant 180 \\ 0, & r > 180. \end{cases} \quad (40)$$

In Fig. 13(d) one can see that the interaction potential in Eq. (40) is able to position nucleosomes on the energy landscape generated using the model in Ref. [23] within the resolution of 3–5 bp. In Sec. IX we analyze in more detail
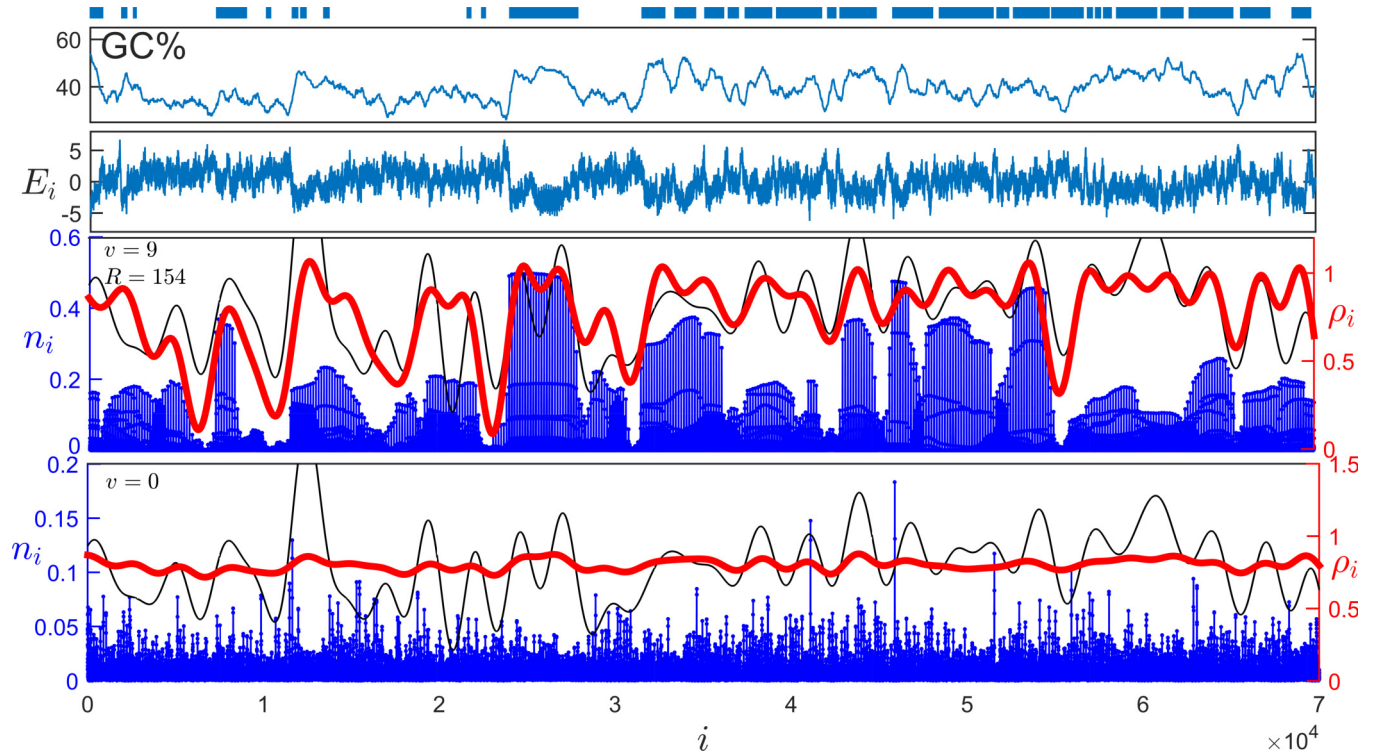
FIG. 12. (Color online) Analysis of the first 7×10^4 bp of *S. cerevisiae*'s first chromosome. The lines on top of the figure depict exons. In panel number (I) the GC content is plotted. (II) Binding energy profile (in units of $k_B T$) calculated using model from Ref. [23]. The additive constant term is such that the mean energy along the chromosome is zero. (III) Calculation of the nucleosome distribution for interacting nucleosomes with $R = 154$ and $v = 9$. The chemical potential is such that $\rho = 0.8$. The lines with dots represent n_i values, while the thick, red line represents occupancy level. The thin black line is the occupancy from the data in Ref. [71]. Both the occupancies are smoothed, such that only 100 lowest Fourier modes are presented. (IV) The same as for the (III) panel but with only hard-core interactions $v = 0$.

robustness and tunability of positioning to different properties of the interacting potential.

In sum, these results indicate that good positioning of nucleosomes is possible even on a realistic energy landscape with narrowly distributed (small σ) and highly autocorrelated (large r_c) energies, provided strong enough interactions between them.

Beyond the positioning of nucleosomes on small length scales, as we described in Sec. VIC, strong interactions between neighboring nucleosomes change DNA occupancy by nucleosomes on large length scales. We turn now to discuss how the theoretical predictions in Sec. VIC are relevant for realistic scenarios.

A. Large-scale fluctuations of occupancy

We start with nucleosomes with only hard-core interactions, as a reference case. In this case, since the energy profile is not well correlated for long distances, the occupancy does not fluctuate significantly on large length scales (above a few nucleosome repeat lengths), as in Figs. 12 and 14. However, as discussed in Sec. VIC, interactions locally crystallize nucleosomes, inducing large-scale fluctuations in occupancy. In Figs. 12 (zoom on first 7×10^4 bp of chromosome I of *S. cerevisiae*) and 14 (whole chromosome I of *S. cerevisiae*), one can see that interacting nucleosomes are distributed nonuniformly along the DNA, in contrast to nucleosomes with only hard-core interactions. Interestingly, this sort of large-length-scale

nonuniformity one also observes on the single-bp resolution data [71] (see Figs. 12 and 14). Moreover, the calculated occupancy seems to follow quite consistently the experimental one on different length scales. This is especially surprising because the data from [71] are based on chemical cleavage, while the model used by us to calculate the energy profile along the DNA was derived based on MNase digestion [23].

These results indicate that the interactions between the nucleosomes help to position them not only on the short length scales (a few bp), but also on the long length scales, inducing large long-scale fluctuations in nucleosomes' occupancy. In other words, strong interactions between the nucleosomes naturally yield long nucleosome diluted and enriched regions along the genome.

We summarize this section with the following conclusions. On a realistic energy landscape, nucleosomes are much better localized in presence of strong interactions. In this case, the calculated results, predicting large-length-scale occupancy fluctuations, agree qualitatively and, surprisingly, quantitatively with the experimental data. We turn now to a more detailed analysis of how properties of the interaction potential affect distribution of nucleosomes along the genome.

IX. TUNABILITY AND ROBUSTNESS OF THE POSITIONING

Several relevant questions are still to be answered. How robust are the obtained results for the made assumptions? What
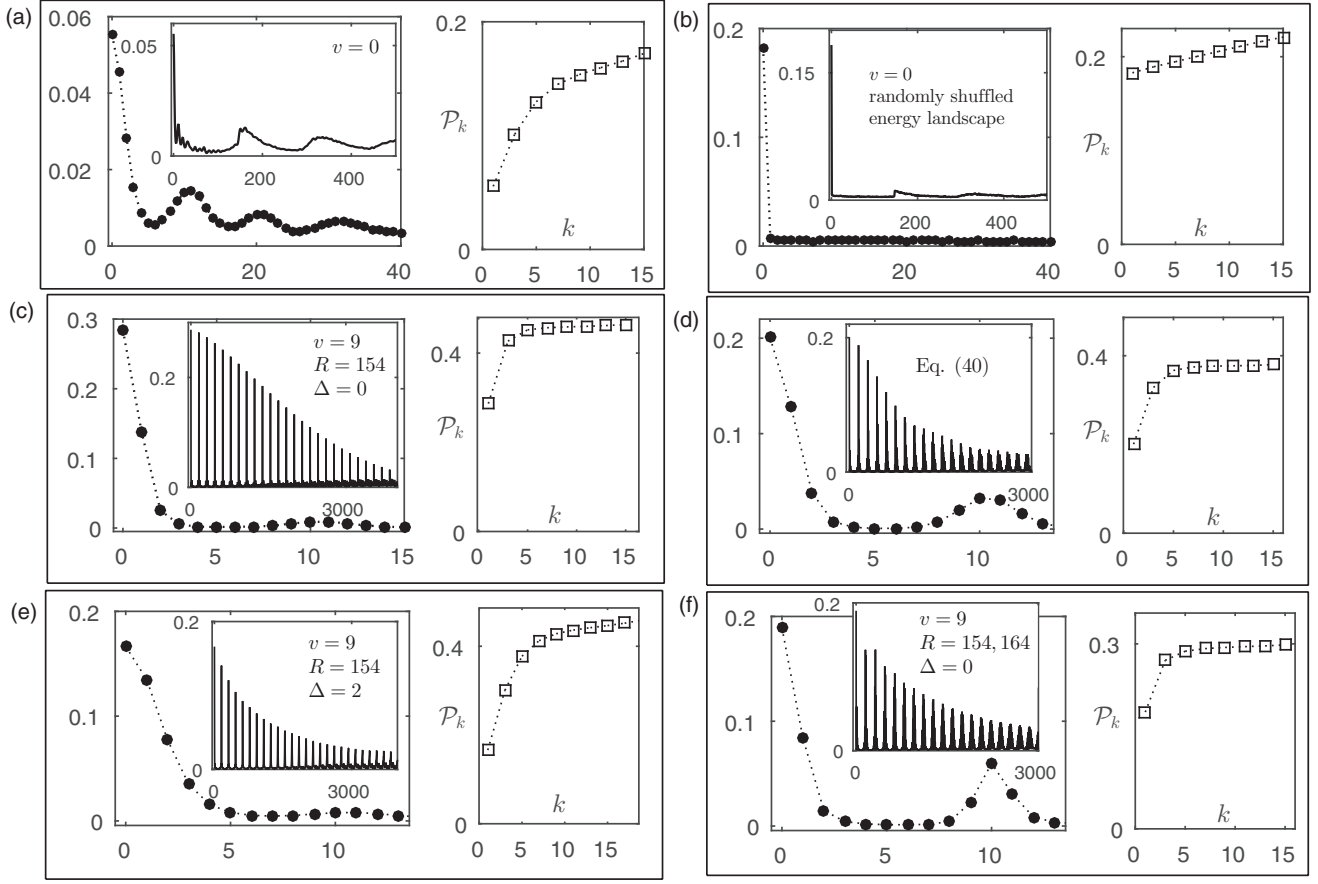
FIG. 13. Positioning goodness in different cases. On the left in each panel the average profile of N peaks of n_i is plotted. Namely, we take N highest peaks $i_1, i_2, \ldots, i_N$ and average n_i around them: $\widetilde{n}_i = \frac{1}{N}(n_{i-i_1} + n_{i-i_2} + \ldots + n_{i-i_N})$. On the y axis is plotted the symmetrized function $1/2(\widetilde{n}_i + \widetilde{n}_{-i})$ vs i. The width of such a profile indicates the resolution of the positioning: a width of an average peak. On the right of each panel we plot the positioning parameter $\mathcal{P}_k$ [see Eq. (5)] vs k. Insets are zoom-out of the main plots. The right plots are the positioning function $\mathcal{P}_k$ vs k. (a) Energy landscape from Ref. [23] with only hard-core interactions between the nucleosomes $v = 0$. (b) The same as in (a) but the energy landscape (not the sequence) is randomly shuffled. (c) The same as in (a) but nucleosomes interact with $v = 9$, $R = 154$, $\Delta = 0$. (d) The same as in (a) but nucleosomes interact with the potential from Eq. (40). (e) The same as in (c) but nucleosomes interact with $\Delta = 2$. (f) The same as in (a) but nucleosomes interact with two well potentials $R = 154$ and 164.

if the interaction potential possesses a certain width [$\Delta > 1$ in Eq. (21)]? How robust are the positions of nucleosomes to a change of parameters, such as the strength of the interactions v, preferable distance R, number of wells, and the width of the interactions potential Δ. Can one "tune" the distribution of nucleosomes along the DNA changing (locally or globally) the parameters of the interaction potential?

We addressed some of these issues above, considering more realistic potential between the nucleosomes, in Eq. (40), and found that the conclusions are quite robust to a particular form of the potential. In this section we address these questions more systematically. We focus here on the energy profile from the previous section, calculated using the model from Ref. [23] and take as a starting point the potential with a single well with $R = 154$, $v = 9$, $\Delta = 0$. In this case, the positioning is reasonably good [see Figs. 12 and 13(c)].

We start by changing the width of the interaction potential Δ. Before we always [except from analyzing the potential form from Eq. (40)] assumed that the interaction potential is sharp to the level of a single bp, $\Delta = 0$. Of course, this assumption is not realistic and real effective interaction

potentials probably possess a finite width and more than one energy well [24,32,72]. However, changing the width to a few base pairs does not change qualitatively the results. Taking $\Delta = 2$ the positioning remains good on the length scale of $2\Delta + 1 = 5$, such that the typical peak of n_i has a height of $\mathcal{P} \simeq 0.3$ and width of 5 bp, as shown in Fig. 13(e). As one can see in Fig. SI1 in the Supplemental Material [73], the occupancy on large length scales does not change much as one tunes the value of Δ. Even on the level of a single bp resolution the Pearson correlation coefficient of ρ_i for a potential with a width of 1 bp ($\Delta = 0$) and 5 bp ($\Delta = 2$) is 0.87. The precise locations of nucleosomes do change, however. Looking on the $N = \rho L/W$ largest values of n_i for both values of Δ we observe an overlap of 18%. This value makes sense because we expect that, upon changing the width of the potential well from 1 to 5 bp, 20% of the nucleosomes will remain in their positions and 80% will move ± 2 bp, within the new potential well. In sum, widening of the potential width does not change the distribution of nucleosomes on a large scale, but makes their position uncertain on the length scale of $2\Delta + 1$.
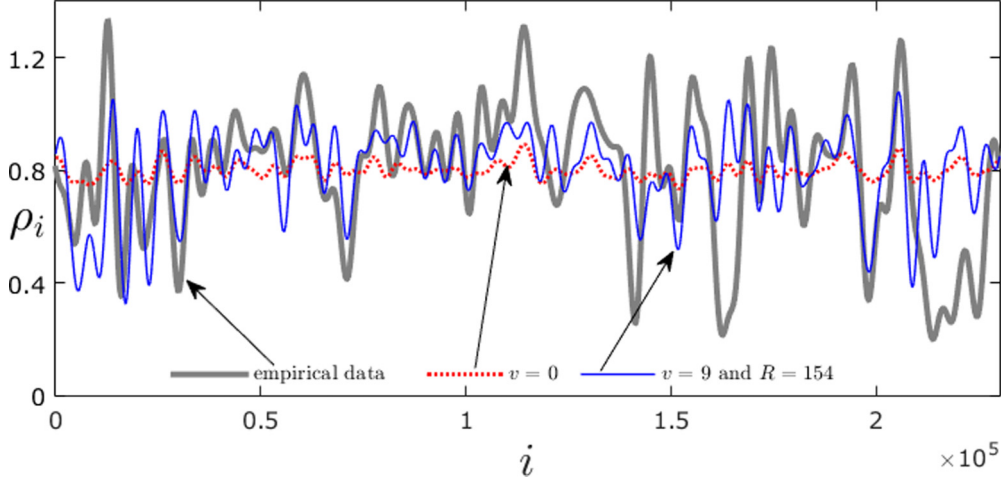
FIG. 14. (Color online) Occupancy of nucleosomes on the global scale of the whole first chromosome of *S. cerevisiae*. Thin, solid line represents calculation of the nucleosome distribution for interacting nucleosomes with $R = 154$ and $v = 9$. Thick, gray line is the occupancy from the data in Ref. [71]. The calculated occupancy for nucleosomes with only hard-core interactions is represented by the dotted, weakly fluctuating line. The chemical potential set for calculations is such that $\rho = 0.8$. All the occupancies are smoothed, such that only 50 lowest Fourier modes are presented.

We add now one more well to the interaction potential at 10 bp from the first one, $R = 154 + 10 = 164$. The positioning gets slightly worse, such that the typical peak of n_i has a height of $\mathcal{P} \simeq 0.2$ and width of 3 bp, as shown in Fig. 13(f). As one can see in Fig. SI2 in the Supplemental Material [73], the occupancy on large length scales does not change much if we add another well in the potential function. Even on the level of a single bp resolution, the Pearson correlation coefficient of ρ_i for a potential with one and two wells is 0.75. The precise locations of nucleosomes do change, dramatically, however. Looking on the $N = \rho L/W$ largest values of n_i for both values of Δ we observe an overlap of only 3%. In sum, another well in the potential width does not change the distribution of nucleosomes on a large scale, but does change their positions on small length scale.

The depth of the potential well does affect the goodness of positioning and has to be strong enough to have any effect. However, once a reasonable positioning is achieved, its value does not change things much. As one can see in Fig. SI3 in the Supplemental Material [73], the occupancy on a large length scale does not change much as one tunes the value of v from 9 to 11. Even on the level of a single bp resolution, the Pearson correlation coefficient of ρ_i for a potential with one and two wells is 0.94. Even the precise locations of nucleosomes do not change, dramatically. Looking on the $N = \rho L/W$ largest values of n_i for both values of v we observe an overlap of 67%.

As one would expect, the position of the well, R, does not affect significantly the positioning properties and distribution of the nucleosomes on a large scale. As one can see in Fig. SI4 in [73], the occupancy on large length scales does not change much as one tunes the value of R from 154 to 160. On the level of a single bp resolution, the Pearson correlation coefficient of ρ_i for a potential with $R = 154$ and 160 is 0.62. However, precise positioning of nucleosomes is very sensitive to the value of R: $N = \rho L/W$ locations with the highest values of n_i for the two cases of $R = 154$ and 160 possess an overlap of only 1%.

We conclude that, in principle, a cell, tuning the preferable distance between nucleosomes, can control their distribution along some part of DNA without changing significantly large-scale properties, such as an average positioning goodness and large-length-scale occupancy.

X. SUMMARY

In this article we focus on goodness of positioning of nucleosomes on the DNA. We make several simplifying assumptions. We assume that we can take into account all the positioning factors by having an effective energy landscape and an effective interaction potential between neighboring nucleosomes. We ignore that nucleosomes can invade each other's DNA territories. In addition, we analyze only the equilibrium distribution of nucleosomes, ignoring very probable nonequilibrium aspects of nucleosome positioning. However, even within this simplified framework we clarify a few aspects of nucleosome positioning, which do not seem to depend on these details.

Looking on a generic energy landscape with some energy distribution width, some energy typical autocorrelation distance, and some interaction potential between neighboring nucleosomes, we derive conditions for a good positioning. We briefly summarize the conditions in the following paragraph.

Assuming that neighboring nucleosomes possess a preferable distance with an affinity $\kappa \gg 1$, relative to other distances, the number of locally crystallized nucleosomes $M \gg 1$ is given by Eq. (29), where N is the average number of nucleosomes and L is the length of the DNA. The positioning is expected to be good on an uncorrelated Gaussian disorder with standard deviation σ if condition Eq. (27) holds. On a correlated energy landscape with a certain correlation distance r_c, the positioning conditioning depends on the required resolution. For a good positioning within k bp it is given by Eq. (38).

Importantly, without strong interactions $\kappa \simeq 1$, the conditions do not seem to hold for realistic parameters, indicating an important role of effective interactions between the

nucleosomes in their positioning. If the positioning, as we suggest, is controlled by interactions, one expects to see long-length-scale fluctuations of nucleosome occupancy. This conclusion agrees with empirical data on occupancy of nucleosomes. Moreover, the derived large-length-scale occupancy profile, derived from an effective energy landscape and interaction potential which is sufficient to position the nucleosomes, is similar to the empirical one. We also analyze the robustness of positioning to parameters of the model. The parameters can vary with time and be different on different parts of the genome. In fact, as we demonstrate, tuning some parameters one can dramatically change the distribution of nucleosomes. In sum, our study emphasizes an important role of interaction between the nucleosomes and indicates the range of parameters needed for it. We expect this knowledge to be important for better understanding of organization of our epigenome.

APPENDIX: NON-GAUSSIAN ENERGY LANDSCAPES

The distribution of the binding energies along the DNA does not have to be Gaussian. Here, we analyze two other possible scenarios for the energy landscapes.

1. Positioning of one nucleosome

Here, we discuss positioning of a single nucleosome on up-exponential and down-exponential energy landscapes.

a. Disordered energy landscape with down-exponential distribution

Consider the scenario with what we denote as *down*-exponential distribution:

$$\Pr(E_i) = \frac{1}{\mathcal{E}} e^{\frac{E_i}{\mathcal{E}}} \; ; \; -\infty < E_i \leqslant 0 \; ; \; \mathcal{E} > 0. \qquad \text{(A1)}$$

In this case the typical minimal energy can be estimated using Eq. (11) to be

$$E_1^{\text{o}} \simeq -\mathcal{E} \ln L. \qquad \text{(A2)}$$

The partition function can be estimated separately in two regimes (two phases in the thermodynamic limit): in the nonfrozen regime $\mathcal{E} \ll 1$ the partition function is given by

$$Z \simeq L \int_{-\infty}^{0} e^{-E} \Pr(E) dE = \frac{L}{1 - \mathcal{E}}. \qquad \text{(A3)}$$

The positioning parameter in this case is given by

$$\mathcal{P} \simeq L^{\mathcal{E}-1} \ll 1, \qquad \text{(A4)}$$

such that the positioning is poor for $\mathcal{E} \ll 1$.

In the opposite regime $\mathcal{E} \gg 1$, the integral in Eq. (A3) diverges and the partition function is dominated by the deepest wells. Thus, it can be estimated using k-minimal energies, given by

$$E_k^{\text{o}} \simeq \mathcal{E} \ln \frac{L}{k}. \qquad \text{(A5)}$$

Therefore,

$$Z \simeq \sum_{k=1}^{\infty} \left(\frac{L}{k} \right)^{\mathcal{E}} = L^{\mathcal{E}} \zeta(\mathcal{E}), \qquad \text{(A6)}$$

such that the positioning parameter in this regime is given by

$$\mathcal{P} \simeq \frac{1}{\zeta(\mathcal{E})} \qquad \text{(A7)}$$

and is larger than $\frac{1}{2}$ and close to 1 for

$$\mathcal{E} \gg \zeta^{-1}(2) \simeq 1.7. \qquad \text{(A8)}$$

In sum, for down-exponential distribution of energies the only requirement for a good positioning is that the parameter $\mathcal{E}$ is larger than one for any length of the DNA.

b. Disordered energy landscape with up-exponential distribution

Consider another scenario with what we denote as *up*-exponential distribution:

$$\Pr(E_i) = \frac{1}{\mathcal{E}} e^{-\frac{E_i}{\mathcal{E}}} \; ; \; 0 \leqslant E_i < \infty \; ; \; \mathcal{E} > 0. \qquad \text{(A9)}$$

In this case, there are many energies which are close to zero, $E_1^{\text{o}} \simeq E_2^{\text{o}} \simeq E_3^{\text{o}} \simeq \ldots \simeq 0$. Positioning is impossible for such energy landscape; for any value of $\mathcal{E}$ one has $\mathcal{P} \ll 1$. However, as we show in the following, interaction between nucleosomes can induce reasonable positioning even on such an energy landscape.

2. Positioning of multiple nucleosomes with only hard-core interactions

Here, we consider positioning of nucleosomes with hard-core interactions on up-exponential and down-exponential energy landscapes.

a. Disordered energy landscape with down-exponential distribution

Consider positioning of N nucleosomes on uncorrelated disordered energy profile down-exponentially distributed with some parameter $\mathcal{E}$ [see Eq. (A1)]. In the regime $\mathcal{E} \ll 1$, the nucleosomes are poorly positioned, while in the opposite regime $\mathcal{E} \gg 1$, the positioning is good.

The derived requirement for positioning may sound weak. However, in fact it means that, say, for $\rho = 70\%$ and $\mathcal{E} = 1.5 k_{\text{B}} T$ (moderate positioning regime, $\mathcal{P} \simeq 0.6$) the typical energy well for a nucleosome is $7.5 \pm 2 k_{\text{B}} T$ deep [see Eq. (A2) with L replaced by L/N], relative to a random DNA sequence.

b. Disordered energy landscape with up-exponential distribution

Consider positioning of N nucleosomes on uncorrelated disordered energy profile up-exponentially distributed with some parameter $\mathcal{E}$ [see Eq. (A9)]. In this case, as for a single nucleosome, the positioning is poor for any value of $\mathcal{E}$.

3. Positioning of strongly interacting nucleosomes

Here, we discuss positioning of strongly interacting nucleosomes on up-exponential and down-exponential energy landscapes.
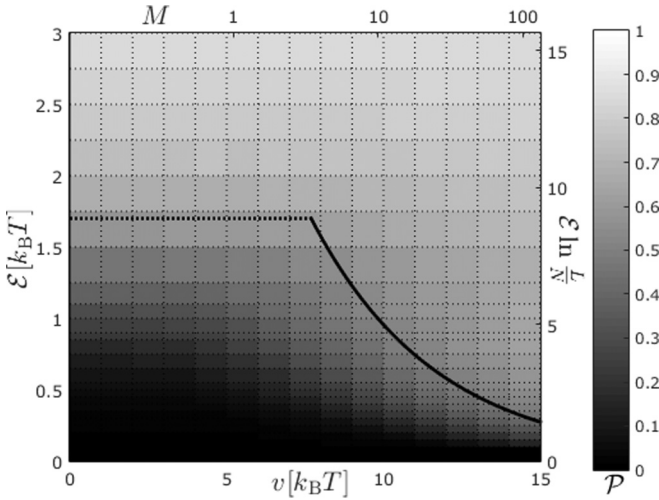
FIG. 15. Positioning of nucleosomes with interactions on down-exponential energy landscape. Positioning parameter $\mathcal{P}$ for the DNA with length $L = 10^4 \times W$ and $W = 147$ for the coverage fraction $\rho = nW/L$ smaller than 80% is plotted vs disorder strength (left axis) and interaction strength (bottom axis) with preferable distance of $R = 148$. On the top one can see the average size of the crystallized cluster of nucleosomes, derived from Eq. (29). On the right the typical binding energy of a nucleosome (relative to the average energy) is shown. The lines represent the analytic conditions for a good positioning [Eqs. (A8) (dotted line) and (A11) (solid line)].

a. Disordered energy landscape with down-exponential distribution

Consider positioning of N nucleosomes on uncorrelated disordered energy profile down-exponentially distributed with some parameter $\mathcal{E}$ [see Eq. (A1)]. In this case, a cluster of $m \gg 1$ crystallized nucleosomes has a Gaussian energy landscape with a standard deviation of $\sqrt{m}\mathcal{E}$. Using the same arguments as for the Gaussian disorder one can derive two conditions for a good positioning. The first is for positioning of weakly interacting nucleosomes, such that $M \simeq 1$. In this case, the condition is given by Eq. (A8).

If this condition is not satisfied, one needs strongly interacting nucleosomes, such that

$$\sqrt{M}\mathcal{E} \gg \sqrt{2\ln R} \qquad (A10)$$

or, using Eq. (29),

$$\kappa \gg \frac{4}{\mathcal{E}^4}\frac{L}{N}\ln^2 R. \qquad (A11)$$
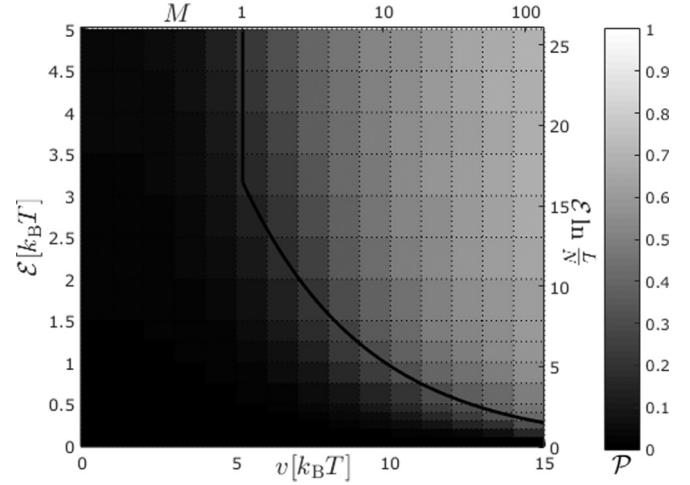

FIG. 16. Positioning of nucleosomes with interactions on up-exponential energy landscape. Positioning parameter $\mathcal{P}$ for the DNA with length $L = 10^4 \times W$ and $W = 147$ for the coverage fraction $\rho = nW/L$ smaller than 80% is plotted vs disorder strength (left axis) and interaction strength (bottom axis) with preferable distance of $R = 148$. On the top one can see the average size of the crystallized cluster of nucleosomes, derived from Eq. (29). On the right the typical energy *peak* (not well) is shown on the length of L/N. The line represents the analytic conditions for a good positioning [Eq. (A12)].

In Fig. 15 one can see that Eqs. (A8) and (A11) predict the border between the two regimes, with good and bad positioning.

b. Disordered energy landscape with up-exponential distribution

Consider positioning of N nucleosomes on uncorrelated disordered energy profile up-exponentially distributed with some parameter $\mathcal{E}$ [see Eq. (A9)]. In this case, as for a single nucleosome, the positioning is poor for any value of $\mathcal{E}$ and one needs strongly interacting nucleosomes for a good positioning.

If nucleosomes strongly interact and form crystallized clusters of $M \gg 1$ nucleosomes on average (and this happens when $\kappa \gg \frac{L}{N}$), the positioning condition is given by Eq. (A11). Thus, the condition for positioning in this case is

$$\kappa \gg \frac{L}{N} \quad \text{and} \quad \frac{4}{\mathcal{E}^4}\frac{L}{N}\ln^2 R. \qquad (A12)$$

In Fig. 16 one can see that Eq. (A12) predicts the border between the two regimes, with good and bad positioning.

[1] R. D. Kornberg and Y. Lorch, Cell **98**, 285 (1999).

[2] T. J. Richmond and C. A. Davey, Nature (London) **423**, 145 (2003).

[3] D. J. Gaffney, G. McVicker, A. A. Pai, Y. N. Fondufe-Mittendorf, N. Lewellen, K. Michelini, J. Widom, Y. Gilad, and J. K. Pritchard, PLoS Genetics **8**, e1003036 (2012).

[4] E. C. Small, L. Xi, J.-P. Wang, J. Widom, and J. D. Licht, Proc. Natl. Acad. Sci. USA **111**, E2462 (2014).

[5] C. Jiang and B. F. Pugh, Nat. Rev. Genet. **10**, 161 (2009).

[6] L. Bai and A. V. Morozov, Trends Genet. **26**, 476 (2010).

[7] Y. Belch, J. Yang, Y. Liu, S. A. Malkaram, R. Liu, J.-J. M. Riethoven, and I. Ladunga, PLoS One **5**, e12984 (2010).

[8] R. E. Thurman, E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, B. Vernot *et al.*, Nature (London) **489**, 75 (2012).

[9] F. H. Lam, D. J. Steger, and E. K. O'Shea, Nature (London) **453**, 246 (2008).

[10] K. Struhl and E. Segal, Nat. Struct. Mol. Biol. **20**, 267 (2013).

[11] E. Trifonov, Nucleic Acids Res. **8**, 4041 (1980).

[12] S. C. Satchwell, H. R. Drew, and A. A. Travers, J. Mol. Biol. **191**, 659 (1986).

[13] A. Thåström, P. Lowary, H. Widlund, H. Cao, M. Kubista, and J. Widom, J. Mol. Biol. **288**, 213 (1999).

[14] A. Thåström, P. Lowary, and J. Widom, Methods **33**, 33 (2004).

[15] A. V. Morozov, K. Fortney, D. A. Gaykalova, V. M. Studitsky, J. Widom, and E. D. Siggia, arXiv:0805.4017.

[16] T. E. Takasuka and A. Stein, Nucleic Acids Res. **38**, 5672 (2010).

[17] D. D. Winkler, U. M. Muthurajan, A. R. Hieb, and K. Luger, J. Biol. Chem. **286**, 41883 (2011).

[18] R. Padinhateeri and J. F. Marko, Proc. Natl. Acad. Sci. USA **108**, 7799 (2011).

[19] M. Vignali, A. H. Hassan, K. E. Neely, and J. L. Workman, Mol. Cell. Biol. **20**, 1899 (2000).

[20] V. B. Teif and K. Rippe, Nucleic Acids Res. **37**, 5641 (2009).

[21] P. Korber, T. Luckenbach, D. Blaschke, and W. Hörz, Mol. Cell. Biol. **24**, 10965 (2004).

[22] Z. Zhang, C. J. Wippo, M. Wal, E. Ward, P. Korber, and B. F. Pugh, Science **332**, 977 (2011).

[23] N. Kaplan, I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, D. Tillo, Y. Field, E. M. LeProust, T. R. Hughes, J. D. Lieb, J. Widom *et al.*, Nature (London) **458**, 362 (2009).

[24] J. Widom, Proc. Natl. Acad. Sci. USA **89**, 1095 (1992).

[25] B. Mergell, R. Everaers, and H. Schiessel, Phys. Rev. E **70**, 011915 (2004).

[26] F. Mühlbacher, H. Schiessel, and C. Holm, Phys. Rev. E **74**, 031919 (2006).

[27] A. Cherstvy and R. Everaers, J. Phys.: Condens. Matter **18**, 11429 (2006).

[28] R. Stehr, N. Kepper, K. Rippe, and G. Wedemann, Biophys. J. **95**, 3677 (2008).

[29] J.-P. Wang, Y. Fondufe-Mittendorf, L. Xi, G.-F. Tsai, E. Segal, and J. Widom, PLoS Comput. Biol. **4**, e1000175 (2008).

[30] S. A. Grigoryev, G. Arya, S. Correll, C. L. Woodcock, and T. Schlick, Proc. Natl. Acad. Sci. USA **106**, 13317 (2009).

[31] Y. Liu, C. Lu, Y. Yang, Y. Fan, R. Yang, C.-F. Liu, N. Korolev, and L. Nordenskiöld, J. Mol. Biol. **414**, 749 (2011).

[32] R. V. Chereji, D. Tolkunov, G. Locke, and A. V. Morozov, Phys. Rev. E **83**, 050903 (2011).

[33] R. V. Chereji and A. V. Morozov, J. Stat. Phys. **144**, 379 (2011).

[34] A. G. Cherstvy and V. B. Teif, Phys. Biol. **11**, 044001 (2014).

[35] D. A. Beshnova, A. G. Cherstvy, Y. Vainshtein, and V. B. Teif, PLoS Comput. Biol. **10**, e1003698 (2014).

[36] M. Smolle and J. L. Workman, Biochim. Biophys. Acta **1829**, 84 (2013).

[37] M. Engeholm, M. de Jager, A. Flaus, R. Brenk, J. van Noort, and T. Owen-Hughes, Nat. Struct. Mol. Biol. **16**, 151 (2009).

[38] W. Möbius, B. Osberg, A. M. Tsankov, O. J. Rando, and U. Gerland, Proc. Natl. Acad. Sci. USA **110**, 5719 (2013).

[39] R. V. Chereji and A. V. Morozov, Proc. Natl. Acad. Sci. USA **111**, 5236 (2014).

[40] V. B. Teif and K. Rippe, J. Phys.: Condens. Matter **22**, 414105 (2010).

[41] C. P. Woodbury, Biopolymers **20**, 2225 (1981).

[42] G. V. Gurskii and A. S. Zasedatelev, Biofizika **23**, 932 (1978).

[43] V. B. Teif and K. Rippe, Briefings Bioinform. **13**, 187 (2012).

[44] B. Derrida, Phys. Rev. Lett. **45**, 79 (1980).

[45] U. Gerland, J. D. Moroz, and T. Hwa, Proc. Natl. Acad. Sci. USA **99**, 12015 (2002).

[46] M. Slutsky and L. A. Mirny, Biophys. J. **87**, 4021 (2004).

[47] M. Sheinman, O. Bénichou, Y. Kafri, and R. Voituriez, Rep. Prog. Phys. **75**, 026601 (2012).

[48] L. De Haan and A. Ferreira, *Extreme Value Theory: An Introduction* (Springer, Berlin, 2007).

[49] *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1964).

[50] A. Sharma and M. Sheinman, J. Phys. D: Appl. Phys. **46**, 125106 (2013).

[51] J. D. McGhee and P. H. von Hippel, J. Mol. Biol. **86**, 469 (1974).

[52] G.-W. Li, O. G. Berg, and J. Elf, Nat. Phys. **5**, 294 (2009).

[53] M. J. Morelli, R. J. Allen, and P. R. Ten Wolde, Biophys. J. **101**, 2882 (2011).

[54] M. Sheinman and Y. Kafri, Phys. Biol. **9**, 056006 (2012).

[55] I. R. Epstein, Biophys. Chem. **8**, 327 (1978).

[56] R. D. Kornberg and L. Stryer, Nucleic Acids Res. **16**, 6677 (1988).

[57] S. Lubliner and E. Segal, Bioinformatics **25**, i348 (2009).

[58] C. Peng, S. Buldyrev, A. Goldberger, S. Havlin, F. Sciortino, M. Simons, H. Stanley *et al.*, Nature (London) **356**, 168 (1992).

[59] B. Derrida, J. Phys. Lett. **46**, 401 (1985).

[60] M. Deserno, http://www.cmu.edu/biolphys/deserno/pdf/corr_gaussian_random.pdf

[61] M. Y. Tolstorukov, V. Choudhary, W. K. Olson, V. B. Zhurkin, and P. J. Park, Bioinformatics **24**, 1456 (2008).

[62] G. Locke, D. Tolkunov, Z. Moqtaderi, K. Struhl, and A. V. Morozov, Proc. Natl. Acad. Sci. USA **107**, 20998 (2010).

[63] T. van der Heijden, J. J. van Vugt, C. Logie, and J. van Noort, Proc. Natl. Acad. Sci. USA **109**, E2514 (2012).

[64] D. Tolkunov and A. V. Morozov, Adv. Protein Chem. Struct. Biol. **79**, 1 (2010).

[65] V. Teif, A. Shkrabkou, V. Egorova, and V. Krot, Mol. Biol. **46**, 1 (2012).

[66] E. N. Trifonov, Phys. Life Rev. **8**, 39 (2011).

[67] H.-R. Chung, I. Dunkel, F. Heise, C. Linke, S. Krobitsch, A. E. Ehrenhofer-Murray, S. R. Sperling, and M. Vingron, PLoS One **5**, e15754 (2010).

[68] K. R. Bradnam, C. Seoighe, P. M. Sharp, and K. H. Wolfe, Mol. Biol. Evolution **16**, 666 (1999).

[69] A. B. Cohanim, Y. Kashi, and E. N. Trifonov, J. Biomol. Struct. Dyn. **23**, 559 (2006).

[70] D. Lohr and K. Van Holde, Proc. Natl. Acad. Sci. USA **76**, 6326 (1979).

[71] K. Brogaard, L. Xi, J.-P. Wang, and J. Widom, Nature (London) **486**, 496 (2012).

[72] F. Strauss and A. Prunell, EMBO J. **2**, 51 (1983).

[73] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.92.022704 for figures for the tuning parameters and robustness of the model.