Nucleosome Switches

David J. Schwab,* Robijn F. Bruinsma, and Joseph Rudnick
Department of Physics and Astronomy, University of California, Los Angeles, Los Angeles, California 90024, USA

Jonathan Widom
Department of Biochemistry, Molecular Biology, and Cell Biology, Northwestern University, Evanston. Illinois 60208, USA
(Received 7 August 2007; published 6 June 2008)

We present a statistical-mechanical model for the positioning of nucleosomes along genomic DNA molecules as a function of the strength of the binding potential and the chemical potential of the nucleosomes. We show that a significant section of the DNA is composed of two-level nucleosome switching regions where the nucleosome distribution undergoes a localized, first-order transition. The location of the nucleosome switches shows a strong correlation with the location of gene-regulation regions.

DOI: 10.1103/PhysRevLett.100.228105 PACS numbers: 87.15.A−, 05.20.Jj, 87.16.Sr

Genetic information in higher organisms is encoded in centimeters-long DNA molecules, which are compacted into chromosomes in the cell nucleus through a hierarchical series of folding steps. In the first level of compaction, short stretches of (negatively charged) DNA [147 base pairs (bp), 50 nm] are wrapped locally in ~2 turns around (positively charged) 3.5 nm radius protein spools, forming nucleosomes (see Fig. 1, top inset). Nucleosomes are separated along the DNA by short stretches of unwrapped "linker" DNA, typically about 10–50 bp in length. Thus, about 75%–90% of each chromosomal DNA molecule is wrapped in nucleosomes. DNA that is wrapped in nucleosomes is less-easily accessible to many gene-regulatory proteins; therefore, the detailed locations of nucleosomes along the DNA have important biological consequences.

In an early model, Kornberg and Stryer [1] treated the nucleosomes as a uniform one-dimensional (1D) liquid of hard rods with an excluded volume of the order of 147 bp. They attributed regularities in nucleosome positioning observed *in vivo* to the decaying density oscillations near a boundary that are characteristic for the 1D liquid of hard rods at higher densities [2]. The "boundaries" would be provided here, for example, by the site-specific binding of transcription factors (TFs) that, when bound to DNA, prevent that stretch of DNA from being wrapped in a nucleosome. Recently, Segal *et al.* showed that nucleosomes in fact prefer certain DNA sequences over others, and that genomes utilize these sequence preferences to bias the preferred locations of nucleosomes. This sequence specificity of DNA/nucleosome binding is due to the sequence-dependent bending stiffness of DNA itself. Certain dinucleotides (one base pair followed by another) greatly enhance the ability of DNA to sharply bend in particular directions relative to the DNA helix axis, as required by the nucleosome. This sharp bending occurs every DNA helical repeat (~10 base pairs), when the major groove of the DNA faces inwards towards the center of curvature, and again, about 5 base pairs out of phase, and with opposite direction, when the major groove faces outward. Bends of each direction are facilitated by specific dinucleotides. Segal *et al.* [3] used an alignment of natural nucleosome DNA sequences to create a statistical profile (a position-dependent Markov model) that represented the nucleosome's dinucleotide sequence preferences at each of the base pair steps within the nucleosome. The resulting Markov model assigns a log-likelihood for an isolated nucleosome starting at a given base pair in the genome, which amounts to an effective on site potential for nucleosomes. They then used a dynamic programming algorithm to exactly solve the equilibrium distribution of hard rods on this potential landscape, and the resulting equilib-

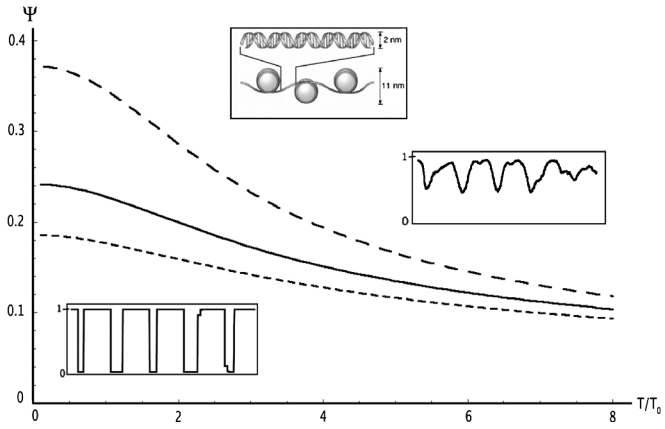

FIG. 1. Temperature dependence of the overlap parameter Ψ with temperature in units of the reference temperature. Short dashed line: $\mu = 62$ with zero temperature occupancy of 89.6%. Solid line: $\mu = 24$ with zero temperature occupancy of 83.3%. Long dashed line: $\mu = 0$ with zero temperature occupancy of 51.5%. Insets: Plot of site occupancy probability between sites 9000 and 10 000 at $\mu = 24$ for $T/T_0 = 0.1$ (bottom left) and $T/T_0 = 8.0$ (top right).

rium density profile of this 1D statistical mechanics model correctly predicted about half of the actual stable nucleosome positions of the yeast genome.

In many organisms, cells of different types or developmental states—all of which have the same genomic DNA sequences—organize their DNA into arrays of nucleosomes at quite different nucleosome densities (concentrations). For example, rat neuronal cells have an average nucleosome spacing of ~165 bp (~18 bp average linker DNA length) at birth, while 30 days later the same cells have increased the spacing to ~218 bp (~71 bp average linker DNA) [4]. The chromatin repeat length of brain cortex and cerebellar neurons changes concomitant with terminal differentiation. If slight changes in average density or nucleosome binding strength would alter the nucleosome density profile in a drastic and chaotic manner, then the predicted nucleosome positions would not be "robust" and any functionality for nucleosome positioning would be in serious doubt. The aim of this Letter is to examine the thermodynamic stability of nucleosome positioning using the model of Segal *et al.* [3].

We work in the grand-canonical ensemble with the N-nucleosome Hamiltonian:

$$H = \sum_{i=1}^{N} U(n_{i+1} - n_i) + \sum_{i=1}^{N} V_{n_i}. \tag{1}$$

Here, n_i is the location of the first base pair blocked by the ith nucleosome. The nucleosome-nucleosome interaction potential $U(m)$ is equal to zero if m exceeds the hard-core size $a = 157$ [5] and infinite if m is less than or equal to a. Next, $V_n \equiv -k_B T_0 \log P_n$ is the nucleosome on site potential energy with T_0 the reference temperature [6]—the characteristic energy scale of the on site potential—and P_n the likelihood for nucleosome binding at site n as obtained from the Markov model analysis for the DNA sequence of yeast chromosme II [7]. The probability ρ_n for site n to be the first site of a stretch of a sites occupied by a nucleosome is obtained from a discrete version of an exact recursion relation derived by Percus describing the statistical mechanics of a liquid of hard rods in an external potential [8]. This recursion relation,

$$h_n = \frac{H_n e^{\beta(\mu - V_n)}}{1 + H_n e^{\beta(\mu - V_n)}}, \tag{2}$$

with $H_n = \prod_{m=2}^{a}(1 - h_{n+m-1})$, is expressed in terms of the set of quantities h_n that we obtain by numerical iteration of Eq. (2) from large to small n. The site probability ρ_n then follows from the relation $h_i = \rho_i(1 - \sum_{j=1}^{a-1} \rho_{i-j})^{-1}$ while the thermodynamic properties follow from the grand partition function $\Xi(\mu, T) \propto \prod_n (\frac{1}{1-h_n})$. Note that Eq. (2) reduces to the classical Langmuir isotherm $\rho_n = \frac{e^{\beta(\mu - V_n)}}{1 + e^{\beta(\mu - V_n)}}$ in the absence of the excluded-volume interactions.

The level of correlation between the nucleosome site probabilities and the position dependence of the site potential V_n will be characterized by an overlap parameter Ψ:

$$\Psi = N^{-1} \frac{\sum_{n=1}^{L} \rho_n (V_n - \langle V \rangle)}{V_{\min} - \langle V \rangle}. \tag{3}$$

Here, $\langle V \rangle$ is the mean on site probability, L is the system size, and $V_{\min}$ is the site potential with the largest (i.e., most attractive) binding energy of the sequence. If all nucleosomes are located at optimal binding sites, where V_n equals $V_{\min}$, then Ψ equals one while Ψ is zero if there are no correlations between site probability and site potential, which occurs as $T/T_0 \to \infty$. Figure 1 shows the dependence of this overlap parameter on the ratio T/T_0 of the ambient and reference temperatures for different values of the chemical potential corresponding to densities in the biologically relevant range of 50%–90% occupancy. In all cases, the overlap parameter decreases towards zero for T/T_0 large compared to one. In this regime, the site occupation probability shows—approximately—sinusoidal density modulations with period equal to the hard rod length (right-hand inset) corresponding to the Kornberg-Stryer density oscillations [1]. The overlap parameter grows as T/T_0 decreases and then saturates for T/T_0 small compared to one. Note, from Fig. 2, that the value of Ψ for small T/T_0 increases with decreasing density. Indeed, in the close-packing limit (i.e., 100% occupancy) the site occupancy probability is constant, which—according to Eq. (2)—means that the overlap parameter must be zero. Nevertheless, Ψ remains surprisingly large, even at 90% occupancy. A typical occupancy plot in this regime, shown in the left-hand inset of Fig. 2, clearly indicates preferred locations for the nucleosomes. It is indeed only in this low-temperature regime that precise nucleosome locations can

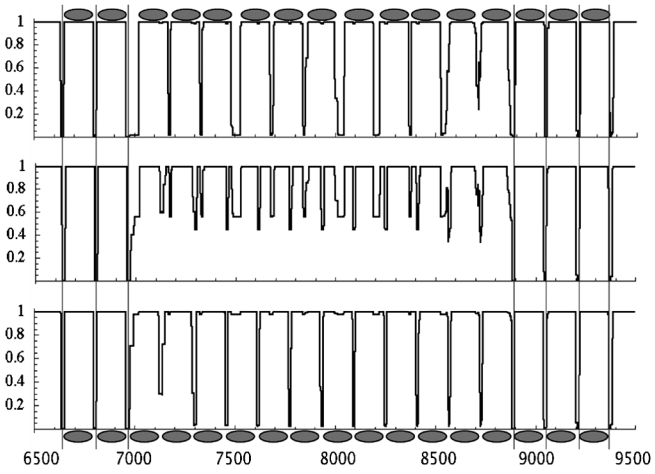

FIG. 2. Site occupancy of an unstable section (bp 6500–9500) at $T/T_0 = 0.5$ for different values of the chemical potential. The mean occupancies are 88.8%, 89.3%, and 90%. For higher, respectively, lower, chemical potentials, the occupation pattern is stable with, respectively, 12 and 11 nucleosomes. The unstable sequence for intermediate chemical potential is the superposition of the two stable patterns.

be predicted. The change between the two regimes, for T/T_0 near one, can be viewed as a freezing transition, but Fig. 1 does not provide evidence for a true thermodynamic phase transition [9].

Turning now to the thermodynamic stability of nucleosome positioning, we find that the occupancy profile for $T/T_0 = 0.5$ contains both stable and unstable sequences. Most sequences exhibit well-defined density profiles of the form shown in Fig. 1, which allow an unambiguous assignment of nucleosome positions and which are stable with respect to small changes in μ. These stable sequences are, however, interspersed by shorter sequences, having lengths of the order of 1000 bp, with poorly defined density profiles. The fraction of unstable sequences is about 15% at $\mu = 24$ for $T/T_0 = 0.5$ while it decreases to zero in the limit of small T/T_0. The middle panel of Fig. 2 shows a typical example of such an unstable sequence with a length of 3000 bp. If one slightly increases the chemical potential (from 89.3% to 90% mean occupancy) then the occupancy profile of this section evolves into a well-defined nucleosome configuration (with $N = 12$ nucleosomes; Fig. 2, top panel). Similarly, after a slight decrease in mean occupancy, to 88.8%, the section again transforms to a well-defined configuration (with $N = 11$ nucleosomes; Fig. 2, bottom panel).

The physical meaning of the ''disordered'' occupancy profile of the middle panel of Fig. 2 becomes clear if one notes that it is the superposition of the stable $N = 11$ and $N = 12$ configurations. This suggests that the transition from the $N = 11$ to the $N = 12$ configuration as a function of the chemical potential should be viewed as a ''micro'' first-order phase transition with the free energies of the $N = 11$ and $N = 12$ states degenerate at the transition point. The middle panel would represent a form of finite-size phase coexistence. Indeed, plots of the mean number of nucleosomes $\langle N \rangle$ of the section as a function of the chemical potential μ for different values of T/T_0 (see Fig. 3) are closely similar to magnetization versus applied field isotherms of a 1D Ising model with a transition temperature T_0. For T/T_0 small compared to one, a well-defined, rapid switch takes place between the $N = 11$ and $N = 12$ states as a function of μ. We performed Monte Carlo simulations [10] of this section of the chain to determine the statistical distribution of the overlap parameter Ψ at $\mu/k_BT_0 \approx 47.9$, where the $N = 11$ and $N = 12$ states are degenerate at low temperatures (see the inset of Fig. 3). For low temperatures ($T/T_0 \ll 1$), the distribution is bimodal and sharply peaked at 0.197 ($N = 12$) and 0.373 ($N = 11$) while for high temperatures ($T/T_0 \gg 1$) there is a single, much broader, Gaussian peak at a somewhat smaller peak position. At intermediate temperatures ($T/T_0 = 2$; see the inset of Fig. 3), the distribution is a superposition of the two peaks at 0.197 and 0.373 coexisting with a reduced Gaussian peak, which further confirms the first-order nature of the transition.

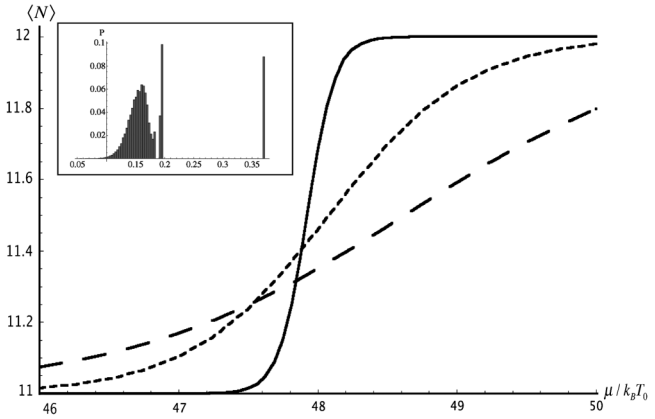

FIG. 3. Average number of nucleosomes between sites 6970 and 8730 of yeast chromosome II as a function of chemical potential. Solid line: $T/T_0 = 0.1$ Short dashed line: $T/T_0 = 0.5$. Long dashed line: $T/T_0 = 1$. Inset: Histogram of the distribution of the overlap parameter Ψ for the same stretch of DNA obtained from Monte Carlo simulations at $\mu \approx 47.9$ where the 11 and 12 nucleosome configurations are energetically nearly degenerate for $T/T_0 = 2$.

The appearance of micro first-order transitions is not an accidental specific of the DNA of yeast chromosome II. Indeed, if we replace V_n with a Gaussian random site potential having the same rms width as the V_n then the fraction of unstable sequences remains comparable to that of yeast chromosome II. The transitions are in fact a generic consequence of the unavoidable competition at higher occupancy levels between arrangements with different density but the same free energy.

The nucleosome conformation in the switching regions is thus exquisitely sensitive to small changes in the chemical potential and the characteristic energy scale k_BT_0 of the DNA-nucleosome interaction. Cells of different types or different developmental states with different nucleosome densities are known to be characterized by chemical modification of the nucleosomes (methylation or acetylation) that would alter this characteristic energy scale. The model predicts that changes in nucleosome density in response to chemical modification will be mainly localized to the switching regions. A corollary of this prediction is the possibility that the location of the switching regions on the chromosome correlates with the location of gene-regulatory sequences, specifically the binding sites of transcription factors. To test this hypothesis, we collected 278 published TF binding locations on chromosome two from the Saccharomyces Genome Database. We equilibrated the system at ~75% and ~90% average occupancies and calculated the absolute value of the difference in the probability of site occupancy. The mean (absolute) change was found to be 22.7%. When we calculated the mean absolute change in occupancy on the restricted set of the 278 TF binding locations, we found it to be 32.9%. To

check for the statistical significance of this result, we randomly chose 250 different sets of 10 base pairs regions and calculated the average absolute change in site occupancy on changing the density. We repeated this procedure and generated a distribution of average absolute changes and found a standard deviation of 2% from the mean. This means that the statistical probability of randomly achieving a mean value of 32.9% is less than 10^{-7}. It follows that at least some TF binding sites are strategically placed on segments of DNA that on average are more likely to reconfigure to changes in nucleosome concentration or affinity.

In summary, we have shown that a simple 1D model for nucleosome positioning, that correctly predicts 50% of the nucleosome locations along yeast chromosome II, is characterized by a sequence of highly localized, first-order rearrangement transitions that occur as a function of the characteristic interaction energy between the DNA and nucleosomes. The transitions are a generic consequence of frustration between the requirements of the excluded-volume interactions and the on site potential energy. The location of the first-order rearrangement transitions correlates with the location of TF binding sites, indicating that the frustration is exploited by nature as gene-regulatory switches that distinguish cells in different states of development.

This work was supported by the NSF under DMR Grant No. 04-04507.

*djs47@physics.ucla.edu

[1] R. Kornberg and L. Stryer, Nucleic Acids Res. **16**, 6677 (1988).
[2] L. Tonks, Phys. Rev. **50**, 955 (1936); J. K. Percus, J. Stat. Phys. **15**, 505 (1976).
[3] E. Segal *et al.*, Nature (London) **442**, 772 (2006).
[4] A. W. Jaeger and C. C. Kuenzle, EMBO J. **1**, 811 (1982).
[5] We are assuming that two nucleosomes at minimum spacing still are separated by a ten base pair sequence that is not in direct contact with either nucleosome.
[6] The reference temperature depends on the choice of DNA sequence. Under *in vitro* conditions it could be tuned by the ambient salt concentration.
[7] The probability distribution of the on site potentials is approximately Gaussian with a width $\Delta V/k_B T_0$ 9.3 and correlation length of about 150 bp.
[8] J. K. Percus, J. Stat. Phys. **15**, 505 (1976).
[9] Even though this is a 1D system, it can exhibit a true thermodynamic freezing transition for the case of a sinusoidal site potential with random phase; see M. Q. Zhang, J. Phys. A **24**, 3949 (1991).
[10] We performed the simulation under a restricted set of 100 000 configurations.