A Lattice Model for Transcription Factor Access to Nucleosomal DNA

Vladimir B. Teif,* Ramona Ettig, and Karsten Rippe*
BioQuant and German Cancer Research Center, Heidelberg, Germany

ABSTRACT Nucleosomes, the basic repeating unit of chromatin, consist of 147 basepairs of DNA that are wrapped in almost two turns around a histone protein octamer core. Because ~3/4 of the human genomic DNA is found within nucleosomes, their position and DNA interaction is an essential determinant for the DNA access of gene-specific transcription factors and other proteins. Here, a DNA lattice model was developed for describing ligand binding in the presence of a nucleosome. The model takes into account intermediate states, in which DNA is partially unwrapped from the histone octamer. This facilitates access of transcription factors to up to 60 DNA basepairs located in the outer turn of nucleosomal DNA, while the inner DNA turn was found to be more resistant to competitive ligand binding. As deduced from quantitative comparisons with recently published experimental data, our model provides a better description than the previously used all-or-none lattice-binding model. Importantly, nucleosome-occupancy maps predicted by the nucleosome-unwrapping model also differed significantly when partial unwrapping of nucleosomal DNA was considered. In addition, large effects on the cooperative binding of transcription factors to multiple binding sites occluded by the nucleosome were apparent. These findings indicate that partial unwrapping of DNA from the histone octamer needs to be taken into account in quantitative models of gene regulation in chromatin.

INTRODUCTION

The eukaryotic genome is organized by histone proteins in a supramolecular nucleoprotein complex referred to as chromatin (1). The DNA sequence can be represented as a one-dimensional lattice encoding for genes that are expressed in a tightly controlled manner that reflects certain cellular states. This process is controlled by the context-dependent binding of regulatory molecules that recognize features of the DNA sequence as well as posttranslational histone modifications and the spatial conformation of chromatin. Many mathematical descriptions of gene regulation are performed in the framework of one-dimensional DNA lattice models, which are conceptually based on the assumptions of Markov (2) or Ising (3) models. In this type of model, a molecular system is divided into elementary units (e.g., basepairs for DNA) and is described by multiple states associated with each unit, e.g., bound/unbound (4–20). The states of the whole system are then represented as different combinations of states of the elementary units. Lattice models commonly consider a protein as a single entity, which can bind a fixed number of DNA sequence units. Different approaches for the description of protein-DNA binding in the context of chromatin have been reviewed recently (21). For simplicity it is usually assumed that a protein can be either bound or not, and that intermediate states are absent. If, for example, a protein covers m nucleotides upon binding to the n^{th} DNA site, all nucleotides from n to $n + m - 1$ are excluded from binding to other proteins. However, in reality, DNA-binding does not follow a simple all-or-none binding mechanism but, instead, proceeds via intermediate states. These may provide the molecular basis for induced-fit recognition and assembly of multisubunit complexes. For example, the human RNA polymerase I transcription factor UBF has several HMG domains. These bind DNA progressively (depending on the DNA distortion induced by already bound HMG domains) and increase DNA bending further (22).

Here, we will focus on the basic building block of chromatin, the nucleosome, for which intermediate binding states are critically involved in its biological function. The nucleosome consists of a 147-bp-long DNA segment that is wrapped around a protein core consisting of eight histones, with two copies each of the histones H2A, H2B, H3, and H4. The nucleosome structure is dynamic and regulates gene expression by modulating DNA accessibility. To mediate access of the nucleosomal DNA for the binding of transcription factors (TFs), nucleosomes are either translocated along the DNA by chromatin remodeling complexes (4,13,23,24), or a partial unwrapping of the DNA from the histone octamer core occurs (25–35) (Fig. 1). Several mechanistic models to describe nucleosome unwrapping have been proposed during the last few years in the context of force spectroscopy experiments (36–39). However, these descriptions do not directly apply to in vivo processes where unwrapping of DNA may occur spontaneously in the absence of a directed force.

Here, we have developed (to our knowledge) the first lattice model in the framework of Ising-Markov approaches that allows treating intermediate states of protein binding to the DNA. The model provides a general approach to DNA-binding problems ranging from simple polyamines to transcription factors and nucleosomes. After formulating it as an extension of the previously suggested transfer matrix formalism (13,40,41), the model is applied to DNA unwrapping from the nucleosome. From a comparison of our

Submitted June 4, 2010, and accepted for publication August 13, 2010.

*Correspondence: vladimir.teif@bioquant.uni-heidelberg.de or karsten.rippe@bioquant.uni-heidelberg.de

Editor: Laura Finzi.

0006-3495/10/10/2597/11 $2.00

doi: 10.1016/j.bpj.2010.08.019

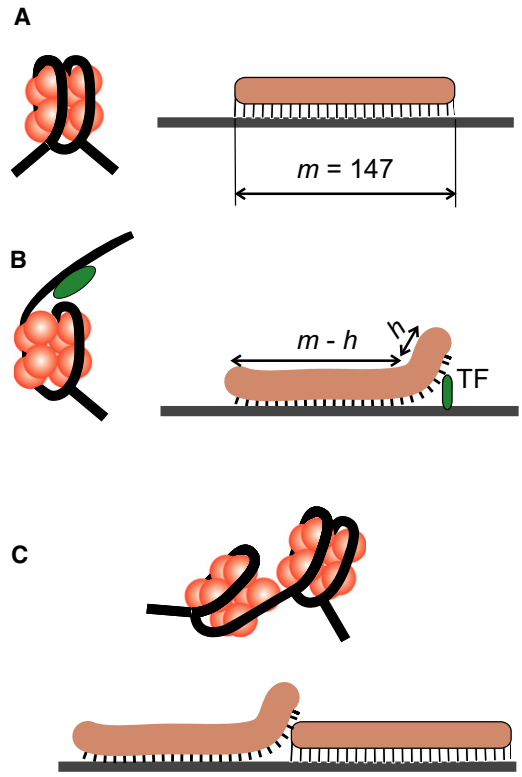

FIGURE 1 A lattice model for unwrapping DNA from the nucleosome. (*A*) Lattice model for the interactions of the DNA with the histone octamer core. The histone core is considered as an extended ligand, which covers $m = 147$ bp upon interactions of the DNA with the protein. If unwrapping is considered, a fraction of interactions at both ends of the protein-covered DNA region can be disrupted. (*B*) A transcription factor can bind the DNA partially unwrapped from the histone octamer. The unwrapping of h basepairs is depicted in the lattice model as unbinding of h corresponding protein-DNA contacts out of a total of 147 bonds. (*C*) If nucleosome unwrapping is allowed, nucleosomes can invade each other's binding sites.

calculations with the available in vitro and in vivo experiments, it is suggested that unwrapping DNA from the nucleosome facilitates access of transcription factors to the DNA sequence situated close to the region where the DNA enters/exits the nucleosome. This approach is extended to evaluate the nucleosome-mediated cooperativity of TF binding. In addition, it is concluded that partial unwrapping of individual nucleosomes can also affect large-scale genomic nucleosome-occupancy maps. Because our model includes conventional all-or-none binding as a particular case, we suggest that it might be considered as a general extension of lattice models currently used in genomic studies for a more faithful description of TF interactions with chromatin.

MATERIALS AND METHODS

Lattice model calculations

Calculation of nucleosome unwrapping, TF-nucleosome competition, and the maps of TF and nucleosome occupancies on the DNA were performed as described in the Theory section.

Molecular dynamics simulation

The interaction enthalpy for each basepair in the nucleosome was calculated using the molecular dynamics (MD) trajectory of a nucleosome simulated for ~20 ns at 1 bar pressure, 300°K, in explicit water with 150 mM NaCl. The simulations and analysis were performed with software package NAMD (Ver 2.6) (42) and the NAMD Energy plugin provided by the visualization software VMD (Ver 1.8.6) (43), respectively. The electrostatic and van der Waals interactions of the 147 single basepairs with histone proteins were calculated every 20 picoseconds, resulting in 956 snapshots of the nucleosome trajectory, which were averaged to determine the average interaction enthalpies for each basepair.

THEORY

We consider reversible binding of proteins of f types to the DNA with the DNA basepairs numbered by index $n = (1, N)$. A protein of type $g = (1, f)$ covers up to m_g lattice units upon binding to the DNA. The value m_g corresponds to the length of the DNA binding site in the conventional all-or-none model (41). However, in our approach, intermediate binding states with effective binding site length $<m_g$ are also allowed. Thus, for a binding event, it is not necessary that all interaction sites of the protein are in contact with the DNA (e.g., one or more protein-DNA contacts may remain in the unbound state). This is represented in the model by splitting the protein-DNA interactions into discrete contacts, one per each DNA lattice unit, which are characterized by their microscopic binding constants. The microscopic binding constants $K(n,g,h)$ assigned for each DNA unit $h = (1, m_g)$ covered by the protein are chosen such that their product yields the macroscopic binding constant $K(n,g)$ for the whole protein binding to its target DNA sequence $(n, n + m_g - 1)$. In the calculations performed below, we define the elementary DNA unit to be one basepair. In the initial calculations, the microscopic protein-DNA binding constant for each lattice unit is given as

$$K(n,g,h) = \sqrt[m_g]{K(n,g)},$$

i.e., interactions are equally distributed over all protein-DNA contacts. Then a potential with local variations of the binding energy is derived here from MD simulations and compared to the homogeneous potential.

In the dilute-solution approximation, protein binding decreases its entropy by

$$R \times \ln [c_0(g)],$$

where $c_o(g)$ is the concentration of free protein of type g. Accordingly, bound protein states are characterized by additional statistical weights $c_o(g)$ while unbound states have weight 1. In the computational model, we assign the weight $c_o(g)$ to the first protein-DNA contact. Thus, the entropic cost for removing the unbound protein from the bulk solution to the DNA is independent of the number of protein-DNA contacts ($c_o(g)$ per protein), while the energy contribution from protein-DNA interactions increases with the number of formed contacts with a value of $K(n,g,h)$

per contact. The energetic gain for an individual protein is maximal when all m_g contacts are formed. On the other hand, incomplete binding leaves more space for the binding of other proteins to the DNA lattice. Thus, the final combinatorial optimization is not trivial and needs to be calculated.

For the nucleosome, incomplete binding corresponds to the unwrapping of the DNA from the histone octamer protein core. The initiation of unwrapping might be assigned an additional weight *Unwrap*, which may depend on the protein type g and the length of unwrapped DNA h: $Unwrap = Unwrap(h,\ g)$. In the calculations reported here, we assume that there are no specific obstacles for unwrapping, i.e., $Unwrap = 1$. Setting $Unwrap = 0$ would make this model equivalent to the all-or-none binding where partial unwrapping is prohibited. Thus, our model can be viewed as a generalization of protein-DNA binding models, including classical all-or-none binding as a specific case. To derive a general description, we also consider the possibility that proteins bound to the DNA interact with each other. This interaction is characterized by a parameter $w(l, g_1, g_2)$ that represents the interaction potential between the proteins of type g_1 and g_2 separated by the distance l. At $l = 0$, w is equivalent to the McGhee-von Hippel contact cooperativity parameter (17,44). Each protein is characterized by the maximum interaction distance V_g. Proteins do not interact at distances larger than V_g, $w\ (l > V_g,\ g_1,\ g_2) = 1$. Anticooperativity corresponds to $w = 0$. For example, one can formulate the condition that nucleosomes cannot come closer than 10 bp to each other as $w(l \le 10, g_1, g_2) = 0$. In the calculations reported below, long-range interactions are not taken into account, and protruding of the protein's DNA interaction site beyond the end of the DNA lattice is prohibited. Furthermore, a steric overlap of two proteins is not allowed unless one of them is bound to the DNA only partially. In the latter case, unbound DNA basepairs belonging to the binding site of the first protein can be involved in DNA-binding of the second protein (Fig. 1 and Fig. S1 in the Supporting Material).

Following previous works (12,13,40,41), all microstates allowed for each individual DNA unit (basepair) are enumerated (Table 1). The transfer matrices contain conditional probabilities $Q_n(i, j)$ of having unit n in state i and unit $(n + 1)$ in state j. The following algorithm allows assigning nonzero elements of the transfer matrix, taking into account intermediate bound states for large proteins ($m_g > 1$):

1. First unit of g-type protein followed by the second unit:

$$1 \le g \le f, i = \sum_{k=1}^{g-1} m_k + 1, j = i + 1: \begin{cases} n = 1: Q_n(i,j) = K(n,g,1) \times c_0(g), \\ 1 < n \le N - \left(m_g - 1\right): Q_n(i,j) = K(n,g,1). \end{cases} \quad (1)$$

2. Bound unit of g-type protein, not at the ends of the protein, followed by another bound unit (if $m > 2$):

$$1 \le g \le f, i = \sum_{k=1}^{g-1} m_k + 1 + h, j = i + 1, 1 < h \le m_g - 2, \\ h < n \le N - \left(m_g - h - 1\right): Q_n(i,j) = K(n,g,h+1). \quad (2)$$

3. Bound unit $(m_g - h)$ of g-type protein followed by a right free DNA end (h protein units are unbound):

$$1 \le g \le f, i = \sum_{k=1}^{g} m_k - h, 0 \le h \le m_g, \\ j = \sum_{g=1}^{f} m_g + 2, m_g - h \le n \le n - h: \quad (3)$$
$$\begin{cases} n = 1: Q_n(i,j) = K\left(n,g,m_g - h\right) \times c_0(g) \times Unwrap(h,g), \\ n > 1:\ Q_n(i,j) = K\left(n,g,m_g - h\right) \times Unwrap(h,g). \end{cases}$$

4. Bound unit $(m_{g1} - h_1)$ of g_1-type protein followed by unit 1 of g_2-protein (no gap between proteins; h_1 units of g_1-protein are unbound):

$$1 \le g_1 \le f, i = \sum_{k=1}^{g_1} m_k - h_1, 0 \le h_1 < m_{g1}, \\ 1 \le g_2 \le f, j = \sum_{k=1}^{g_2-1} m_k + 1,\ m_{g1} - h_1 \le n \le N - m_{g2}: \\ \begin{cases} n = 1:\ Q_n(i,j) = K\left(n,g_1,m_{g1} - h_1\right) \times w(0,g_1,g_2) \times \\ \quad Unwrap(h_1,g_1) \times c_0(g_1) \times c_0(g_2), \\ n > 1:\ Q_n(i,j) = K\left(n,g_1,m_{g1} - h_1\right) \times w(0,g_1,g_2) \times \\ \quad Unwrap(h_1,g_1) \times c_0(g_1). \end{cases} \quad (4)$$

5. Last unit of g_1-type protein followed by unit h_2+1 of g_2-protein (no gap between proteins; h_2 units of g_2-protein are unbound):

TABLE 1 State enumeration for a DNA unit

State number	State description
1	First unit of type 1 protein
…	…
m_1	Last unit of type 1 protein
…	…
$\sum_{1}^{f-1} m_g + 1$	First unit of type f protein
…	…
$\sum_{1}^{f} m_g$	Last unit of type f protein
$\sum_{g=1}^{f} m_g + 1$	Left free DNA end
$\sum_{g=1}^{f} m_g + 2$	Right free DNA end
…	…
$\sum_{g=1}^{f} m_g + 2 + \sum_{g=1}^{g_2-1} V_g + l$	g_1-l-g_2 gap (l free units before next g_2 protein), $l \le V_{g2}$
…	…
$\sum_{g=1}^{f} (m_g + V_g) + 2 + l$	g_1-l-g_2 gap (l free units before next g_2 protein), $l > V_{g2}$
…	…
$\sum_{g=1}^{f} (m_g + V_g) + 2 + \max(V_g) + 1$	Free unit out of protein-protein interactions, not at the DNA ends

$$1 \le g1 \le f;\ i = \sum_{k=1}^{g_1} m_k,\ 1 \le g_2 \le f,\ j = \sum_{k=1}^{g_{2-1}} m_k + 1 + h_2,$$
$$1 \le h_2 < m_{g2}, m_{g1} \le n \le N - \left(m_{g2} - h_2\right):$$
$$Q_n(i,j) = K\left(n, g_1, m_{g1}\right) \times w(0, g_1, g_2) \times Unwrap(h_2, g_2) \times c_0(g_2). \quad (5)$$

6. Left free DNA end continues:

$$i = j = \sum_{g=1}^{f} m_g + 1 : Q_n(i,j) = w(0,0,0) = 1. \quad (6)$$

7. Right free DNA end continues:

$$i = j = \sum_{g=1}^{f} m_g + 2, n > 1 : Q_n(i,j) = w(0,0,0) = 1. \quad (7)$$

8. Left free DNA end followed by bound unit $h+1$ of g-type protein (h unbound protein units):

$$1 \le g \le f,\ i = \sum_{g=1}^{f} m_g + 1, j = \sum_{1}^{g-1} m_k + 1 + h,$$
$$0 \le h < m_g, n \le N - \left(m_g - h\right):$$
$$Q_n(i,\ j) = w(0,\ 0, g) \times Unwrap(h, g) \times c_0(g). \quad (8)$$

9. Bound unit ($m_g - h$) of g-type protein followed by a non-interacting gap longer V_g (h unbound protein units):

$$1 \le g \le f, i = \sum_{1}^{g} m_k - 1 - h, 0 \le h < m_g;$$
$$j = \sum_{g=1}^{f} \left(m_g + V_g\right) + 2, m_g - h \le n < N:$$
$$\begin{cases} n = 1:\ Q_n(i,j) = w\left(V_g + 1, g, 0\right) \times K\left(n, g, m_g - h\right) \\ \quad \times Unwrap(h, g) \times c_0(g), \\ n > 1:\ Q_n(i,j) = w\left(V_g + 1, g, 0\right) \times K\left(n, g, m_g - h\right) \\ \quad \times Unwrap(h, g). \end{cases} \quad (9)$$

10. Large noninteracting gap continues, units before $max(V_g)$:

$$i = \sum_{g=1}^{f} \left(m_g + V_g\right) + 2, 1 \le k \le max\left(V_g\right),$$
$$j = i + k, : Q_n(j, j+1) = 1. \quad (10)$$

11. Large noninteracting gap continues, units after max (V_g):

$$i = j = \sum_{g=1}^{f} \left(m_g + V_g\right) + 2 + max\left(V_g\right) + 1,$$
$$1 < n < N : Q_n(i,j) = 1. \quad (11)$$

12. Large noninteracting gap followed by bound unit $h+1$ of g-type protein (h unbound protein units):

$$i = \sum_{g=1}^{f} \left(m_g + V_g\right) + 2 + max\left(V_g\right) + 1,$$
$$j = \sum_{1}^{g-1} m_k + 1 + h, 0 \le h < m_g, 1 < n \le N - \left(m_g - h_2\right): \quad (12)$$
$$Q_n(i,j) = Unwrap(h, g) \times c_0(g).$$

13. Unit ($m_{g1} - h$) of g_1-type protein followed by l-unit gap followed by g_2-protein:

$$i = \sum_{k=1}^{g_1} m_k - h, 0 \le h < m_{g1},\ j = \sum_{g=1}^{f} m_g + 2 + \sum_{g=1}^{g_2-1} V_g + l,$$
$$1 \le l \le V_{g2} \text{ for } V_{g2} > 0, m_{g1} \le n \le N:$$
$$\begin{cases} n = 1:\ Q_n(i,j) = w(l, g_1, g_2) \times K\left(n, g_1, m_{g1} - h\right) \\ \quad \times Unwrap(h, g_1) \times c_0(g_1), \\ n > 1:\ Q_n(i,j) = w(l, g_1, g_2) \times K\left(n, g_1, m_{g1} - h\right) \\ \quad \times Unwrap(h, g_1). \end{cases} \quad (13)$$

14. g_1-l-g_2 gap continues (l free units before g_2-type protein):

$$2 \le l \le V_{g2} \text{ for } V_{g2} > 0,\ i = \sum_{g=1}^{f} m_g + 2 + \sum_{g=1}^{g_2-1} V_g + l,$$
$$j = i - 1,\ 1 \le n \le N : Q_n(i,j) = 1. \quad (14)$$

15. g_1-l-g_2 gap followed by bound unit $h + 1$ of g_2-type protein:

$$i = \sum_{g=1}^{i} m_g + 2 + \sum_{g=1}^{g_2-1} V_g,\ j = \sum_{1}^{g_2-1} m_k + 1 + h,$$
$$0 \le h < m_{g2}, 1 < n < N - m_{g2} - h:$$
$$Q_n(i,\ j) = Unwrap(h,\ g_2) \times c_0(g_2). \quad (15)$$

Following standard mathematical notations, in all equations above we assume that the sum equals 0 in the case that the upper summation limit is less than the lower limit.

The partition function Z and its derivatives are calculated recursively, according to Eqs. 16 and 17 (18):

$$Z = A_N \times \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix},$$
$$A_i = A_{i-1} \times Q_n,\ A_0 = (1\ \ 1\ \ \dots\ \ 1), \quad (16)$$

$$\frac{\partial Z}{\partial K(n,g)} = \frac{\partial A_N}{\partial K(n,g)} \times \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix},$$
$$\frac{\partial A_n}{\partial K(n,g)} = \frac{\partial A_{n-1}}{\partial K(n,g)} \times Q_n + A_{n-1} \times \frac{\partial Q_n}{\partial K(n,g)}, \quad (17)$$
$$A_0 = (1\ \ 1\ \ \dots\ \ 1).$$

Since $K(n,g)$ values enter the elements of the transfer matrix linearly, the probability $c(n,g)$ that the DNA site n is covered by a protein of type g can be expressed as follows (41):

$$c(n,g) = \frac{\partial Z}{\partial K(n,g)} \times \frac{K(n,g)}{Z}. \quad (18)$$

Fig. S1 shows exemplary calculations for the binding of a ligand of length $m = 3$ to a DNA lattice of length $N = 4$ with binding constant K and total concentration c_0 set to 1. This type of ligand could represent, for example, the flexible linear polyamine spermidine^{3+}, which can contact the DNA double helix either by one, two, or three amino-groups (45–47). In the all-or-none binding model only three states exist, one free and two bound states (Fig. S1 *A*). Therefore, the probability that the first basepair is covered by the protein is 1/3, while the probability that the second basepair is covered by the protein is 2/3. When incomplete binding is taken into account, 25 new possible states have to be considered (Fig. S1 *B*). The inspection of all 28 possible states shows that there are 14 states where the first basepair is covered and 16 states where the second basepair is covered. Therefore the probability that the first basepair is occupied by the ligand is 14/28 = 1/2, while the probability that the second basepair is bound is 16/28 = 4/7. Calculations according to Eqs. 1–18 provide exactly the same values (Fig. S1 *C*), demonstrating that the algorithm described above is correct.

RESULTS AND DISCUSSION

In the nucleosome model, the histone octamer core particle interacts with a continuous stretch of $m = 147$ DNA basepairs. This value is used for all calculations presented here. Although physically the DNA wraps around the protein core, in terms of the model it is equivalent to assuming a flexible protein entity binding the DNA and covering 147 basepairs upon binding (Fig. 1). Nucleosome unwrapping is represented by one or more unbound DNA units at the ends of the protein's DNA interaction site. In the lattice formalism, all distances are measured in elementary DNA units. Throughout this study, the elementary unit is one basepair. However, the unit size may be set larger than one basepair for more coarse-grained calculations and to speed up calculations. For example, assuming that an elementary unit comprises 10 basepairs would correspond to a nucleosome length of $m = 15$ units, and accelerate calculations by an order of magnitude. Because Eqs. 1–18 are applicable to any number of protein types f, the nucleosomes can be represented by one protein of type $g = 1$, while other proteins, such as linker histones, nonhistone architectural proteins, and transcription factors, can be represented by additional proteins of type $g = (2, f)$.

NUCLEOSOME UNWRAPPING REDISTRIBUTES THE DNA OCCUPANCY IN FAVOR OF THE INNER DNA TURN

For a short DNA fragment of $N = 147$ bp in the presence of nucleosomes at a midsaturating concentration ($K \times c_0 = 1$), the histone core particle has only one potential binding site in the all-or-none model. Therefore, the probability that any DNA unit is bound by the histone is equal to 0.5 (Fig. 2 *A*). The nucleosome-unwrapping model predicts a different behavior. The binding probability is maximal in the middle of the nucleosome and smoothly decreases toward the edges of the DNA segment (Fig. 2 *A*). Thus,

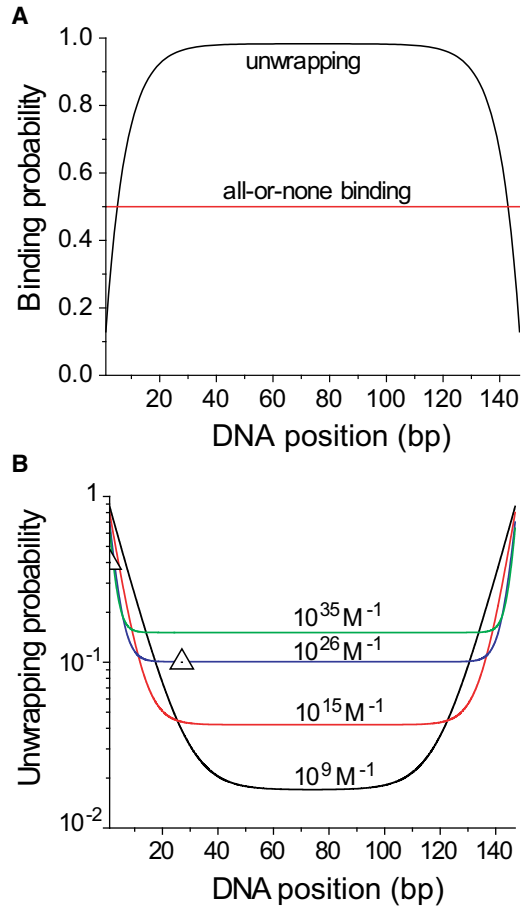

FIGURE 2 DNA interactions with the nucleosome core calculated with the theoretical framework presented here for $N = 147$ bp of DNA. The product of equilibrium binding constant and free protein concentration was kept constant at $K \times c_0 = 1$. (*A*) Binding probability calculated for the all-or-none model (intermediate states are prohibited) and for the nucleosome-unwrapping model, in which intermediate states are allowed for $K = 10^9$ M^{-1}. (*B*) Unwrapping probability calculated for different K values are indicated in the figure. (*Triangles*) Experimental values of unwrapping probabilities from Koopmans et al. (33). Note that K represents the total protein-DNA binding energy in the nucleosome (~1–2 k_BT per 1 nm of DNA equivalent to 0.6–1.2 kcal mol^{-1} nm^{-1} at room temperature). Thus, the partial dissociation of DNA at the ends would occur spontaneously.

several basepairs at the DNA entry/exit site of the nucleosome can unwrap more easily. This is observed experimentally, for example, in fluorescent spectroscopy (28–33) and atomic force microscopy studies (27,48). A somewhat surprising result is that our model predicts also that unwrapped DNA segments at the ends of the nucleosome effectively help to stabilize the inner turn of the DNA in the nucleosome because they increase the occupancy of this region (Fig. 2 *A*). The latter effect is of entropic origin, as can be understood from a more simple case depicted in Fig. S1. Because nucleosome unwrapping decreases the effective length over which histone-DNA interactions occur, the number of possible bound configurations of the nucleosome increases. Additional binding configurations include translational sliding. However, even in the absence of translational sliding (e.g., if the nucleosome center is frozen at some position along the DNA lattice), the number of configurations increases due to differences in the degree of unwrapping. Both these effects increase the probability for the ~80 bp of the inner DNA turn of the nucleosome being associated with DNA.

The magnitude of the effects depicted in Fig. 2 *A* depends on the nucleosome-binding constant K and the effective free histone octamer concentration c_0. Because the latter is affected in the cell by many processes such as the activity of histone chaperones and chromatin remodeling complexes, only rough estimates and no exact values of these parameters are currently available (21). Fig. 2 *B* shows calculations performed for different values of K and c_0. Following a recent experimental study, nucleosome unwrapping is quantitatively characterized by the unwrapping probability (or the equilibrium constant) for wrapping/unwrapping of a given DNA segment as a function of the distance to the DNA entry/exit site at the nucleosome (30). In our calculations, several cases ranging from half-saturation ($K \times c_0 = 1$) to a complete saturation ($K \times c_0 = 1000$) were considered. The experimental data obtained in the study of Koopmans et al. (33) can be fitted with the value of $K = 10^{26}\ \mathrm{M}^{-1}$ for the equilibrium binding constant in the half-saturation regime. This would correspond to an interaction energy of ~1–2 k_BT per nm of DNA, which is consistent with other estimates (39). The nucleosome unwrapping probability of ~0.1 at a location 27 basepairs inside the nucleosome reported by Koopmans et al. (30) is in agreement with the value 0.02–0.1 reported by Li and Widom (49) for the physiological ionic conditions. However, it should be noted that the two experimental points in Fig. 2 *B* are not sufficient for a reliable determination of the total unwrapping energy. Furthermore, in vivo the process of nucleosome disassembly is likely to proceed via dissociation of one and then two histone H2A·H2B dimers (50). The binding/dissociation of one H2A·H2B dimer is characterized by a dissociation constant in the nanomolar range and could be considered as the initial nucleosome disassembly step.

NUCLEOSOME UNWRAPPING CHANGES GENOMIC NUCLEOSOME MAPS

The nucleosome-unwrapping model can significantly change nucleosome occupancy maps as compared to those calculated for the all-or-none binding model (Fig. 3). In Fig. 3, *A* and *B*, calculations for DNA lattices of 200 and 1000 bp are shown, which represent the characteristic length scale of promoter regions for which the nucleosome position has a crucial effect on transcription initiation (51,52). Furthermore, we conducted a comparison with a recent atomic force microscopy (AFM) study where the exact nucleosome positions on a yeast DNA sequence were measured via high-resolution imaging and averaged to derive the nucleosome probability distribution (48). Fig. 3 *C* shows the resulting experimental data for a DNA region of 394 bp from chromosome 3 in comparison with the unwrapping and the all-or-none model. The all-or-none model predicts that a given 394-bp DNA fragment will most probably have two nucleosomes close to the DNA ends, and a nucleosome-depleted region in the middle. While the two-wing structure of this curve describes the experimental distribution quite well, it does not explain the experimentally observed increase of nucleosome-probability in the middle of the DNA. However, if nucleosome unwrapping is taken into account, the calculated nucleosome probability displays an increase in the middle of the DNA segment, in agreement with the experimental results. Neither the all-or-none model curve, nor the nucleosome-unwrapping curve completely coincides with the experimental curve, most likely because DNA sequence-specificity of binding was not considered. The latter predicts an energy barrier for nucleosome formation in the middle of this DNA segment (48), which would favor depletion, not enrichment, of the nucleosome probability in the middle of the DNA. The experimentally observed enrichment of the nucleosome probability in Fig. 3 *C* can be explained by the nucleosome-unwrapping model, but not by the all-or-none binding model.

NUCLEOSOME UNWRAPPING FACILITATES TRANSCRIPTION FACTOR DNA-BINDING IN THE PRESENCE OF A NUCLEOSOME

Interaction of transcription factors with nucleosomes is an integral part of any gene-regulatory process.

One simple and insightful concept for the linkage of TF and histone octamer binding is the so-called "collaborative competition" model (53). The nonspecific binding of the histone octamer to the DNA competes with a sequence-specific binding of transcription factors. For the case of two TF binding sites near to each other, the binding of the first TF may stabilize the partial unwrapping of nucleosomal DNA so that the binding of the second TF nearby is facilitated. Because multiple regulatory DNA sites often occur within a nucleosome-length distance, cooperativity effects of this type are to be expected in vivo, and have been

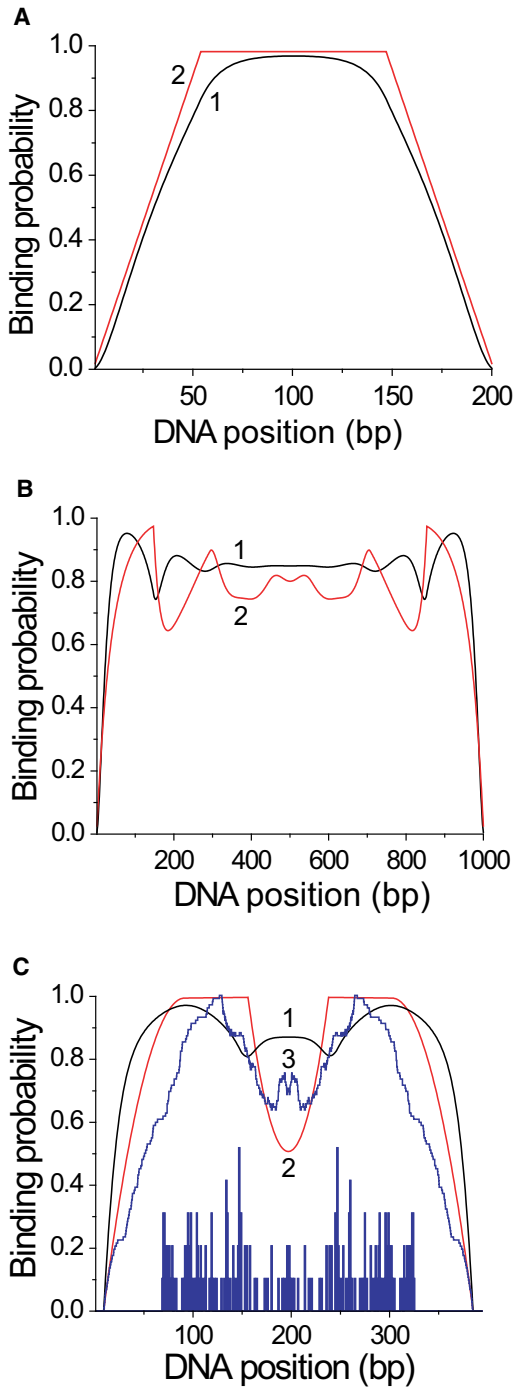

FIGURE 3 Influence of partial DNA unwrapping on positioning multiple nucleosomes. The nucleosome-unwrapping model (*line 1*) was compared to all-or-none binding (*line 2*) for $K = 10^9$ M^{-1} with a value of $c_0 = 10^{-9}$ M in panels *A* and *B*. (*A*) DNA length $N = 200$ bp. (*B*) DNA length $N = 1000$ bp. (*C*) The experimentally observed nucleosome distribution from Fig. 2*A* of Milani et al. (48) (*line 3*) for a DNA region of $N = 394$ bp (*S. cerevisiae* genomic DNA from chromosome 3 pos. 249050-249443) was compared with the binding probability calculated for this DNA segment assuming all-or-none binding (*line 2*) or allowing partial nucleosome unwrapping (*line 1*). The bars shown in the bottom correspond to the raw data from the AFM measurements of the centers of nucleosome positions that were rebinned here by summing up all counts within each 1-bp window.

observed experimentally in vitro (53,54). To examine the collaborative competition effect in the context of the all-or-none versus the nucleosome-unwrapping model, the calculations depicted in Fig. 4 were conducted. For a mononucleosome with one specific TF-binding site at $n = 10$ and a binding site length of 10 bp, the competitive binding under saturating conditions of histone octamer binding was evaluated ($K_{nuc} = 10^9$ M^{-1}, $c_{0,nuc} = 10^{-6}$ M; $K_{TF} = 10^{10}$ M^{-1}, $c_{0,TF} = 10^{-9}$ M) (Fig. 4 *A*). In the absence of TF, the nucleosome is formed in this region with a probability close to 1. For the all-or-none binding model, TF binding is strongly suppressed by the nucleosome under these conditions (the probability of TF binding is <0.5%). In contrast, a TF occupancy of ~60% is computed for this binding site with the nucleosome-unwrapping model.

To evaluate the cooperativity of TF binding, a second TF binding site was added to the same DNA segment, separated by 5 bp from the first TF binding site (Fig. 4 *B*). In the frame of the all-or-none binding model, both TFs bind with a probability of ~40%. This is much higher than for a single site in Fig. 4 *A* and results from the fact, that binding the second TF becomes much easier as soon as the first TF has bound the DNA. Because the nucleosome binds as a single entity, both sites have equal probabilities for TF binding. The nucleosome-unwrapping model yields a significantly different behavior. The TF occupancy is 80% for the left binding site, while the second binding site is occupied to ~60% (which is still higher than 40% occupancy in the all-or-none model). Thus, both the all-or-none and the unwrapping model predict the collaborative competition effect, but the unwrapping model leads to a much higher cooperativity of TF binding.

In a third scenario, competitive binding of TF and histone octamer was studied on a longer DNA fragment of 500 bp (Fig. 4 *C*). As in the previous calculations, nucleosomes form nonspecifically whereas TFs have a specific recognition sequence. In the all-or-none binding model, the collaborative competition effect weakens with the length of the DNA segment. TF occupancy at the double binding sites decreases from 40% in the case of 147-bp DNA (Fig. 4 *B*) to 10% in the case of 500 bp DNA (Fig. 4 *C*). In the nucleosome-unwrapping model, the two TF binding sites are characterized by TF occupancy at ~70% and hardly decrease with DNA length. Thus, in the nucleosome-unwrapping model, TFs always outcompete nucleosomes at these sites for the given parameter set. This does not mean that in all cases TF binding probability is higher than that of a nucleosome at a given site. It rather indicates that the nucleosome probability at a given site has a noticeable reduction with respect to its average value due to TF binding. In high-throughput experiments, such nucleosome-depleted regions are usually interpreted as sequences with low affinity to bind a nucleosome.

A value of $c_0 = 8 \times 10^{-9}$ M was used in the calculations to improve the fit to the experimental dataset.

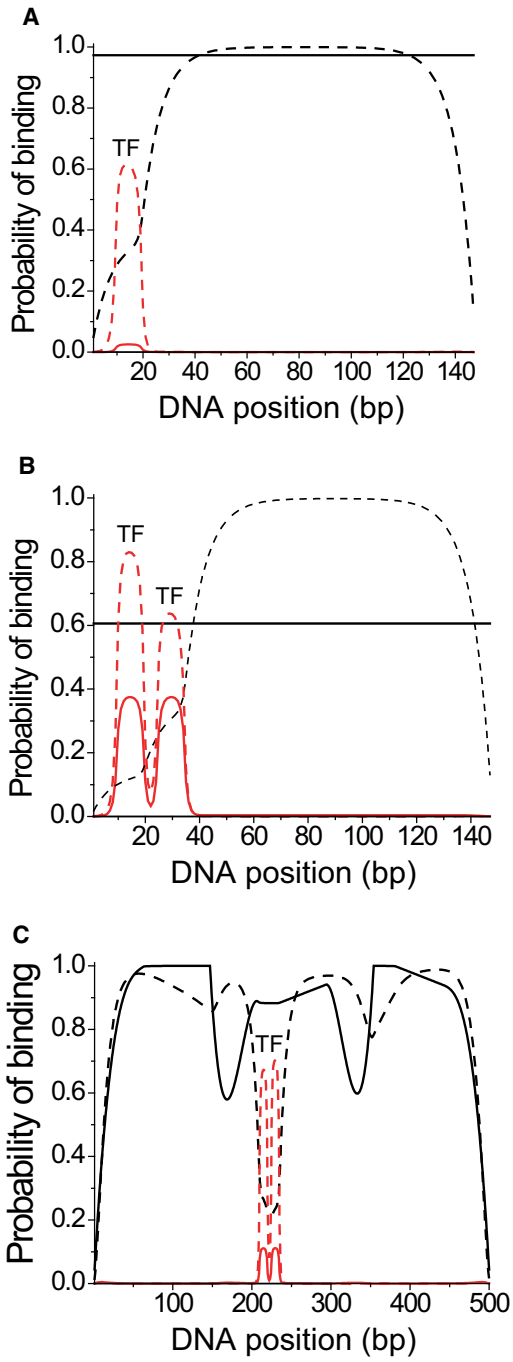

FIGURE 4 Competitive binding of histone octamer and transcription factors to DNA. (*Solid lines*) Calculated for the all-or-none binding model. (*Dashed lines*) Nucleosome-unwrapping model. Nucleosome occupancies in the absence of TF are 1.0 for all positions. TF binding site length $m_{TF} = 10$ bp, $K(n, \text{nuc}) = 10^9\ \text{M}^{-1}$ for any n, $c_0(\text{nuc}) = 10^{-6}$ M, and $c_0(\text{TF}) = 10^{-9}$ M. (*A*) $N = 147$ bp, $K(n = 10, \text{TF}) = 10^{10}\ \text{M}^{-1}$, and $K(n \neq 10, \text{TF}) = 10^6\ \text{M}^{-1}$. (*B*) $N = 147$ bp, $K(n = [10, 15], \text{TF}) = 10^{10}\ \text{M}^{-1}$, and $K(n \neq [10, 25], \text{TF}) = 10^6\ \text{M}^{-1}$. (*C*) $N = 500$ bp, $K(n = [210, 225], \text{TF}) = 10^{10}\ \text{M}^{-1}$, and $K(n \neq [210, 225], \text{TF}) = 10^6\ \text{M}^{-1}$. The index *nuc* corresponds to the nucleosome core particle while *TF* designates the transcription factor.

However, our analysis suggests that they might as well represent regions of TF binding sites.

Experimentally, it is found that promoters of transcriptionally active genes are usually depleted from nucleosomes, while promoters of transcriptionally inactive genes are occluded by nucleosomes and remodeled in the process of gene activation (51,52). Many promoters possess both the nucleosome-excluding signals (such as poly(dA·dT) repeats) and the TF- or RNAP-positioning signals such as TATA boxes (24,55,56). Thus, the egg-or-chicken question remains unsolved: It is not clear whether the region upstream of the transcription start site is nucleosome-depleted at active promoters due to the inherently low affinity of the histone octamer to the underlying DNA sequence (57), or due to the displacement by transcription factors (58) combined with the action of chromatin remodelers (13). The calculations performed in the context of Fig. 4 support the view that TF sequence preferences are of the first order of importance, while nucleosome sequence-preferences are of second order.

INCLUDING A SINGLE-BASE RESOLUTION UNWRAPPING POTENTIAL FROM MOLECULAR DYNAMICS SIMULATIONS IN THE LATTICE MODEL

The calculations above were performed assuming a homogenous histone-DNA interaction potential with the energy equally distributed among all contacts inside the nucleosome:

$$K(n, g, h) = \sqrt[m_g]{K(n, g)}.$$

However, this is not the case in the real nucleosome, where at least 14 distinct interaction centers have been identified previously in the crystal structure (59). In order to reproduce the histone-DNA interaction potential inside the nucleosome, we performed a series of molecular dynamics (MD) calculations (Fig. 5 *A*). Nucleosomal MD trajectories were analyzed in terms of the interaction enthalpies between the DNA and the histone proteins. This resulted in the single-base resolution interaction energy profile presented in Fig. 5 *B*. Based on the relative energy differences derived from the MD simulations, microscopic binding constants $K(n, g, h)$ were derived as follows:

$$\begin{cases} K(n, g, h) = C \times \exp(-\beta \Delta H) \\ \sum_{h=1}^{147} K(n, g, h) = K(n, g) \end{cases}. \qquad (19)$$

Here $K(n, g)$ is the macroscopic binding constant (the same one as used in the previous calculations with homogenous potential), and C is the normalization constant which disappears after solving the expressions in Eq. 19. In Fig. 5 *C*, calculations performed for the histone-DNA interaction potential determined by the expressions in Eq. 19 are

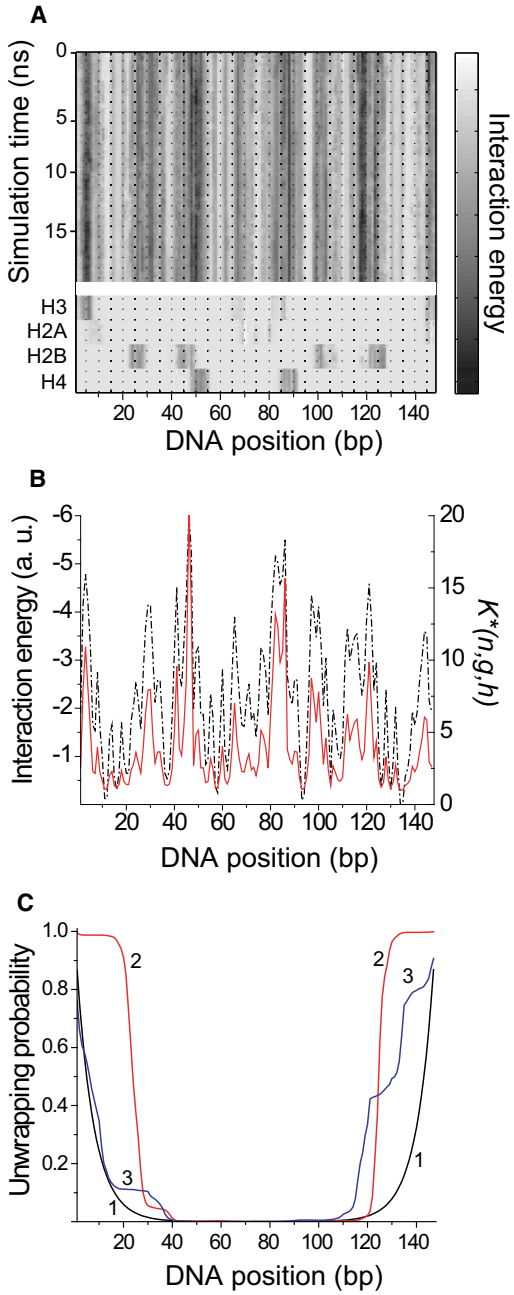

FIGURE 5 Effect of local variations in histone octamer-DNA interaction energies on the unwrapping behavior. Calculations in panels *B* and *C* were conducted with $K(n,g) = 10^9$ M^{-1}, $c_0 = 10^{-6}$ M. (*A*) Histone-DNA interaction map determined from all-atom molecular dynamics simulations. The bottom panel shows the contribution of the unstructured N-terminal histone tails to the nucleosomal DNA interactions. Residues in these tails can be modified by posttranslational modifications, which could modulate the DNA interaction affinity. (*B*) The relative interaction energy derived from the MD simulations of the nucleosome (*dashed lines*) and the corresponding microscopic binding constants $K(n,g,h)$ (*solid lines*). (*C*) Comparison of the nucleosome unwrapping profile calculated for different DNA-histone interaction potentials: 1), Homogeneous interaction energies $K(n,g,h) = \sqrt[ms]{K(n,g)}$ at all DNA positions. 2), Microscopic binding constants $K(n,g,h)$ estimated from MD simulations in panel *A*. 3), Random microscopic binding constants $K(n,g,h)$ scaled so that their total product gives $K(n,g)$.

presented. As for the homogenous potential, the unwrapping profile is characterized by the destabilized outer DNA turn and stabilized inner DNA turn. However, for the more realistic interaction potential, the boundary between the inner and outer turns becomes even more abrupt: ~60 bp of the outer nucleosome turn can easily unwrap, while the rest of the DNA has a near-zero probability of unwrapping (it could be freed only when the nucleosome dissociates as a whole or some of the nucleosome subunits dissociate). For comparison, we calculated nucleosome unwrapping assuming that the microscopic binding constants $K(n, g, h)$ take random values (provided their product gives the same macroscopic binding constant $K(n, g)$). These did not reproduce the abrupt transition of nucleosome unwrapping calculated for the histone-DNA interaction potential determined from the MD simulations (Fig. 5 *C*). Thus, while nucleosome unwrapping is dominated by the general energy-entropy competition described above, specific interactions inside the nucleosome shape the dissociation behavior of the DNA to form a rather sharp transition between the outer and inner turn of the DNA wrapped around the histone octamer.

CONCLUSIONS

Here we have introduced what we believe to be the first lattice model for protein-DNA binding that takes into account intermediate protein binding states. This model provides a significantly more accurate description for the interaction of the histone octamer with DNA in the nucleosome complex than the all-or-none binding model currently used in the analysis of DNA-protein binding in vitro and in vivo (9,11–18,41,57,58,60,61). Within the limited available experimental data sets, the shape of the predicted unwrapping curves is consistent with FRET measurements of the unbinding probability dependence on the length of unwrapped DNA (30). The nucleosome binding constant estimated from fitting of these experimental data gives an unwrapping energy of 1–2 k_BT per nm DNA, in agreement with other studies (36–39). Taking into account partial unwrapping of nucleosomal DNA leads to a redistribution of the DNA occupancies. The outer turn of the nucleosomal DNA becomes more accessible and the inner turn less accessible for the binding of transcription factors due to entropic reasons (Fig. 2). Thus, access of transcription factors to the DNA close to the site where it enters/exits the nucleosome is facilitated over binding of the regions of the nucleosomal DNA.

Taking nucleosome unwrapping into account provides a more faithful transcription of competitive DNA binding of TFs and the histone octamer core as well as an explanation for the nucleosome-mediated cooperativity of TF binding proposed previously (Fig. 3). It also significantly affects positioning of multiple nucleosomes and thus is relevant for analyzing their distribution at a given genomic

region. Significant changes of nucleosome binding maps are expected by taking unwrapping of nucleosomal DNA into account, as inferred from the comparison of our calculations with the AFM experiments of Milani et al. (48). It was shown that the general conclusions obtained with the homogenous potential remain valid for a more realistic histone-DNA potential derived from MD simulations. In addition, a more abrupt boundary between the weakly bound outer DNA turn and the strongly bound inner DNA turn of the nucleosome was apparent. For the calculation of nucleosome and TF positioning, one should note that the microscopic binding constants $K(n,g,h)$ assigned to each DNA unit depend on the combination of three parameters: the position of the protein-binding site along the DNA, n; the type of nucleosome, g; and the position of a given unit of nucleosomal DNA with respect to the histone octamer, h. The dependence on the DNA sequence positions as given by parameter n has been characterized in detail by various bioinformatics approaches (4,13,23,24). It can be incorporated into theoretical descriptions as described elsewhere (13). The dependence on the internal nucleosome structure, i.e., the dependence on parameter h, was derived here from molecular dynamics simulation (Fig. 5).

It is to be expected that an additional layer of variations in the local histone-DNA interaction strength would be included by posttranslational modifications of histone tails (Fig. 5 *A*). In particular for acetylation of histone lysines, a direct effect on the stability of the nucleosome core particle as well as its higher-order interactions has been reported because the positively charged lysine is neutralized in the acetylated state (62,63). Likewise, also for DNA methylation a direct effect on the nucleosome stability has been inferred from spectroscopic in vitro studies (64). Given appropriate data sets, these effects could be integrated also into the parameter g to improve quantitative models of epigenetic regulation (21).

SUPPORTING MATERIAL

One figure, a test of the proposed calculation algorithm against an analytical solution for a short ligand, is available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(10)00988-4.

We thank Pascale Milani for providing the raw data of the published AFM measurements of nucleosome positioning and Easwar Ramakrishnan for the help with the data analysis in the Origin 8.0 software.

This work was supported by a fellowship of the German CellNetworks Cluster of Excellence (No. EXC81) and grant No. B10M-060 of the Belarusian Foundation of Fundamental Research to V.B.T. and by Deutsche Forschungsgemeinschaft grant No. Ri 1283/8-1 to KR.

REFERENCES

1. van Holde, K. E. 1989. Chromatin. Springer-Verlag, New York.
2. Markov, A. A. 1907. Investigation of a specific case of dependent observations. *Izv. Imper. Akad. Nauk (St.-Petersburg).* 3:61–80.
3. Ising, E. 1925. Contribution to the theory of ferromagnetism [Beitrag zur Theorie des Ferromagnetismus]. *Z. Phys.* 31:253–258.
4. Segal, E., and J. Widom. 2009. From DNA sequence to transcriptional behavior: a quantitative approach. *Nat. Rev. Genet.* 10:443–456.
5. Trifonov, E. N. 2010. Nucleosome positioning by sequence, state of the art and apparent finale. *J. Biomol. Struct. Dyn.* 27:741–746.
6. Roider, H. G., A. Kanhere, …, M. Vingron. 2007. Predicting transcription factor affinities to DNA from a biophysical model. *Bioinformatics.* 23:134–141.
7. Kornberg, R. D., and L. Stryer. 1988. Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.* 16(14A):6677–6690.
8. Nechipurenko, Y. D., and G. V. Gursky. 1986. Cooperative effects on binding of proteins to DNA. *Biophys. Chem.* 24:195–209.
9. Nechipurenko, I. D., and M. V. Vol'kenshteĭn. 1986. [Analysis of the nucleosome arrangement on satellite DNA]. *Dokl. Akad. Nauk SSSR.* 286:216–220.
10. Epstein, I. R. 1978. Cooperative and non-cooperative binding of large ligands to a finite one-dimensional lattice. A model for ligand-oligonucleotide interactions. *Biophys. Chem.* 8:327–339.
11. Raveh-Sadka, T., M. Levo, and E. Segal. 2009. Incorporating nucleosomes into thermodynamic models of transcription regulation. *Genome Res.* 19:1480–1496.
12. Teif, V. B. 2010. Predicting gene-regulation functions: lessons from temperate bacteriophages. *Biophys. J.* 98:1247–1256.
13. Teif, V. B., and K. Rippe. 2009. Predicting nucleosome positions on the DNA: combining intrinsic sequence preferences and remodeler activities. *Nucleic Acids Res.* 37:5641–5655.
14. Pérez, A. G., V. E. Angarica, …, A. T. Vasconcelos. 2009. From sequence to dynamics: the effects of transcription factor and polymerase concentration changes on activated and repressed promoters. *BMC Mol. Biol.* 10:92.
15. Wasson, T., and A. J. Hartemink. 2009. An ensemble model of competitive multi-factor binding of the genome. *Genome Res.* 19:2101–2112.
16. Granek, J. A., and N. D. Clarke. 2005. Explicit equilibrium modeling of transcription-factor binding and gene regulation. *Genome Biol.* 6:R87.
17. McGhee, J. D., and P. H. von Hippel. 1974. Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J. Mol. Biol.* 86:469–489.
18. Gurskii, G. V., A. S. Zasedatelev, and M. V. Vol'kenshtein. 1972. Theory of single-dimensional absorption. II. Adsorption of small molecules on a heteropolymer. *Mol. Biol. (Moscow).* 6:385–393.
19. Cera, E. D., and P. E. Phillipson. 1996. Map analysis of ligand binding to a linear lattice. *Biophys. Chem.* 61:125–129.
20. Djordjevic, M., A. M. Sengupta, and B. I. Shraiman. 2003. A biophysical approach to transcription factor binding site discovery. *Genome Res.* 13:2381–2390.
21. Teif, V., and K. Rippe. 2010. Statistical-mechanical lattice models for protein-DNA binding in chromatin. *J. Phys.: Condens. Matter.* 22:414105.
22. Stefanovsky, V. Y., G. Pelletier, …, T. Moss. 2001. DNA looping in the RNA polymerase I enhancesome is the result of non-cooperative in-phase bending by two UBF molecules. *Nucleic Acids Res.* 29:3241–3247.
23. Schones, D. E., and K. Zhao. 2008. Genome-wide approaches to studying chromatin modifications. *Nat. Rev. Genet.* 9:179–191.
24. Radman-Livaja, M., and O. J. Rando. 2010. Nucleosome positioning: how is it established, and why does it matter? *Dev. Biol.* 339:258–266.
25. Zlatanova, J., T. C. Bishop, …, K. van Holde. 2009. The nucleosome family: dynamic and growing. *Structure.* 17:160–171.
26. Suzuki, Y., Y. Higuchi, …, K. Takeyasu. 2010. Molecular dynamics of DNA and nucleosomes in solution studied by fast-scanning atomic force microscopy. *Ultramicroscopy.* 110:682–688.

27. Shlyakhtenko, L. S., A. Y. Lushnikov, and Y. L. Lyubchenko. 2009. Dynamics of nucleosomes revealed by time-lapse atomic force microscopy. *Biochemistry.* 48:7842–7848.

28. Poirier, M. G., M. Bussiek, …, J. Widom. 2008. Spontaneous access to DNA target sites in folded chromatin fibers. *J. Mol. Biol.* 379:772–786.

29. Poirier, M. G., E. Oh, …, J. Widom. 2009. Dynamics and function of compact nucleosome arrays. *Nat. Struct. Mol. Biol.* 16:938–944.

30. Gansen, A., A. Valeri, …, C. A. Seidel. 2009. Nucleosome disassembly intermediates characterized by single-molecule FRET. *Proc. Natl. Acad. Sci. USA.* 106:15308–15313.

31. Bucceri, A., K. Kapitza, and F. Thoma. 2006. Rapid accessibility of nucleosomal DNA in yeast on a second time scale. *EMBO J.* 25:3123–3132.

32. Li, G., M. Levitus, …, J. Widom. 2005. Rapid spontaneous accessibility of nucleosomal DNA. *Nat. Struct. Mol. Biol.* 12:46–53.

33. Koopmans, W. J., R. Buning, …, J. van Noort. 2009. spFRET using alternating excitation and FCS reveals progressive DNA unwrapping in nucleosomes. *Biophys. J.* 97:195–204.

34. Anderson, J. D., A. Thåström, and J. Widom. 2002. Spontaneous access of proteins to buried nucleosomal DNA target sites occurs via a mechanism that is distinct from nucleosome translocation. *Mol. Cell. Biol.* 22:7147–7157.

35. Anderson, J. D., and J. Widom. 2000. Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *J. Mol. Biol.* 296:979–987.

36. Möbius, W., R. A. Neher, and U. Gerland. 2006. Kinetic accessibility of buried DNA sites in nucleosomes. *Phys. Rev. Lett.* 97:208102.

37. Qamhieh, K., T. Nylander, and M. L. Ainalem. 2009. Analytical model study of dendrimer/DNA complexes. *Biomacromolecules.* 10: 1720–1726.

38. Brower-Toland, B. D., C. L. Smith, …, M. D. Wang. 2002. Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA. *Proc. Natl. Acad. Sci. USA.* 99:1960–1965.

39. Kulić, I. M., and H. Schiessel. 2004. DNA spools under tension. *Phys. Rev. Lett.* 92:228101.

40. 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.

41. Teif, V. B. 2007. General transfer matrix formalism to calculate DNA-protein-drug binding in gene regulation: application to O_R operator of phage λ. *Nucleic Acids Res.* 35:e80.

42. Phillips, J. C., R. Braun, …, K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.

43. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:327–338.

44. Teif, V. B., S. G. Haroutiunian, …, D. Y. Lando. 2002. Short-range interactions and size of ligands bound to DNA strongly influence adsorptive phase transition caused by long-range interactions. *J. Biomol. Struct. Dyn.* 19:1093–1100.

45. Todd, B. A., and D. C. Rau. 2008. Interplay of ion binding and attraction in DNA condensed by multivalent cations. *Nucleic Acids Res.* 36:501–510.

46. Teif, V. B. 2005. Ligand-induced DNA condensation: choosing the model. *Biophys. J.* 89:2574–2587.

47. Teif, V. B., and K. Bohinc. 2010. Condensed DNA: condensing the concepts. *Prog. Biophys. Mol. Biol.*, in press. Available online July 16, 2010. DOI: 10.1016/j.pbiomolbio.2010.1007.1002.

48. Milani, P., G. Chevereau, …, A. Arneodo. 2009. Nucleosome positioning by genomic excluding-energy barriers. *Proc. Natl. Acad. Sci. USA.* 106:22257–22262.

49. Li, G., and J. Widom. 2004. Nucleosomes facilitate their own invasion. *Nat. Struct. Mol. Biol.* 11:763–769.

50. Rippe, K., J. Mazurkiewicz, and N. Kepper. 2008. Interactions of histones with DNA: nucleosome assembly, stability and dynamics. *In* DNA Interactions with Polymers and Surfactants. R. S. Dias, and B. Lindman, editors. Wiley, London. 135–172.

51. Schones, D. E., K. Cui, …, K. Zhao. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell.* 132:887–898.

52. Kaplan, N., I. K. Moore, …, E. Segal. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature.* 458:362–366.

53. Polach, K. J., and J. Widom. 1996. A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *J. Mol. Biol.* 258:800–812.

54. Adams, C. C., and J. L. Workman. 1995. Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol. Cell. Biol.* 15:1405–1421.

55. Zhang, Y., Z. Moqtaderi, …, K. Struhl. 2009. Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat. Struct. Mol. Biol.* 16:847–852.

56. Segal, E., and J. Widom. 2009. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* 19:65–71.

57. Segal, E., Y. Fondufe-Mittendorf, …, J. Widom. 2006. A genomic code for nucleosome positioning. *Nature.* 442:772–778.

58. Morozov, A. V., K. Fortney, …, E. D. Siggia. 2009. Using DNA mechanics to predict in vitro nucleosome positions and formation energies. *Nucleic Acids Res.* 37:4707–4722.

59. Davey, C. A., D. F. Sargent, …, T. J. Richmond. 2002. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* 319:1097–1113.

60. Crothers, D. M. 1968. Calculation of binding isotherms for heterogenous polymers. *Biopolymers.* 6:575–584.

61. Akhrem, A. A., A. S. Fridman, and D. Y. Lando. 1985. Effect of ligand interaction with the boundaries between different forms of DNA on intramolecular transitions. [Trans.]. *Dokl. Akad. Nauk SSSR.* 284: 212–216.

62. Shogren-Knaak, M., H. Ishii, …, C. L. Peterson. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science.* 311:844–847.

63. Wang, X., and J. J. Hayes. 2006. Physical methods used to study core histone tail structures and interactions in solution. *Biochem. Cell Biol.* 84:578–588.

64. Choy, J. S., S. Wei, …, T. H. Lee. 2010. DNA methylation increases nucleosome compaction and rigidity. *J. Am. Chem. Soc.* 132:1782–1783.