Biophysical Journal
Article

Biophysical Society

Effects of size, cooperativity, and competitive binding on protein positioning on DNA

Leo S. McCormack,[1,2] Artem K. Efremov,[2,*] and Jie Yan[2,3,*]
[1]Department of Physics, Imperial College London, London, United Kingdom; [2]Mechanobiology Institute and [3]Department of Physics, National University of Singapore, Singapore, Singapore

ABSTRACT Accurate positioning of proteins on chromosomal DNA is crucial for its proper organization as well as gene transcription regulation. Recent experiments revealed existence of periodic patterns of nucleoprotein complexes on DNA, which frequently cannot be explained by sequence-dependent binding of proteins. Previous theoretical studies suggest that such patterns typically emerge as a result of the proteins' volume-exclusion effect. However, the role of other physical factors in patterns' formation, such as the length of DNA, its sequence heterogeneity, and protein binding cooperativity/binding competition to DNA, remains unclear. To address these less understood yet important aspects, we investigated potential effects of these factors on protein positioning on finite-size DNA by using transfer-matrix calculations. It has been found that upon binding to DNA, proteins form oscillatory patterns that span over the length of up to ~10 times the size of the protein binding site, with the shape of the patterns being strongly dependent on the length of DNA and the proteins' binding cooperativity to DNA. Furthermore, calculations showed that small variations in the proteins' affinity to DNA due to its sequence heterogeneity do not much change the main geometric characteristics of the observed protein patterns. Finally, competition between two different types of proteins for binding to DNA has been found to lead to formation of highly diverse and complex alternating positioning of the two proteins. Altogether, these results provide new insights into the roles of physicochemical properties of proteins, the DNA length, and DNA-binding competition between proteins in formation of protein positioning patterns on DNA.

SIGNIFICANCE Organization of chromosomal DNA in living cells requires an intricate interplay among different types of DNA-binding proteins, which frequently form regular patterns on DNA that have a strong impact on its resulting conformation as well as the global gene transcription profile. Only recently, with the help of statistical physics methods, have we started to gain understanding regarding the physical principles that govern the formation of such protein patterns on chromosomal DNA. In this study, we investigate in detail the role of physicochemical properties of DNA-binding proteins as well as DNA-binding competition between different types of proteins in the formation of the protein positioning patterns on finite-size DNA.

INTRODUCTION

In bacterial and mammalian cells, genomic DNA is organized into a compact chromatin structure by a variety of DNA architectural proteins. It has been estimated that ~75% of DNA in human cells is occupied by nucleosomes (1,2), whose positioning on chromosomal DNA has a strong effect on its accessibility to transcription factors as well as DNA and RNA polymerases, all of which play a vital role in maintenance and processing of genetic information stored in the DNA. For this reason, one of the focuses of previous studies has been the understanding of physicochemical and biological factors governing protein positioning on DNA.

To this aim, a number of theoretical and experimental tools have been developed during past years aimed at prediction and measurement of protein positioning patterns along DNA, including, but not limited to, such theoretical methods as molecular and Brownian dynamics simulations (3,4) as well as Metropolis Monte Carlo and transfer-matrix calculations (2,5,6), whose main goal was interpretation of experimental data obtained in protein footprint assays (7–10), atomic force microscopy (11–13), and single-DNA manipulation assays (14–16). Multiple studies based on these methods have revealed that one of the most important contributions to the observed localization of proteins on

Submitted November 24, 2020, and accepted for publication March 18, 2021.

*Correspondence: phyyj@nus.edu.sg or mbiay@nus.edu.sg

Editor: Mark C. Williams.
https://doi.org/10.1016/j.bpj.2021.03.016

chromosomal DNA is made by the intrinsic sequence-specific binding affinity of proteins to various parts of DNA (17). This finding resulted in the development of numerous bioinformatic and biophysical approaches based on sequence-encoded effects of nucleosome positioning, which have been shown to be relatively effective in modeling of the large-scale genomic organization, correctly predicting the positions of ~50% of the nucleosomes on chromosomal DNA (18–21).

Although providing, on average, more or less accurate results on the large genome scale, it has been found that in many cases, nucleosome positioning on specific segments of chromosomal DNA shows patterns beyond the prediction by sequence-dependent binding of histone octamers. A notable example is the periodic positioning of nucleosomes observed upstream of transcription start sites (22,23), where oscillating patterns with sharp peaks are observed, with decreasing amplitudes further away, regardless of the sequence-specific affinity of histones. Such oscillatory patterns arise because of a steric hindrance formed by transcription factors that forbid nucleosomes from binding to transcription start sites, leading to emergence of the abnormal nucleosome positioning next to them. At a distance of ~10 nucleosome sizes, the positioning pattern of nucleosomes again correlates well with the intrinsic sequence-specific binding affinity of histones to DNA (24,25). Subsequently, theoretical studies using statistical physics methods have shown that protein positioning on DNA is not solely determined by proteins binding strength to various parts of DNA but in fact is an outcome of the tug of war between protein sequence-specific affinities and positioning entropy effects (26). This finding helped to explain experimentally observed oscillatory positioning patterns of nucleosomes near large protein complexes, showing very good agreement between theoretical predictions and experimental measurements.

Despite the successful explanation of many experimentally observed phenomena regarding the positioning of DNA-binding proteins, it should be noted that previous theoretical studies have not explored in detail the role of several important factors that determine DNA organization by such proteins in living cells. Namely, it is known that maintenance and organization of chromosomal DNA is performed by a number of protein species, which frequently compete with each other for DNA binding. Indeed, experimental and theoretical studies have shown that DNA-binding competition between histones and other proteins plays an important role in the formation of protein positioning patterns, by which some regions become abundant with smaller proteins and others are depleted (27,28). Such an effect may have a major role in gene transcription regulation, as it has been revealed that formation of nucleosomes near promoter regions frequently has a strong effect on the transcription level of downstream genes (29). Thus, existing experimental studies suggest that DNA-binding competition between different types of proteins plays a crucial role in their positioning on DNA that has a drastic effect on DNA organization and gene transcription regulation. However, there is a lack of theoretical works addressing this important point, and as a result, the role of DNA-binding competition and DNA-binding cooperativity between proteins in the formation of such patterns still remains unclear.

To fill the gap in understanding the role of aforementioned physical effects in formation of protein positioning patterns on DNA, we have used a semi-analytic transfer-matrix approach (2). As a result, it has been found that DNA-binding proteins create oscillatory positioning patterns spanning the length of several protein sizes next to steric barriers that can be formed either by the DNA ends or by large protein complexes. Variation of the model parameters further revealed that the shapes of such positioning patterns are not only sensitive to the DNA-binding affinity and cooperative binding energy of proteins to DNA but also to the ratio between the size of nucleoprotein complexes and the length of DNA segment available for the complexes' formation. Major spatial characteristics of such patterns have been found to be independent from small fluctuations in the protein binding energy to DNA arising from the sequence inhomogeneity of the latter, showing that protein positioning patterns are robust to small random binding energy variations. The effects of these parameters are also probed for the case of two different protein species competing for binding to DNA, in which it is found that the observed positioning patterns for both proteins have strong sensitivity to the parameter values, whose variation leads to significant changes in the proteins' distribution on the DNA.

METHODS

To calculate occupancy of finite-length DNA by nonoverlapping proteins, we employed a transfer-matrix approach, which has been previously described in detail by Teif (28) in regard to sequence-dependent positioning of proteins on DNA. In this work, we use it to investigate the "tug of war" between the three main physical forces that govern proteins' positioning on DNA: protein-DNA and protein-protein interactions, as well as the system entropy, which tends to be maximized in thermodynamic equilibrium.

To this aim, DNA in this study is modeled as a linear polymer composed of M segments, each of which has a length corresponding to one DNA base-pair (bp). In the calculations, we explore the following two scenarios: 1) DNA interacting with a single species of protein and 2) competitive binding of two different types of proteins, referred to as A and B, to the same DNA. For the sake of generality, here we show only formulas describing the more complex second case, from which one can easily derive mathematical expressions for the first scenario simply by removing protein B from the equations.

To investigate the role of protein-DNA and protein-protein interactions in formation of the positioning patterns on DNA, in this study it is assumed that proteins A and B bind to DNA with free energies $\varepsilon_A = k_B T \ln(c_A/K_{A,d})$ and $\varepsilon_B = k_B T \ln(c_B/K_{B,d})$ and that upon binding, they occupy K_A and K_B DNA segments (i.e., DNA bps), respectively. Here, c_A, c_B, $K_{A,d}$, and $K_{B,d}$ are concentrations of proteins A and B in solution and their equilibrium binding constants to DNA, respectively.

To take into account possible binding cooperativities between adjacent proteins on DNA, we also introduced into the calculations cooperative

binding energies μ_{AA}, μ_{BB}, and μ_{AB} (= μ_{BA}) denoting the interactions between two adjacent proteins of the corresponding types (AA, AB, or BB).

In the calculations, multiple proteins are allowed to bind to the same DNA as soon as the resulting nucleoprotein complexes do not overlap with each other. According to the transfer-matrix method, the physical configuration of protein-covered DNA can then be completely specified by the set of M indexes, (s_1, …, s_M), each of which indicates the state of the corresponding DNA segment. Thus, for example, index s_i denotes the state of the i^{th} DNA segment. Because proteins A and B occupy K_A and K_B segments upon binding to DNA, it is clear that each DNA segment can exist in one of the $K_A + K_B + 1$ states, which we enumerate as 0, 1_A, 2_A, 3_A, …, K_A, 1_B, 2_B, 3_B, …, K_B. Specifically, $s_i = 0$ indicates that the i^{th} DNA segment is free and is not bound to any of the proteins, A or B, whereas $s_i = 1_A$, …, K_A and $s_i = 1_B$, …, K_B correspond to situations in which the i^{th} DNA segment interacts with the respective binding site on the surface of one of the proteins A or B (see Fig. 1 for more details).

By knowing the state of each DNA segment, the total conformational energy of DNA can be written as a sum of the individual protein's binding and cooperative binding energies:

$$E_{\text{tot}}(s_1, s_2, \ldots, s_M) = -\sum_{i=1}^{M}\sum_{u=A,B}\left[\frac{\varepsilon_u}{K_u}\times\sum_{j=1_u}^{K_u}\delta_{s_i j}\right] - \sum_{i=1}^{M-1}\sum_{u,v=A,B}\mu_{uv}\,\delta_{s_i K_u}\delta_{s_{i+1}1_v}. \quad (1)$$

where $\delta_{s_i s_j}$ is the Kronecker delta ($\delta_{s_i s_j} = 1$ if $s_i = s_j$ and 0 otherwise). In the above equation, the first term describes the sum of all of the proteins' binding energies to DNA in a given protein-DNA configuration, where for simplicity it is assumed that the binding energy of each protein is equally distributed between the protein-bound DNA segments. Therefore, the average binding energy distributed to one site is ε_A/K_A or ε_B/K_B for protein A or B, respectively. As for the second term, it describes the sum of nearest-neighbor interactions between all of the adjacent proteins bound to DNA. It should be noted that Eq. 1 automatically takes into account the volume-exclusion effect between neighboring proteins, as none of the DNA segments can be in a state in which two different proteins are bound to it in an overlapping fashion.

For the sake of simplicity and clarity of the results, in this study we assumed that all the binding sites on the surface of a protein are fully engaged with DNA once the protein is bound to it. In the model, this is achieved by imposing an infinite energy penalty to the following cases of partially unfolded nucleoprotein complexes (see (30)): 1) if $s_i = 0$ and $s_{i+1} = 2_u$, …, K_u, where u = A or B; 2) if $s_i = K_u$ and $s_{i+1} = 2_v$, …, K_v, where u, v = A or B; and 3) if $s_i = j_u$, where $j_u = 1_u$, …, $K_u - 1$ and $s_{i+1} \neq (j+1)_u$. For more details, see Fig. 1 and Supporting materials and methods.

Although the final expression for the total conformational energy of DNA (Eqs. S1–S11) looks somewhat complicated, it is not hard to show that it can be presented in a simple mathematical form:

$$E_{\text{tot}}(s_1, s_2, \ldots, s_M) = E_{s_1} + \sum_{i=1}^{M-1} E_{s_i, s_{i+1}} + E_{s_M}, \quad (2)$$

where $E_{s_i, s_{i+1}}$ are the nearest-neighbor energy terms comprising interactions between DNA and proteins (if any) bound to the i^{th} and $(i+1)^{\text{th}}$ DNA segments. As for the first and last DNA segments, $i = 1$ and $i = M$, they stand separately from other DNA segments because the first DNA segment does not have an upstream neighbor, whereas the last DNA segment does not have a downstream neighbor. Thus, the number of states in which these two segments can possibly be is limited in comparison to other DNA segments, and the E_{s_1} and E_{s_M} energy terms in Eq. 2 simply describe the respective boundary conditions imposed on the DNA ends (see Eqs. S11 and S12, a–c).

Using the fact that the total DNA energy can be represented as a sum of energy terms corresponding to neighboring DNA segments, it is then straightforward to calculate the partition function of DNA by employing the transfer-matrix formalism. Namely, by introducing a transfer-matrix $T_i(s_i, s_{i+1}) = e^{-\beta E_{s_i s_{i+1}}}$ associated with each pair of neighboring DNA segments, i and $i+1$, it can be shown that the partition function of DNA can be written as the following matrix product (see Eq. S15): $Z = V_1^{\text{T}} T_1 T_2 \ldots T_{M-1} V_M$, where V_1^{T} is a transposed vector describing the boundary conditions imposed on the first DNA segment (namely, s_1 can be either in state 0 or 1_u, u = A or B) and V_M is the vector describing the boundary conditions imposed on the last DNA segment (s_M can be either in state 0 or K_u, u = A or B). Here, $\beta = 1/k_B T$, where k_B is Boltzmann's constant and T is temperature.

By using the notion of a matrix trace, Tr, and its invariability under cyclic matrix permutations, the above formula for the DNA partition function can be further rewritten as

$$Z = \sum_{s_1,\ldots,s_M} e^{-\beta E_{\text{tot}}(s_1, s_2, \ldots, s_M)}\xi(s_1, s_M) = \text{Tr}(W T_1 T_2 \cdots T_{M-1}), \quad (3)$$

where $\xi(s_1, s_M)$ is a boundary condition function imposing restriction on the states of the DNA end segments and $W = V_M V_1^{\text{T}}$ is the corresponding boundary condition matrix.

Having in hand a mathematical expression for the DNA partition function, the probability of a particular DNA confirmation can be calculated as $p(s_1, s_2, \ldots, s_M) = \frac{1}{Z} e^{-\beta E_{\text{tot}}(s_1, s_2, \ldots, s_M)}$. Furthermore, it is easy to calculate the probability $p(s_i = \alpha)$ of the i^{th} DNA segment to be in a particular state, α, as

$$p(s_i = \alpha) = \frac{\text{Tr}\left(W T_1 T_2 \cdots \tilde{T}_i \cdots T_{M-1}\right)}{\text{Tr}(W T_1 T_2 \cdots T_i \cdots T_{M-1})}, \quad (4)$$

where $\tilde{T}_i$ is a modified transfer-matrix associated with the i^{th} DNA segment, whose entries are all equal to 0 except for the α^{th} row, which is kept the same as in the original transfer-matrix T_i; see Supporting materials and methods for more details.

Using these formulas, it is then straightforward to calculate the following quantities in addition to $p(s_i = \alpha)$, which were studied in this work: 1) the site-occupancy probability for the i^{th} DNA segment to be bound to any protein, $p_{b,i} = 1 - p(s_i = 0)$; and 2) the average DNA occupancy fraction by proteins, $\theta = \sum_{i=1}^{M} p_{b,i}/M$.

The focus of the study was to understand effects of the sizes of DNA and proteins, as well as the DNA sequence heterogeneity and the proteins' binding energy, on the above quantities.

To study the behavior of the site-occupancy probability, $p_{b,i}$, as a function of the DNA site index, i, which frequently forms oscillating patterns, we determined the period of each resulting pattern for $p_{b,i}$ by measuring valley-to-valley distances between the peaks of the pattern. In addition, we also numerically calculated its Fourier spectrum by applying a discrete Fourier transform (using the numpy.fft package in the Python coding language), in which the strongest mode present in the Fourier spectrum determines the main oscillation frequency of the pattern and hence gives a good indication of its wavelength, or in other words, the period of the pattern.

RESULTS

DNA site-dependent occupancy probability

To find out how the finite size of proteins and DNA influences formation of nucleoprotein complexes, we calculated

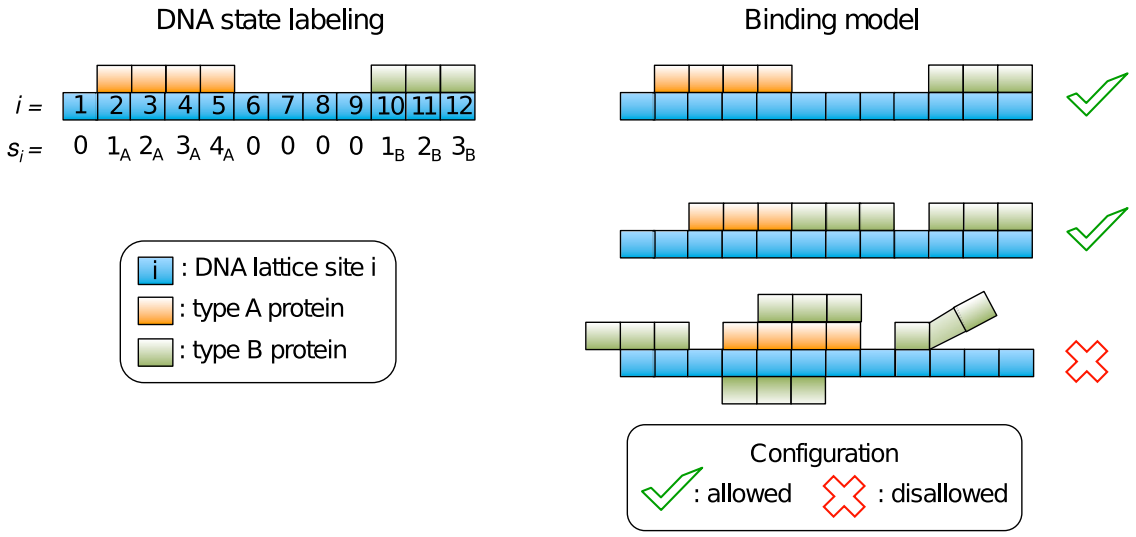

FIGURE 1 Schematic picture of allowed and disallowed DNA-protein conformations in transfer-matrix calculations. For the sake of simplicity, the figure shows the case of a DNA molecule having the length of $M = 12$ bp and two types of proteins, A and B, interacting with it. The binding site size of protein A is $K_A = 4$ bp, and for protein B, it is $K_B = 3$ bp. In the model, each DNA lattice site may only bind to one protein at a time, and each nucleoprotein complex results from protein-DNA interaction such that all of the protein sites are occupied by the DNA (i.e., there are no partially unfolded nucleoprotein complexes). To see this figure in color, go online.

the site-dependent occupancy probability, $p_{b,i}$, for different lengths of DNA. In these calculations, we considered DNA interaction with a single type of DNA-binding protein (referenced to in the model as protein A), which was assumed to occupy 15 bps upon binding to DNA—a typical size of the binding site for many known DNA-interacting proteins (31).

Fig. 2, *A* and *B* show the result obtained on 100 and 200 bp DNA in the case when the protein has zero binding free energy ($\varepsilon_A = 0\ k_BT$, i.e., the protein concentration is chosen at the value of the equilibrium dissociation constant) and zero cooperative binding energy ($\mu_{AA} = 0\ k_BT$) to all DNA segments. Chemical kinetics theory predicts that at these conditions the average occupancy fraction of DNA, θ, must be 50% if a protein binds to a single DNA site (i.e., a single bp). However, in the case in which the protein binds to a DNA segment that spans several DNA bp, the equilibrium DNA site-occupancy probability is expected to be higher. This is because if one bp is bound to a protein, then there must be multiple bp nearby bound to the same protein, and as a result, "artificial" cooperativity in DNA bp occupancies appears in the model in the case of proteins whose binding site spans more than one bp. Consistently, our transfer-matrix calculations show that the average site-occupancy fraction reaches $\theta \sim 67\%$ (Fig. 2, *A* and *B*) for protein A occupying $K_A = 15$ bp upon binding to DNA.

What is, however, interesting is that the occupancy of DNA sites is not uniform, and an oscillating pattern created by DNA-bound proteins can be clearly seen in Fig. 2, *A* and *B* near the two DNA ends. Fourier transformation of the site-occupancy probability reveals that the main frequency of the oscillations is $\sim 0.05\ \text{bp}^{-1}$ (Fig. S1), which approximately equals the inverse of the protein-DNA binding segment size (15 bp). This suggests that the oscillation frequency is correlated with the protein-bound segment size, which is further confirmed in our calculations for proteins and DNA molecules of different sizes (Fig. S2).

Although being strongly pronounced at the DNA ends, the oscillation pattern quickly decays toward the middle of the DNA, and already at distances of >4 protein sizes, practically no DNA occupancy oscillations can be seen. Therefore, one would expect that oscillatory patterns span the whole DNA only for shorter DNA molecules. This is indeed the case as shown in Fig. 2 *A*, displaying the site-occupancy fraction of 100 bp DNA, at which the oscillation persists into the middle of the DNA. Thus, it can be concluded that nonuniform protein positioning patterns cover the entire DNA only if the DNA/protein size ratio is less than $\sim 10:1$, which is supported by calculations (see Fig. S3 for multiple sizes of DNA and protein). Together, these results indicate that proteins' positioning on DNA is not as random on short DNA as on long DNA molecules. To our knowledge, this point has been underappreciated in previous studies.

To obtain further insights into the role of other physical factors in formation of oscillatory protein positioning patterns on DNA, we investigated the effects of the proteins' binding free energy on short DNA segments. Fig. 2, *A* and *B* show site-dependent occupancy probability of DNA, $p_{b,i}$, calculated for two DNA-binding free energies of the protein, $\varepsilon_A = 0.5\ k_BT$ and $2.0\ k_BT$, interacting with either 100 bp DNA or 200 bp DNA.

From Fig. 2, it can be clearly seen that the protein binding free energy has two important effects on the DNA site-dependent occupancy fraction. First, higher binding

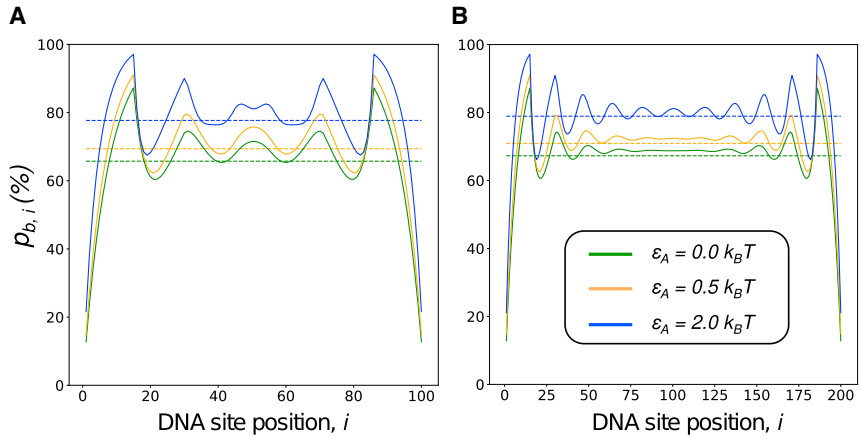

FIGURE 2 Site-dependent occupancies of homogeneous 100 (*A*) and 200 bp (*B*) DNA by proteins of binding site size $K_A = 15$ bp. The binding cooperativity of proteins was set to 0 in the calculations ($\mu_{AA} = 0\ k_BT$), and the protein binding energy had values of $\varepsilon_A = 0.0\ k_BT$ (*green curve*), $0.5\ k_BT$ (*orange curve*), and $2.0\ k_BT$ (*blue curve*). Site-dependent oscillations of the DNA occupancy by proteins, $p_{b,i}$, can be clearly seen from the graphs. Dashed lines shown in the graphs indicate the mean DNA occupancy level, θ, by proteins. To see this figure in color, go online.

energy leads to a larger average occupancy fraction of the DNA, θ, which increases from $\theta \sim 67\%$ for $\varepsilon_A = 0\ k_BT$ to $\theta \sim 70\%$ for $\varepsilon_A = 0.5\ k_BT$ and to $\theta \sim 78\%$ for $\varepsilon_A = 2\ k_BT$ for both lengths of DNA. The second effect that can be observed from the data obtained for the 200 bp DNA is an increase in the spanning range of the oscillatory protein pattern at higher protein binding free energies (see also Fig. 5 *A*). This result suggests that besides the ratio between the DNA length and the protein binding site size, the binding free energy is one of the key factors affecting formation of the periodic protein positioning patterns on DNA.

To obtain further insights into the role of DNA-binding energy in determining the spatial characteristics of the oscillating protein positioning patterns, we investigated how the DNA sequence inhomogeneity may potentially affect the pattern formation. Indeed, from experiments it is known that proteins often bind to DNA in a sequence-dependent manner (17,18). Thus, there could be a stretch of DNA of a particular sequence that has a high affinity to the protein, directing the binding of the proteins. In addition, there could also be stretches of DNA with sequences that have very low affinity to the proteins. To understand how such high-affinity and low-affinity sequences impact the protein positioning on short DNA segments, we inserted a low-affinity binding zone ($\varepsilon_A = -10\ k_BT$) (Fig. 3 *A*) or a high-affinity zone ($\varepsilon_A = 10\ k_BT$) (Fig. 3 *B*) into a 200 bp DNA, occupying the middle DNA segments (numbers 95–109). For the rest of the DNA segments, the binding free energy was set to be equal to $0\ k_BT$. Our results show that the presence of such DNA inserts leads to appearance of protein oscillation patterns not only close to the two DNA ends but also near the protein high-affinity and low-affinity zones, which in the case of the 200 bp DNA results in the formation of oscillatory protein patterns throughout the whole DNA (see Fig. 3, *A* and *B*).

In the above example, it is assumed that the protein has the same binding free energy to all the DNA segments except for the central attracting or repulsing part of the DNA. However, in actual experiments in which DNA has very different bp composition, the site-dependent binding energy of proteins may be inhomogeneous. To examine how sequence inhomogeneity affects positioning of proteins on DNA, we introduced random site-dependent protein-binding free energy distributed according to the Gaussian law around the mean value of $0\ k_BT$ with the standard deviation, σ. The latter was varied from $\sigma = 0.05\ k_BT$ per DNA bp to $\sigma = 0.5\ k_BT$ per DNA bp, which is comparable to the distribution of the protein binding energies per single DNA bp measured in experiments (32). As shown in Fig. S4, although becoming noisier, the periodic positioning pattern is retained with the same characteristic periodicity over the tested range of σ, suggesting that the pattern formation by DNA-bound proteins is a very robust phenomenon.

We next investigated whether oscillations in protein positioning patterns result from volume-excluded competitive binding of the same type of proteins. Fixing the protein size at $K_A = 15$ bp, we calculated the site-occupancy probability for DNA of different sizes: 20, 40, and 70 bps (see Fig. 4). Because the 20 bp DNA can host no more than one protein, it is clear that there cannot be competition for DNA-binding sites between proteins. As a result, the DNA occupancy fraction graph looks flat, reaching a plateau of $\sim 85\%$, and no oscillatory positioning pattern can be observed. In contrast, 40 bp DNA can host up to two proteins, which leads to the appearance of the oscillatory positioning pattern; DNA-bound proteins tend to avoid the middle region, indicating the existence of an apparent repulsion between the proteins appearing because of the volume-exclusion effect. Consistently, on 70 bp DNA, even more protein positioning oscillations can be seen, as this DNA can host up to four DNA-bound proteins. These examples clearly demonstrate that the volume exclusion and competitive binding between proteins is the main driving force responsible for the emergence of oscillatory protein positioning patterns on DNA.

We further investigated the effects of DNA-binding cooperativity between proteins on the oscillatory protein

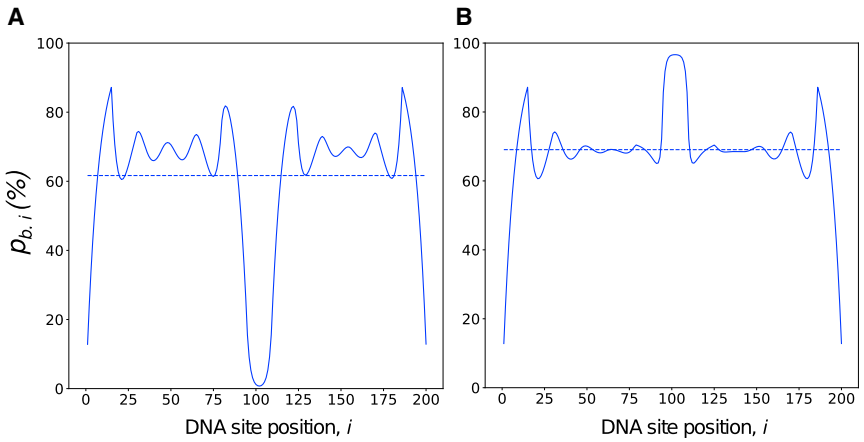

FIGURE 3 Site-dependent occupancy of inhomogeneous 200 bp DNA interacting with proteins that have binding site size of $K_A = 15$ bp. The binding energy of the proteins to DNA was set equal to 0 in the calculations ($\varepsilon_A = 0\ k_BT$) for all sites except the middle part of the DNA molecule. For the latter, we used $\varepsilon_A = -10\ k_BT$ (*A*) and $\varepsilon_A = +10\ k_BT$ (*B*). The cooperative binding energy between the proteins in the above calculations was set to 0: $\mu_{AA} = 0\ k_BT$. To see this figure in color, go online.

positioning on DNA. Fig. 4 demonstrates the results obtained on DNA molecules of various lengths in the case of the positive ($\mu_{AA} = +2\ k_BT$) and negative ($\mu_{AA} = -2\ k_BT$) cooperative binding energies of proteins to DNA. This figure shows that μ_{AA} has two important effects on the DNA-protein interaction. First, the most obvious one is a clearly visible increase in the average DNA occupancy by proteins at positive DNA-binding cooperativity (see the *dashed lines* shown in Fig. 4), which appears to be due to the support of the protein binding to DNA by nucleoprotein complexes already formed on the DNA. Second, the resulting protein positioning pattern has more peaks, but less oscillation amplitude, at positive binding cooperativity than negative binding cooperativity when DNA is long enough to host several DNA-bound proteins. Such phenomena may appear because of a larger number of proteins

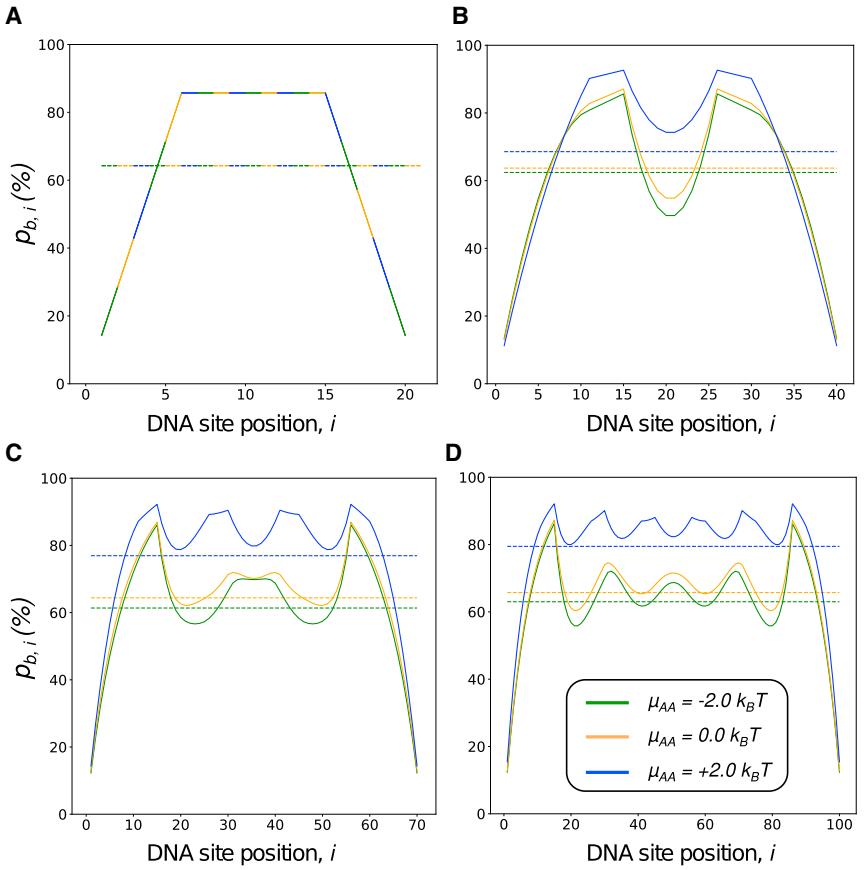

FIGURE 4 Site-dependent occupancies of homogeneous DNA molecules of length 20 bp (*A*), 40 bp (*B*), 70 bp (*C*), and 100 bp (*D*) by proteins that have the binding site size of $K_A = 15$ bp at three different values of the protein cooperative binding energy: $\mu_{AA} = -2\ k_BT$ (*green curve*), $\mu_{AA} = 0\ k_BT$ (*orange curve*), and $\mu_{AA} = +2\ k_BT$ (*blue curve*). In all of the calculations, the binding energy of the protein to DNA was set equal to 0 ($\varepsilon_A = 0\ k_BT$). To see this figure in color, go online.

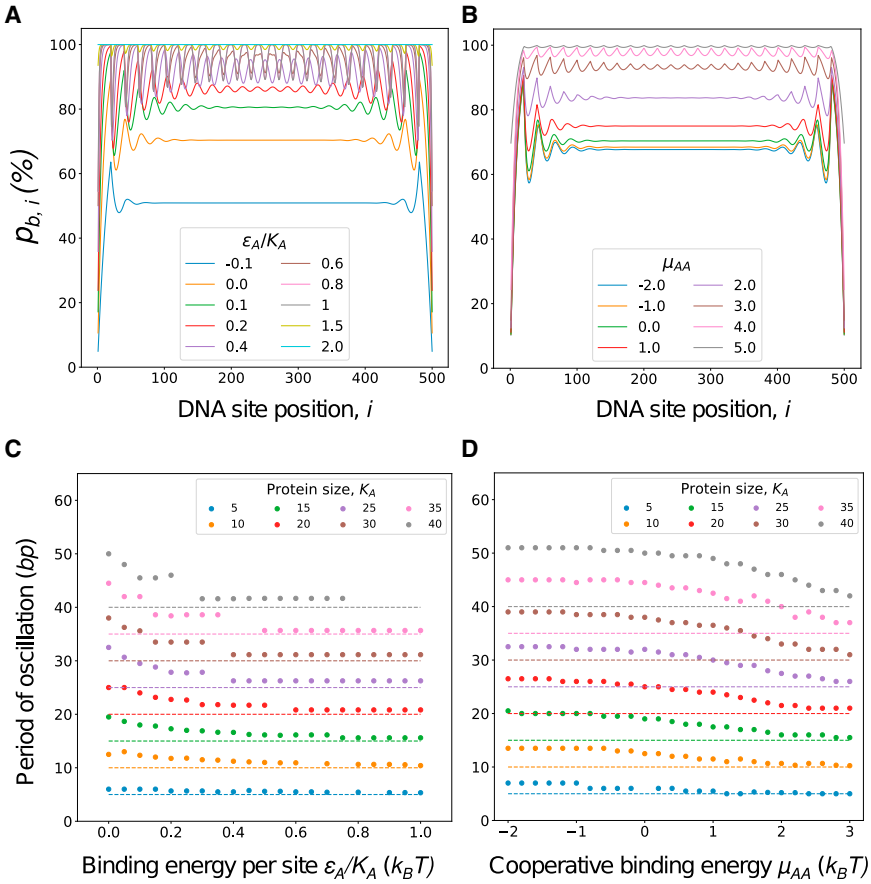

FIGURE 5 Effects of the binding free energy and cooperativity energy of the proteins on the profiles of site-dependent DNA occupancy patterns (*A* and *B*) and their periodicity (*C* and *D*). In all of the calculations, the size of the DNA molecule was $M = 500$ bp. For the graphs shown in (*A*) and (*B*), the protein binding site size was set equal to $K_A = 20$ bp. As for the proteins' binding free energy (ε_A) and cooperative binding energy (μ_{AA}), in (*A*) and (*C*), μ_{AA} was fixed at 0 k_BT value and ε_A/K_A was varied from -0.1 k_BT to 2.0 k_BT, whereas in (*B*) and (*D*), ε_A was fixed at 0 k_BT and μ_{AA} was varied from -2 k_BT to 5 k_BT. The periodicity of the oscillation patterns in (*C*) and (*D*) was determined by calculating the distance between neighboring valleys of the corresponding DNA occupancy probability curves (data points for which the standard deviation of the period of oscillations was larger than 10% were eliminated from the analysis). To see this figure in color, go online.

bound to DNA at positive DNA-binding cooperativity that support the assembly of nucleoprotein complexes on DNA.

Interestingly, in contrast to the positive DNA-binding cooperativity case, negative DNA-binding cooperativity of proteins has a largely negligible effect on the formation of protein positioning patterns, as can be seen from Fig. 4. The main reason behind this is the simple fact that negative cooperativity occludes proteins' binding only to neighboring DNA sites—i.e., only short DNA segments flanking DNA-bound proteins are protected against protein binding, whereas all the remaining DNA-binding sites can interact with proteins without any hindrance. Thus, only a very small number of DNA sites are sequestered from interaction with proteins. As a result, it can be concluded that only positive DNA-binding cooperativity of proteins has a strong effect on the formation of protein positioning patterns on DNA, whereas negative DNA-binding cooperativity of proteins has a negligible effect (see also Fig. 5 *B*).

Finally, Fig. 5, *C* and *D* provide a summary of the dependence of the oscillation period of the DNA occupancy graphs on the binding free energy and cooperative binding energy of proteins to DNA shown in Fig. 5, *A* and *B*. It can be seen from Fig. 5, *C* and *D* that the period of oscillations on average gradually increases above the binding site size of the protein with decreasing protein binding or cooperative binding energies, ε_A and μ_{AA}. Such a phenomenon is caused by a smaller density of proteins on DNA at lower values of the protein binding and/or cooperative binding energies, as well as by a stronger relative contribution of the proteins' positioning entropy to the total free energy of the system, which forces proteins to spread more or less uniformly along the DNA, increasing distance between them. These findings are in good agreement with previous theoretical studies, which predict that the period of nucleosome positioning on DNA in living cells reduces with increasing histone concentration (33).

Pattern formation by different competing DNA-binding proteins

Because experimental data show that in living cells, the chromosomal DNA is organized through cooperation between different types of DNA architectural proteins that frequently compete with each other for binding to DNA, we studied the role of such DNA-binding competition of proteins in pattern formation on DNA.

For the sake of simplicity, in our work we investigated the DNA-binding competition between two different types of proteins, which were designated as proteins A and B, by varying their binding and cooperative binding energies to

DNA as well as their binding site sizes to understand effects of these parameters on the emergence of protein positioning patterns on DNA.

First, we checked how DNA occupancy by nucleoprotein complexes is affected by the proteins' binding free energies. For a reference point, we used transfer-matrix calculations in which the binding site sizes of the proteins were set equal to $K_A = 10$ bp and $K_B = 25$ bp and the binding energies of the proteins were set at $\varepsilon_A = 0\ k_BT$ and $\varepsilon_B = 0\ k_BT$, respectively. Based on these energies, one might expect that there would be no difference in the probabilities of formation of A and B nucleoprotein complexes on DNA. However, as Fig. 6 shows, this is not the case; the smaller size protein A has a much higher probability to form nucleoprotein complexes on the DNA compared with the larger protein B, despite the fact that they have equal binding free energies to DNA per single bp.

This behavior of the proteins has a purely entropic origin. Indeed, a fixed-length DNA can host more complexes of smaller size, resulting in a larger number of DNA-protein states. According to the famous Boltzmann formula, this leads to a higher entropy of the system, meaning that DNA interacting with smaller proteins will have a smaller free energy in comparison to DNA interacting with larger proteins. This eventually results in favor of DNA interaction with the smaller protein A, which can be clearly seen from Fig. 6.

The above entropy-based effect can be partially counteracted by increase in the proteins' binding free energies. For example, if the binding free energies of proteins A and B are both set equal to 1 k_BT per DNA bp (i.e., $\varepsilon_A = 10\ k_BT$ and $\varepsilon_B = 25\ k_BT$), then, as can be seen from Fig. 7, the difference between the average DNA occupancies by the two proteins becomes much smaller than in the case of zero binding energies of the proteins (compare Fig. 6 to the *top row* of Fig. 7). Such an effect is not simply caused by the larger binding energy associated with the larger protein, considering the fact that when the DNA is fully covered by either the larger protein or the smaller protein, the total binding energy of the proteins is the same in both cases. Instead, it is created by interplay between the proteins' binding free energies and positioning entropies. Namely, at zero binding free energies of the proteins, the positioning entropy of the smaller protein A is sufficiently high to help it outcompete the larger protein B in occupying the DNA. However, at higher protein binding energies, DNA becomes heavily occupied by nucleoprotein complexes, leading to a strong decrease in the positioning entropy of both proteins due to the smaller number of possible DNA-protein conformations available at a high occupancy level of the DNA. As the contribution made by the positioning entropy into the total free energy of the DNA drops down, this results in more equal levels of the DNA occupancy by both types of proteins, A and B, as can be seen from the top row of Fig. 7.

Based on these results, one can predict that by further reducing the binding energy of the smaller protein A, the larger protein B will outcompete A in the formation of nucleoprotein complexes on DNA. This is indeed the case, as shown in Fig. 7. By decreasing the binding free energy (ε_A) of the smaller protein A, its domination in shaping the proteins' binding profile can be reversed such that protein B takes the lead, demonstrating a higher average binding probability to DNA. Respective changes in the Fourier spectra of the resulting protein patterns can be seen in Figs. S5–S7, showing how the pattern period switches from the smaller protein to a larger one as the DNA-binding competition becomes dominated by the latter.

Transfer-matrix calculations indicate that emergence of complex protein positioning patterns on DNA can be also enhanced by positive DNA-binding cooperation between proteins A and B, μ_{AB}. Indeed, as can be seen from Fig. 8, variation of μ_{AB} in the range from 0 k_BT to 2 k_BT leads to strong changes in the DNA occupancy level by

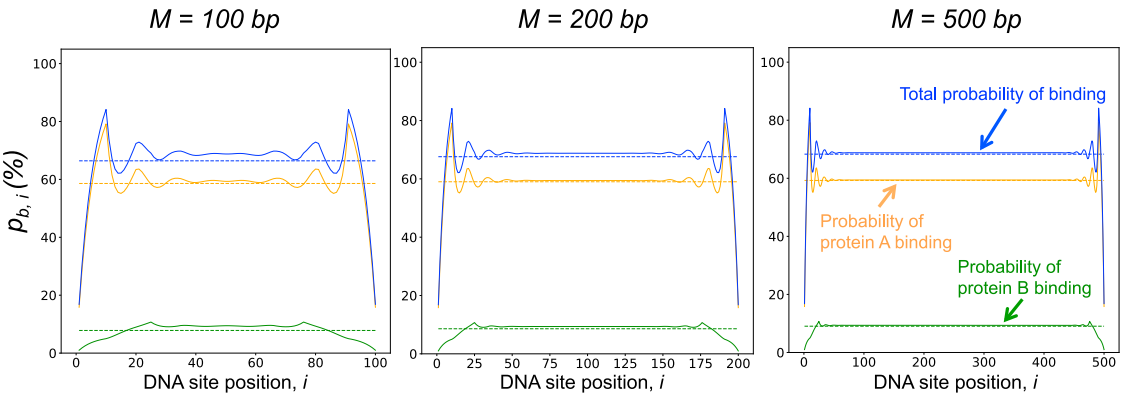

FIGURE 6 (*A–C*) Site-dependent occupancies of homogeneous 100, 200, and 500 bp DNA by two types of proteins, A (with $K_A = 10$ bp) and B (with $K_B = 25$ bp), that compete with each other for DNA-binding sites. In calculations, the protein binding energies and cooperative binding energies were all set to 0 ($\varepsilon_A = \varepsilon_B = \mu_{AA} = \mu_{AB} = \mu_{BB} = 0\ k_BT$). Solid curves in the graphs indicate site-dependent occupancies of DNA by each type of protein (*orange*, protein A; *green*, protein B; and *blue*, both proteins combined). Dashed lines show the average DNA occupancy level in each of the three cases. To see this figure in color, go online.

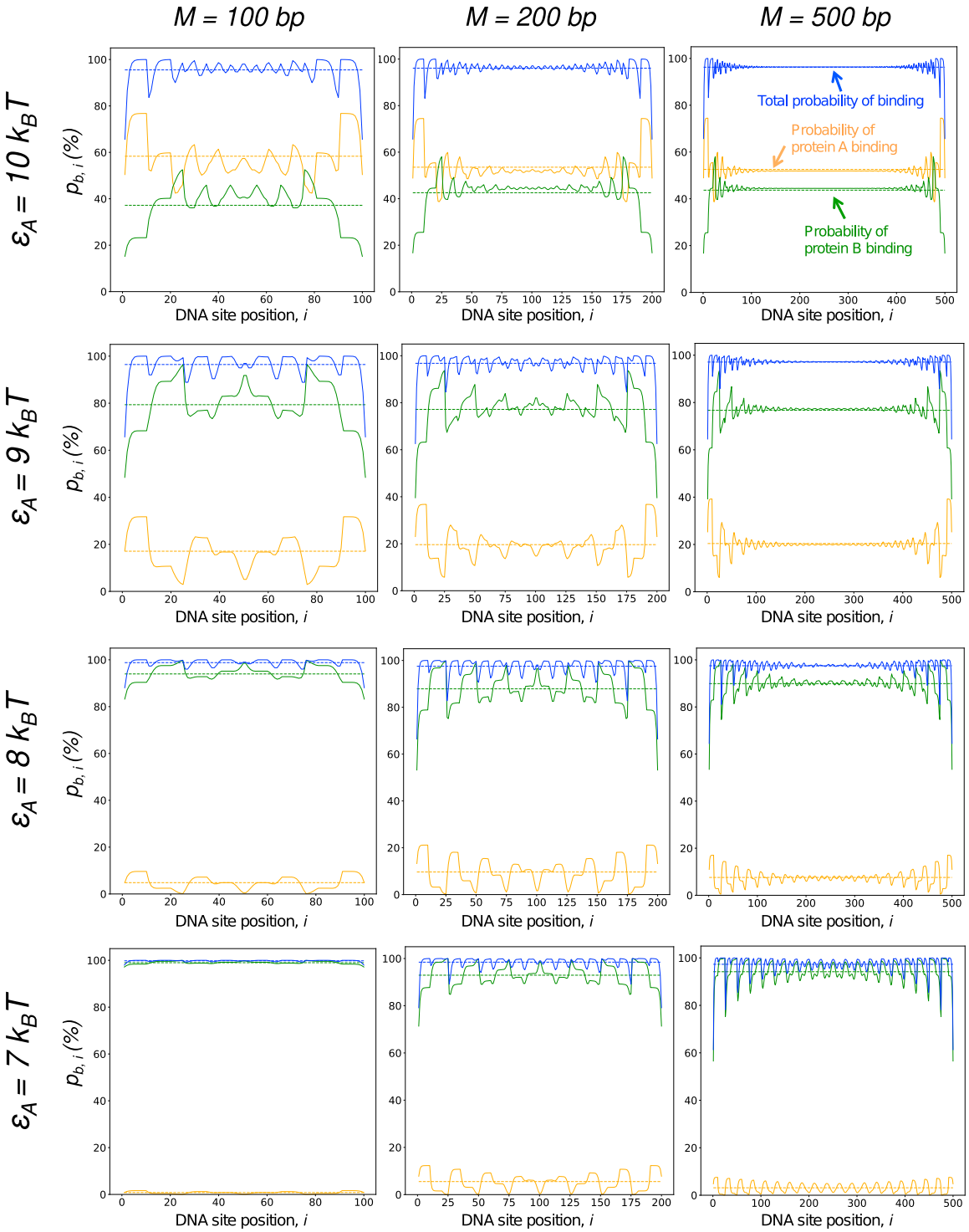

FIGURE 7 Site-dependent occupancies of homogeneous 100, 200, and 500 bp DNA by two types of proteins, A and B, that compete with each other for DNA binding. The binding site size of protein A was $K_A = 10$ bp, and the binding site size of protein B was $K_B = 25$ bp. The binding free energy of protein A (ε_A) was varied from 7 k_BT to 10 k_BT, whereas the binding free energy of protein B (ε_B) was kept constant at 25 k_BT. All of the cooperative binding energies between proteins A and B were set equal to 0: $\mu_{AA} = \mu_{AB} = \mu_{BB} = 0\ k_BT$. Solid curves in the graphs indicate site-dependent occupancies of DNA by each type of protein (*orange*, protein A; *green*, protein B; and *blue*, both proteins combined). Dashed lines show the average DNA occupancy level in each of the three cases. To see this figure in color, go online.

the two types of proteins for all tested sizes of DNA molecule—the higher the value of the DNA-binding cooperativity between the proteins, the more complex the protein positioning patterns on DNA and the larger the oscillation amplitude of these patterns. Furthermore, from the graphs it can be clearly seen that positive DNA-binding

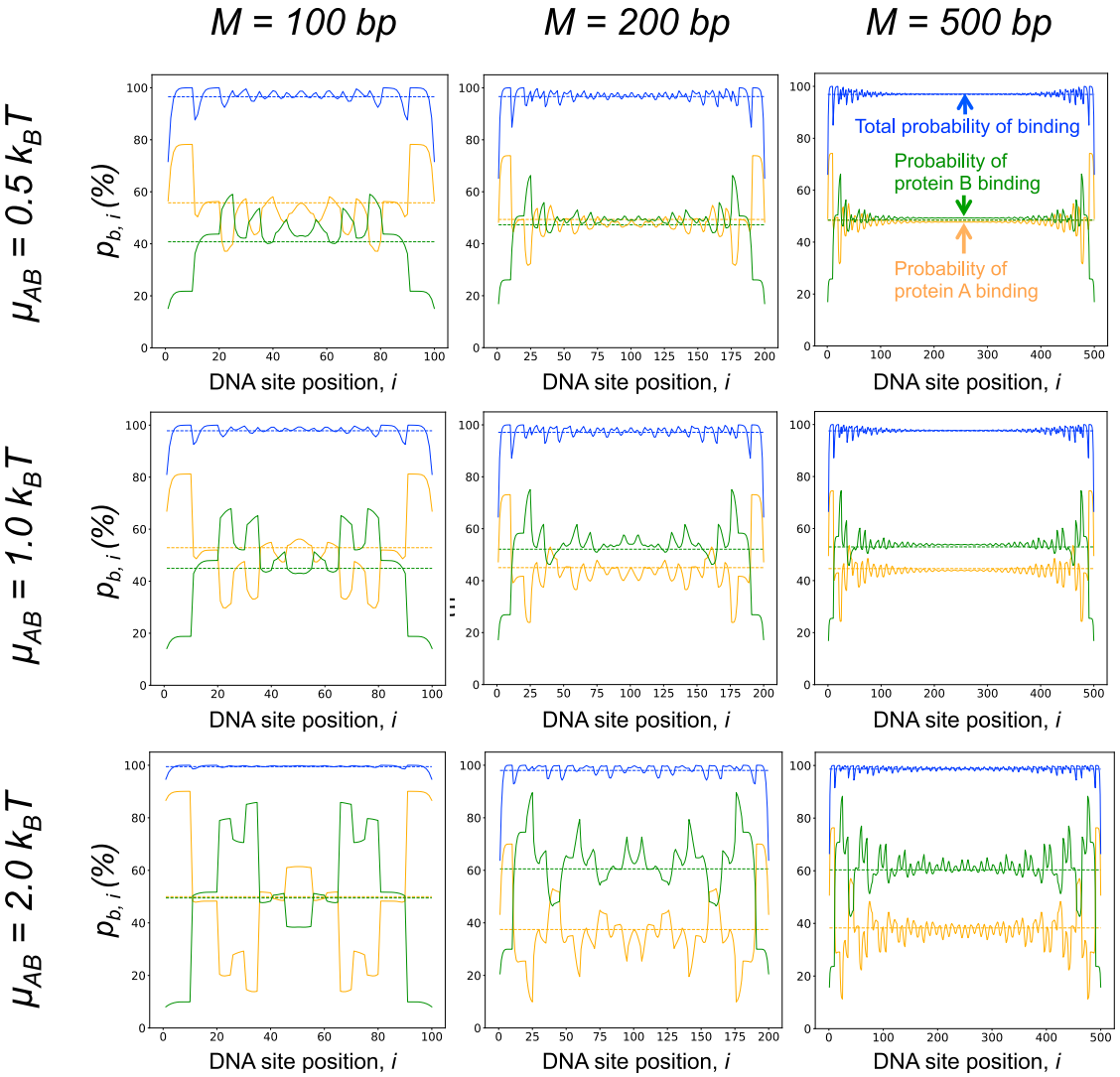

FIGURE 8 Site-dependent occupancies of homogeneous 100, 200, and 500 bp DNA by two type of proteins, A and B, at different levels of the proteins' binding cooperativity, μ_{AB}. In calculations, the binding site size of protein A was $K_A = 10$ bp and of protein B $K_B = 25$ bp. The protein binding energies were set to the same value of 1 k_BT per single DNA site (i.e., $\varepsilon_A = 10\ k_BT$ and $\varepsilon_B = 25\ k_BT$). The cooperative binding energy between proteins A and B, μ_{AB}, was varied from 0.5 to 2 k_BT; the rest of the cooperative binding energies were all set equal to 0: $\mu_{AA} = \mu_{BB} = 0\ k_BT$. Solid curves in the graphs indicate site-dependent occupancies of DNA by each type of protein (*orange*, protein A; *green*, protein B; and *blue*, both proteins combined). Dashed lines show the average DNA occupancy level in each of the three cases. To see this figure in color, go online.

cooperativity between the proteins also results in more spread patterns, spanning larger sections of the DNA.

In contrast, negative DNA-binding cooperativity between proteins A and B does not have any strong effect on the protein positioning pattern on DNA compared with the control case of $\mu_{AB} = 0\ k_BT$ (compare with Figs. S8 and S9). The main reason for such behavior is the same as the one discussed at the end of the previous section—negative DNA-binding cooperativity simply results in occlusion of the nucleoprotein complexes' formation by different types of proteins on neighboring DNA sites but does not have any long-range effect on protein binding to DNA.

Finally, to investigate in detail how the difference in the protein sizes affects the formation of protein positioning patterns on DNA, we varied the binding site size of protein B from 15 to 100 bp while keeping constant the rest of the model parameters in the calculations. As can be seen from Fig. 9 (for $M = 200$ bp) and Fig. S10 (for $M = 500$ bp), the most obvious effect of the protein size change is a gradual decrease of the mean DNA occupancy level by protein B, which again suggests that the entropy associated with the proteins' positioning on the DNA has a stronger effect on the DNA-binding competition between proteins with a large binding site size ratio. There is also another and perhaps more interesting effect—despite the fact that the DNA occupancy by larger protein B experiences a strong drop, it can be seen that this protein still has a strong effect on the pattern formation by smaller protein A. Specifically,

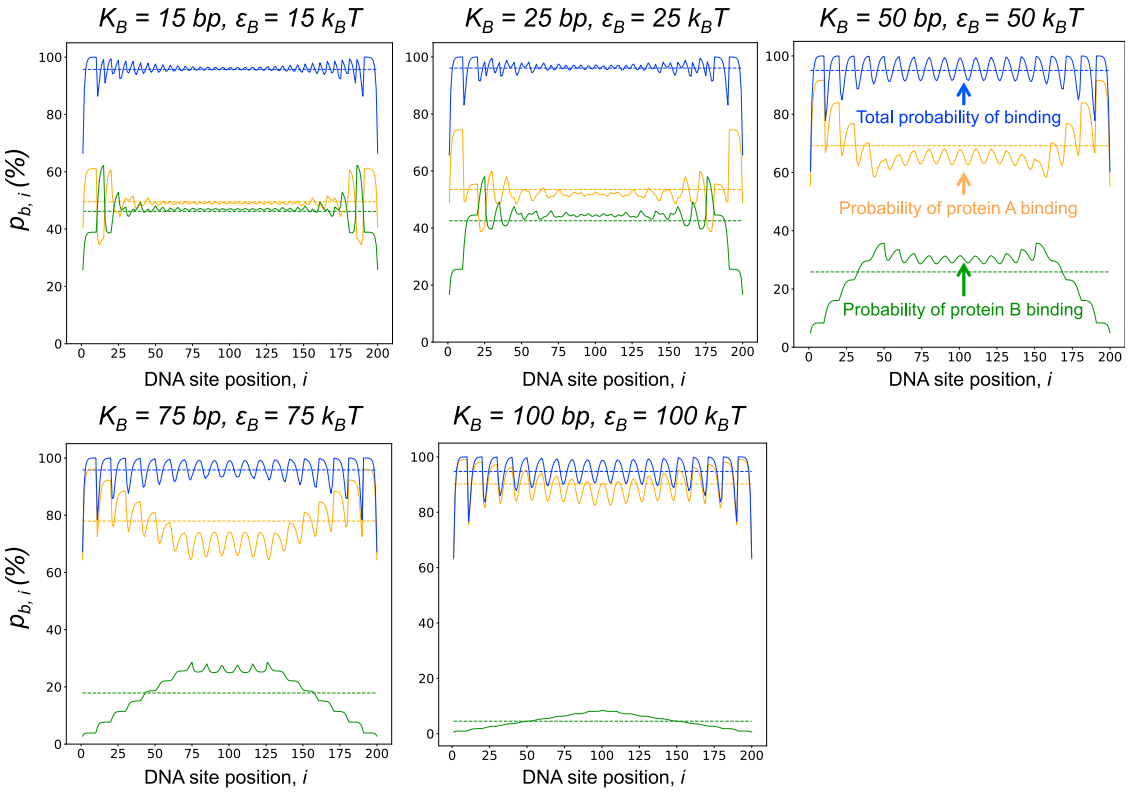

FIGURE 9 Site-dependent occupancies of homogeneous 200 bp DNA by two type of proteins, A and B, at different ratios of the protein sizes. In calculations, the binding site size of protein A, $K_A = 10$ bp, was kept constant, and the binding site size of protein B, K_B, was varied from 15 to 100 bp. In all cases, the protein binding energies were set to the same value of 1 k_BT per single DNA site; ε_A had a constant value of 10 k_BT, and ε_B was set equal to 1 $k_BT \times K_B$. All of the cooperative binding energies between proteins A and B were set equal to 0: $\mu_{AA} = \mu_{AB} = \mu_{BB} = 0$ k_BT. Solid curves in the graphs indicate site-dependent occupancies of DNA by each type of protein (*orange*, protein A; *green*, protein B; and *blue*, both proteins combined). Dashed lines show the average DNA occupancy level in each of the three cases. To see this figure in color, go online.

the protein oscillatory positioning pattern of A becomes more spread, spanning larger regions of DNA, and with increased amplitude of oscillation.

Thus, it can be concluded that not only the proteins' binding and cooperative binding energies but also the sizes of formed nucleoprotein complexes play very important roles in determining the shape of emerging protein positioning patterns on DNA. Although the mean occupancy of DNA by larger proteins may not seem to be significant, these proteins still play a significant role in pattern formation by smaller proteins.

DISCUSSION

In this study, we have investigated the role of a number of physical factors in formation of protein positioning patterns on DNA. It has been found that the size of DNA and nucleoprotein complexes, as well as DNA-binding competition and/or DNA-binding cooperativity between the proteins, has the strongest effect on the shape of such patterns. Interestingly, whereas the protein binding energy showed strong relation to the average DNA occupancy by proteins, small random perturbations in the protein binding energy that may arise from sequence heterogeneity of DNA demonstrated little effect on the main geometric characteristics of the pattern such as its oscillation period, indicating robustness of the molecular mechanisms responsible for the pattern formation.

The strong sensitivity of protein positioning patterns to DNA length found in this work may have several important implications for both in vitro and in vivo studies of DNA-protein interactions. One of the examples is footprint assays, which are typically used for identification of DNA sequence motifs that serve as specific binding sites for various types of DNA-binding proteins in in vitro experiments. The main idea of such assays is based on the mechanism of DNA protection against nuclease digestion by proteins covering DNA. Namely, after allowing proteins to bind to short pieces of DNA, the latter become exposed to finite-time digestion by one of the DNA nucleases, such as MNase or DNase I. The regions of DNA that are bound to proteins are usually less exposed to nuclease digestion and hence have a greater probability of survival in the digestion assay. As a result, sequencing of the surviving DNA pieces can be used to identify a specific binding sequence motif for a studied protein. Although footprint assays seem to be rather

straightforward, the results obtained in our study indicate that sequence motifs discovered by using such techniques may in fact incorrectly reflect the binding affinity of the studied protein to various DNA sequences, as the protein positioning on short DNA segments is strongly influenced by the finite-size DNA effect. As a result, footprint assays may lead to inaccurate prediction of binding sequence motifs for studied proteins, so their interpretation should be done with due care.

The same is also true for in vivo footprint assays, in which it is chromatin extracted from living cells that becomes digested by DNA proteases. Indeed, existing experimental data show that in eukaryotic cells, nucleosomes are packed in tight linear arrays, with only short stretches of nucleosome-free DNA of 20–90 bp length between them (sometimes called DNA linkers) that can host interactions with other DNA-binding proteins (34–36). Because nucleosomes have low spontaneous mobility on DNA unless being moved by nucleosome translocases (37), they can be considered as steric barriers over sufficiently long timescales for other proteins that bind to nucleosome-free DNA linkers. As a result, findings made in our study suggest that binding of small proteins to the short DNA linkers between adjacent nucleosomes should exhibit strong periodic positioning effects dependent on the length of the linkers. This may not only affect results obtained in in vivo footprint assays but also have potential influence on DNA organization, such as the formation of chromatosomes by binding of H1 histones to the DNA linkers near histone octamers (38).

Interestingly, our results show that in the presence of DNA-binding competition between two species of proteins, the smaller of the proteins dictates the positioning of both the smaller and the larger proteins in the case if both of the proteins have comparable DNA-binding free energies per single bp (see Figs. 6 and 7). Furthermore, transfer-matrix calculations also demonstrate that in the case when the binding energy of the smaller protein per single bp becomes less than that of the larger protein, the dominating role of the smaller protein in defining the protein position pattern on the DNA becomes lost (see Fig. 7). Applying these results to nucleosomes and other smaller DNA-binding proteins in eukaryotic cells, it can be seen that linear arrays of nucleosomes, which form on chromosomal DNA with a periodicity close to the nucleosome size, suggest that highly positively charged histone octamers possess much larger DNA-binding affinity than most of the other proteins competing with histones for DNA binding. By such a mechanism, nucleosomes form a general background by serving either as steric barriers or targeted protein complexes on which other DNA-binding proteins operate to provide high dynamic plasticity of DNA organization.

It should be noted that although in this study, we used the most basic model of DNA-protein interactions for transfer-matrix calculations, because of the high universality of the transfer-matrix formalism, it is not very hard to extend the model presented in this work to encompass even more complex features of DNA-protein interactions that have been discovered in recent experimental studies. For example, one of the limitations of this study is the assumption that proteins bind to DNA by forming complete nucleoprotein complexes, i.e., proteins fully engage with DNA by binding maximal possible amount of DNA bps. However, experiments show that this is not always the case, and proteins may engage with DNA to form partially bound intermediate states, which in general are energetically unfavorable under many experimental conditions but could become favorable at high protein concentrations or in the case when proteins have strong binding affinity to DNA. For example, in a recent experimental study (39), it has been shown that transcription factors can invade a nucleosome binding region containing a high-affinity binding site to the factors, forcing transition of the nucleosome into a partially unfolded conformation without full dissociation from DNA. In fact, such partially bound intermediate states of nucleoprotein complexes have been found to be rather widespread among DNA-binding proteins, providing a physical basis for a recently discovered effect of facilitated protein dissociation from DNA by other proteins freely diffusing in surrounding solution (40,41). It is not hard to show that such partially unfolded conformations of nucleoprotein complexes can be easily included into transfer-matrix calculations like those used in our study by introducing additional states of DNA-bound proteins into the model, see, for example, (27). As demonstrated in that study, partial binding of histone octamers to DNA leads to partial blurring of the histone binding patterns on DNA and results in greater DNA accessibility to other DNA-binding proteins, such as various transcription factors.

Finally, it should be noted that although in this study, we mainly focused on the description of physicochemical processes responsible for protein positioning on DNA, the same physical principles based on the tug of war between entropic and enthalpic energy terms will be applicable to understanding diverse biological features in other cellular systems in which steric constraints play the central role. Indeed, periodic binding patterns have been observed in several other biological systems, such as the striped pattern of collagen fibrils (42) or the periodic arrangement of myosin II and α-actinin along bundles of actin filaments (43), whose formation can be studied by using the same transfer-matrix approach as the one described in this work. Furthermore, it has been previously shown that by combining one-dimensional lattice models with the scaling approaches of polymer physics, it is possible to extend the application of transfer-matrix calculations from linear polymers to two-dimensional lipid membranes to gain insights into the physical processes guiding the binding and positioning of flexible peptides on the cell membrane (44), suggesting high universality of this theoretical method.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2021.03.016.

AUTHOR CONTRIBUTIONS

A.K.E. and J.Y. designed the research. L.S.M. carried out all simulations. L.S.M., A.K.E., and J.Y. analyzed the data. L.S.M., A.K.E., and J.Y. wrote the article.

ACKNOWLEDGMENTS

This work is supported by the Ministry of Education Academic Research Fund Tier 1, and the National Research Foundation, Prime Minister's Office, Singapore and the Ministry of Education under the Research Centres of Excellence programme.

REFERENCES

1. Teĭf, V. B., A. V. Shkrobkov, …, V. I. Krot. 2012. [Nucleosomes in gene regulation: theoretical approaches]. *Mol. Biol. (Mosk.).* 46:3–13.
2. Teif, V. B., and K. Rippe. 2010. Statistical-mechanical lattice models for protein-DNA binding in chromatin. *J. Phys. Condens. Matter.* 22:414105.
3. Ettig, R., N. Kepper, …, K. Rippe. 2011. Dissecting DNA-histone interactions in the nucleosome by molecular dynamics simulations of DNA unwrapping. *Biophys. J.* 101:1999–2008.
4. Portella, G., F. Battistini, and M. Orozco. 2013. Understanding the connection between epigenetic DNA methylation and nucleosome positioning from computer simulations. *PLoS Comput. Biol.* 9:e1003354.
5. Parmar, J. J., J. F. Marko, and R. Padinhateeri. 2014. Nucleosome positioning and kinetics near transcription-start-site barriers are controlled by interplay between active remodeling and DNA sequence. *Nucleic Acids Res.* 42:128–136.
6. Schöpflin, R., V. B. Teif, …, G. Wedemann. 2013. Modeling nucleosome position distributions from experimental nucleosome positioning maps. *Bioinformatics.* 29:2380–2386.
7. Mahony, S., and B. F. Pugh. 2015. Protein-DNA binding in high-resolution. *Crit. Rev. Biochem. Mol. Biol.* 50:269–283.
8. Carone, B. R., J.-H. Hung, …, O. J. Rando. 2014. High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Dev. Cell.* 30:11–22.
9. Kerrigan, L. A., and J. T. Kadonaga. 1992. Periodic binding of individual core histones to DNA: inadvertent purification of the core histone H2B as a putative enhancer-binding factor. *Nucleic Acids Res.* 20:6673–6680.
10. Jain, S. S., and T. D. Tullius. 2008. Footprinting protein-DNA complexes using the hydroxyl radical. *Nat. Protoc.* 3:1092–1100.
11. Beckwitt, E. C., M. Kong, and B. Van Houten. 2018. Studying protein-DNA interactions using atomic force microscopy. *Semin. Cell Dev. Biol.* 73:220–230.
12. Milani, P., G. Chevereau, …, A. Arneodo. 2009. Nucleosome positioning by genomic excluding-energy barriers. *Proc. Natl. Acad. Sci. USA.* 106:22257–22262.
13. Pisano, S., E. Pascucci, …, M. Savino. 2006. AFM imaging and theoretical modeling studies of sequence-dependent nucleosome positioning. *Biophys. Chem.* 124:81–89.
14. Zhao, X., S. Guo, …, J. Yan. 2019. Single-molecule manipulation quantification of site-specific DNA binding. *Curr. Opin. Chem. Biol.* 53:106–117.
15. Hall, M. A., A. Shundrovsky, …, M. D. Wang. 2009. High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. *Nat. Struct. Mol. Biol.* 16:124–129.
16. Li, M., and M. D. Wang. 2012. Unzipping single DNA molecules to study nucleosome structure and dynamics. *Methods Enzymol.* 513:29–58.
17. Trifonov, E. N. 1980. Sequence-dependent deformational anisotropy of chromatin DNA. *Nucleic Acids Res.* 8:4041–4053.
18. Segal, E., Y. Fondufe-Mittendorf, …, J. Widom. 2006. A genomic code for nucleosome positioning. *Nature.* 442:772–778.
19. Field, Y., N. Kaplan, …, E. Segal. 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput. Biol.* 4:e1000216.
20. Kaplan, N., I. K. Moore, …, E. Segal. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature.* 458:362–366.
21. Ioshikhes, I. P., I. Albert, …, B. F. Pugh. 2006. Nucleosome positions predicted through comparative genomics. *Nat. Genet.* 38:1210–1215.
22. Gaffney, D. J., G. McVicker, …, J. K. Pritchard. 2012. Controls of nucleosome positioning in the human genome. *PLoS Genet.* 8:e1003036.
23. Jiang, C., and B. F. Pugh. 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* 10:161–172.
24. Möbius, W., and U. Gerland. 2010. Quantitative test of the barrier nucleosome model for statistical positioning of nucleosomes up- and downstream of transcription start sites. *PLoS Comput. Biol.* 6:e1000891.
25. Buschle, J., P. Maass, and W. Dieterich. 2000. Wall-induced density profiles and density correlations in confined Takahashi lattice gases. *J. Stat. Phys.* 99:273–312.
26. Kornberg, R. D., and L. Stryer. 1988. Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.* 16:6677–6690.
27. Teif, V. B., R. Ettig, and K. Rippe. 2010. A lattice model for transcription factor access to nucleosomal DNA. *Biophys. J.* 99:2597–2607.
28. Teif, V. B. 2007. General transfer matrix formalism to calculate DNA-protein-drug binding in gene regulation: application to OR operator of phage lambda. *Nucleic Acids Res.* 35:e80.
29. Kim, H. D., and E. K. O'Shea. 2008. A quantitative model of transcription factor-activated gene expression. *Nat. Struct. Mol. Biol.* 15:1192–1198.
30. Efremov, A. K., and J. Yan. 2018. Transfer-matrix calculations of the effects of tension and torque constraints on DNA-protein interactions. *Nucleic Acids Res.* 46:6504–6527.
31. Harrison, S. C., and A. K. Aggarwal. 1990. DNA recognition by proteins with the helix-turn-helix motif. *Annu. Rev. Biochem.* 59:933–969.
32. Afek, A., J. L. Schipper, …, D. B. Lukatsky. 2014. Protein-DNA binding in the absence of specific base-pair recognition. *Proc. Natl. Acad. Sci. USA.* 111:17140–17145.
33. Beshnova, D. A., A. G. Cherstvy, …, V. B. Teif. 2014. Regulation of the nucleosome repeat length in vivo by the DNA sequence, protein concentrations and long-range interactions. *PLoS Comput. Biol.* 10:e1003698.
34. Clark, D. J. 2010. Nucleosome positioning, nucleosome spacing and the nucleosome code. *J. Biomol. Struct. Dyn.* 27:781–793.
35. Mavrich, T. N., I. P. Ioshikhes, …, B. F. Pugh. 2008. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.* 18:1073–1083.
36. Szerlong, H. J., and J. C. Hansen. 2011. Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. *Biochem. Cell Biol.* 89:24–34.
37. Pennings, S., G. Meersseman, and E. M. Bradbury. 1991. Mobility of positioned nucleosomes on 5 S rDNA. *J. Mol. Biol.* 220:101–110.

38. Hu, J., L. Gu, …, C. Jiang. 2018. Dynamic placement of the linker histone H1 associated with nucleosome arrangement and gene transcription in early Drosophila embryonic development. *Cell Death Dis.* 9:765.

39. Luo, Y., J. A. North, …, M. G. Poirier. 2014. Nucleosomes accelerate transcription factor dissociation. *Nucleic Acids Res.* 42:3017–3027.

40. Kamar, R. I., E. J. Banigan, …, J. F. Marko. 2017. Facilitated dissociation of transcription factors from single DNA binding sites. *Proc. Natl. Acad. Sci. USA.* 114:E3251–E3257.

41. Erbaş, A., and J. F. Marko. 2019. How do DNA-bound proteins leave their binding sites? The role of facilitated dissociation. *Curr. Opin. Chem. Biol.* 53:118–124.

42. Hafner, A. E., N. G. Gyori, …, A. Šarić. 2020. Modeling fibrillogenesis of collagen-mimetic molecules. *Biophys. J.* 119:1791–1799.

43. Hu, S., K. Dasbiswas, …, A. D. Bershadsky. 2017. Long-range self-organization of cytoskeletal myosin II filament stacks. *Nat. Cell Biol.* 19:133–141.

44. Teif, V. B., D. Harries, …, A. Ben-Shaul. 2008. Matrix formalism for site-specific binding of unstructured proteins to multicomponent lipid membranes. *J. Pept. Sci.* 14:368–373.