Nucleosome-mediated cooperativity between transcription factors

Nucleosome-mediated cooperativity between transcription factors

Leonid A. Mirny[1]

Harvard–MIT Division of Health Sciences and Technology, and Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited* by José N. Onuchic, University of California San Diego, La Jolla, CA, and approved October 12, 2010 (received for review December 3, 2009)

Cooperative binding of transcription factors (TFs) to promoters and other regulatory regions is essential for precise gene expression. The classical model of cooperativity requires direct interactions between TFs, thus constraining the arrangement of TF sites in regulatory regions. Recent genomic and functional studies, however, demonstrate a great deal of flexibility in such arrangements with variable distances, numbers of sites, and identities of TF sites located in cis-regulatory regions. Such flexibility is inconsistent with cooperativity by direct interactions between TFs. Here, we demonstrate that strong cooperativity among noninteracting TFs can be achieved by their competition with nucleosomes. We find that the mechanism of nucleosome-mediated cooperativity is analogous to cooperativity in another multimolecular complex: hemoglobin. This surprising analogy provides deep insights, with parallels between the heterotropic regulation of hemoglobin (e.g., the Bohr effect) and the roles of nucleosome-positioning sequences and chromatin modifications in gene expression. Nucleosome-mediated cooperativity is consistent with several experimental studies, is equally applicable to repressors and activators, allows substantial flexibility in and modularity of regulatory regions, and provides a rationale for a broad range of genomic and evolutionary observations. Striking parallels between cooperativity in hemoglobin and in transcriptional regulation point to a general mechanism that can be used in various biological systems.

protein–DNA interactions | promoter | enhancer | histone | Monod–Wyman–Changeux

In higher eukaryotes, cis-regulatory regions are 200 to 3,000 base pairs (bps) in length and may contain clusters of 3 to 50 TF binding sites (TFBSs) (1–4). The arrangement, identity, and affinity of the sites determine the function of the regulatory region. Cooperative binding of TFs to regulatory regions leads to highly cooperative gene activation and is essential for development (5) and other vital processes (6).

Cooperative binding is traditionally explained by protein–protein interactions among TFs (7, 8). While this mechanism finds support in bacterial and some eukaryotic systems (9), several functional and genomic observations are inconsistent with it. Cooperativity by protein–protein interactions (directly or via DNA looping) (7, 8, 10) can significantly constrain arrangements of TFBSs, allowing only those that provide the correct orientation, order, and distance between TFs. On the contrary, recent evolutionary analysis of *Drosophila* enhancers revealed massive turnover and rearrangements of TFBSs (11, 12). Furthermore, functional studies demonstrated that cis-regulatory regions could tolerate incorporation of new binding sites (promiscuity) and significant alterations in TFBS placement while retaining in vivo functionality (11, 13, 14). The few mechanisms that have been proposed to explain flexible arrangements of TFBSs and promiscuity are based on the idea of transcriptional synergy (i.e., cooperative recognition or simultaneous contact between TFs and some part of the transcription machinery) (13, 14), rather than cooperative binding of TFs to DNA.

An alternative mechanism of cooperativity considered here is based on synergistic binding of noninteracting TFs mediated by a nucleosome. The phenomenon of nucleosome-mediated cooperativity has been documented by a series of in vivo and in vitro experiments (6, 15–17), which demonstrated synergistic binding and gene activation by nonendogenous TFs (e.g., Gal4 and LexA) that occupied sites on nucleosomal DNA (Fig. 1). Such cooperativity requires only the DNA-binding domains of TFs, suggesting that it does not involve chromatin modification or direct protein–protein interactions (18). Experimental studies (16) and an earlier model (19) of synergistic binding to nucleosomal DNA considered only two close-by (~30 bps) sites (Fig. 1*B*) that interact through an assisted unwrapping mechanism: binding of the first TF leads to partial unwrapping of nucleosomal DNA, thus making the site of the second TF more accessible (19).

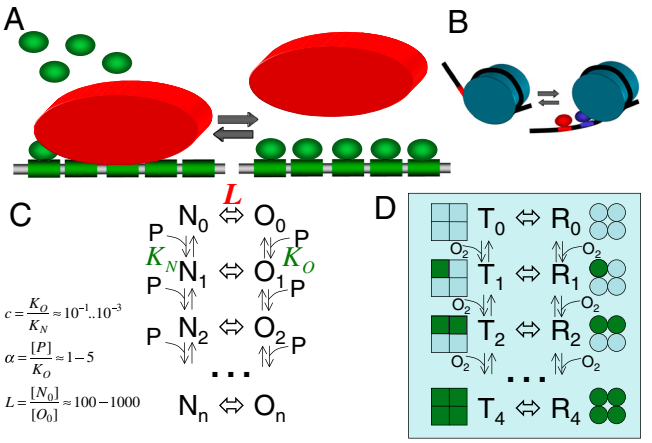

Fig. 1. The model of nucleosome-mediated cooperativity. (*A*) DNA region containing an array of *n* sites (green boxes) that can be bound by a histone core (red oval), thus becoming nucleosomal DNA, or remain naked. In either the nucleosomal (*N*) or the open (*O*) state, the DNA can be bound by transcription factors (TFs, green ovals). Binding of TFs to the nucleosomal DNA is diminished as compared to naked DNA but is possible due to transient, partial unwinding of the DNA (shown in *B*) (36). The equilibrium of the system is fully characterized by the scheme in *C*. The states of the system: nucleosomal (N_i) and open (O_i), with *i* being the number of TFs bound. In this form, the nucleosome-TF system is identical to the Monod–Wyman–Changeux (MWC) model of cooperativity in hemoglobin (*D*). The *N* and *O* forms of the DNA correspond to the *T* and *R* states of hemoglobin; TF binding is equivalent to O_2 binding (see Table S1). Like the MWC model, the nucleosome-TF system is determined by three dimensionless parameters: *L*, *c* and α (see text). (*B*) The model of Polach and Widom (19): synergistic binding by two TFs to nearby sites through partial unwinding of nucleosomal DNA. The mechanism requires the constant presence of a nucleosome and does not lead to nucleosome eviction. In our model, binding of multiple TFs can evict a nucleosome completely thus allowing more distant sites to interact, more sites to be involved and hence a higher Hill coefficient.

Author contributions: L.M. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]E-mail: leonid@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.0913805107/-/DCSupplemental.

Here we introduce and analyze a different mechanism of nucleosome-mediated cooperativity. We show that competition between histones and TFs for a region of DNA that bears an array of TFBSs induces strongly cooperative binding of TFs and cooperative nucleosome eviction. We find this mechanism of cooperativity is identical to the Monod–Wyman–Changeux (MWC) model of cooperativity in hemoglobin (20). Using this analogy, we gain deeper insights into a range of phenomena such as the role of nucleosome-positioning sequences and histone modifications, low-affinity TFBS, and TFBS clustering (see Table S1). Finally, we review experimental evidences in support of our nucleosome-mediated mechanism (Table S2).

Presented mechanism of nucleosome-mediated cooperativity is sufficiently general, provides a high Hill coefficient (21), and leads to passive nucleosome eviction. These aspects distinguish it from assisted unwrapping (19), which requires that a nucleosome remains in place but cannot achieve a high Hill coefficient due to gradual unwrapping of nucleosomal DNA (see *SI Text*). Our framework, in essence, integrates DNA unwrapping as a mechanism of TF access to nucleosomal DNA, with the possibility of nucleosome eviction. The high energetic cost of nucleosome eviction and greater length of DNA that becomes accessible to TF binding upon eviction provide high cooperativity for TF binding. Note that our model does not consider active (ATP-dependent) nucleosome modification by recruited TFs. This important mechanism of chromatin remodeling may follow the initial cooperative binding of TFs and passive nucleosome eviction. Similarly, we do not consider the effect of nucleosome eviction on the positioning of neighboring nucleosomes (22, 23). A recent study (24) modeled the effect of two TFs binding their sites in nucleosomal DNA, demonstrating cooperative binding, which, however, do not exhibit a high Hill coefficient (*SI Text*, Table S3, and Fig. S3).

Proposed mechanism requires several TFBSs located within nucleosomal DNA (see refs. 7, 25, and 26 for examples). As such, it is more applicable to gene regulation in multicellular eukaryotes where the sites are shorter and more sites are clustered together to form a regulatory region than to yeast where one to two sites can be sufficient (27).

In general, the input of a regulatory module is an external stimulus (e.g., a ligand or activation of an upstream kinase) and the output is gene expression. Here, the input is a concentration of TFs activated by such stimuli and poised to bind their sites. The output is TF occupancy of a regulatory region, rather than gene expression, because expression can be a complex function of the occupancy. Our model is equally applicable to repressors and activators, providing cooperative binding irrespective of the effect of bound TFs on gene expression.

Results

The Model of Nucleosome-Mediated Cooperativity. We consider interactions of TFs with a stable nucleosome, containing an array of n TFBSs within its DNA footprint (147 bps) (Fig. 1). This region of DNA can be in one of two states: the nucleosome (N) state and the open (O) state, in which histones are absent from the region. While histones limit access of other proteins to nucleosomal DNA, the nucleosome is highly dynamic, with DNA unwrapping and wrapping at a high rate, thus making nucleosomal DNA at least partially accessible to TFs (28, 29). TFBSs can be occupied by TFs in either the N or O state leading to $2n$ states labeled N_i and O_i ($i = 0,1,\ldots n$), where i is the number of occupied sites. The equilibrium between the N and O states in the absence of bound TFs is characterized by the constant $L = [N_0]/[O_0]$, where $L \gg 1$ for a stable nucleosome. The affinity of TFs for the sites depends on the state, with binding constants K_N and K_O ($K_O \ll K_N$ due to higher affinity in the open state). Suppression of TF binding to nucleosomal DNA is reflected by the free energy cost of DNA unwrapping required to accommodate one more TF (30) and is taken into account by parameter $c \equiv K_O/K_N \ll 1$.

For simplicity of presentation, we assume all TFBSs to have the same affinity and experience the same suppression by the nucleosome. More complicated models that drop these assumptions and consider partial unwrapping are considered in *SI Text* but yield similar results.

TFs can be activated by different mechanisms: increased intranuclear concentration of a TF (e.g., by facilitating its nuclear localization or inhibiting degradation) or elevated TF affinity for its cognate sites (e.g., through a conformational change upon a modification or binding to a ligand). Both factors are taken into account by a dimensionless parameter $\alpha = [P]/K_O$, where $[P]$ is the intranuclear concentration of the activated TF. In equilibrium, the system is fully determined by three dimensionless parameters: c, L, and α (Fig. 1). We study equilibrium properties, assuming that rapid exchange of TFs and histones leads to fast equilibration.

Cooperative Binding and Nucleosomal Occupancy. The two quantities of primary biological interest are TF occupancy per site (Y) and the occupancy by the nucleosome (Y_N). Using statistical mechanics, and analogy to the MWC model, we obtain expressions for these quantities (see *SI Text*):

$$Y = \alpha \frac{(1+\alpha)^{n-1} + Lc(1+c\alpha)^{n-1}}{(1+\alpha)^n + L(1+c\alpha)^n}, \quad [1]$$

$$Y_N = \frac{L(1+c\alpha)^n}{(1+\alpha)^n + L(1+c\alpha)^n}. \quad [2]$$

Fig. 2*A* presents TF and nucleosome occupancies as a function of TF concentration, computed using parameters inferred from experiments with $\alpha = 0$–10, $L \approx 10^3$, $c \approx 10^{-2}$ (see *Materials and Methods*). Strikingly, nucleosome-mediated cooperativity results in a sharp transition with a twofold increase in TF concentration leading to a more than eightfold increase in the occupancy. The nucleosome occupancy also changes cooperatively, dropping from about 65% to less than 10% due to a twofold change in TF concentration.

Complex cis-regulatory elements of higher eukaryotes may require several activating TFs (or several copies of the same TF) to be bound for initiation of gene expression (4, 31). To take this into account, we calculate another quantity, the probability of having at least k TF bound, P_k, as a measure of occupancy, which also shows a significant cooperativity (Fig. 2*B* and *SI Text*).

Importantly, as in the case of hemoglobin, the cooperativity stems from suppression of TF binding at low TF concentration (Fig. 2*A*, dotted vs. solid green lines). This behavior is distinct from cooperative binding enhanced by attractive protein–protein interactions. Cooperative suppression of TF binding by nucleosomes, however, could be advantageous in multicellular eukaryotes that contain an overwhelming number of spurious binding sites in their genomes (12, 27).

Surprisingly, our model suggests that nucleosome destabilization should have opposite effects on cooperativity and on TF binding. Factors that destabilize nucleosomes, [e.g., histone modification, poly(dA:dT), etc.] facilitate TF binding but make this binding less cooperative. Similarly, factors that stabilize nucleosomes suppress binding at a low TF concentration, making binding more cooperative. In *Discussion*, we review recent genomics findings that reveal signatures of nucleosome stabilization in human regulatory regions (26).

Analogy to Hemoglobin. Strikingly, the system of TFs and a nucleosome is identical to the scheme of cooperativity in hemoglobin described by the classical MWC model (20) (see Fig. 1 *C* and *D*). Table S1 summarizes equivalent parameters and analogous phenomena between the two systems. The MWC model consid-

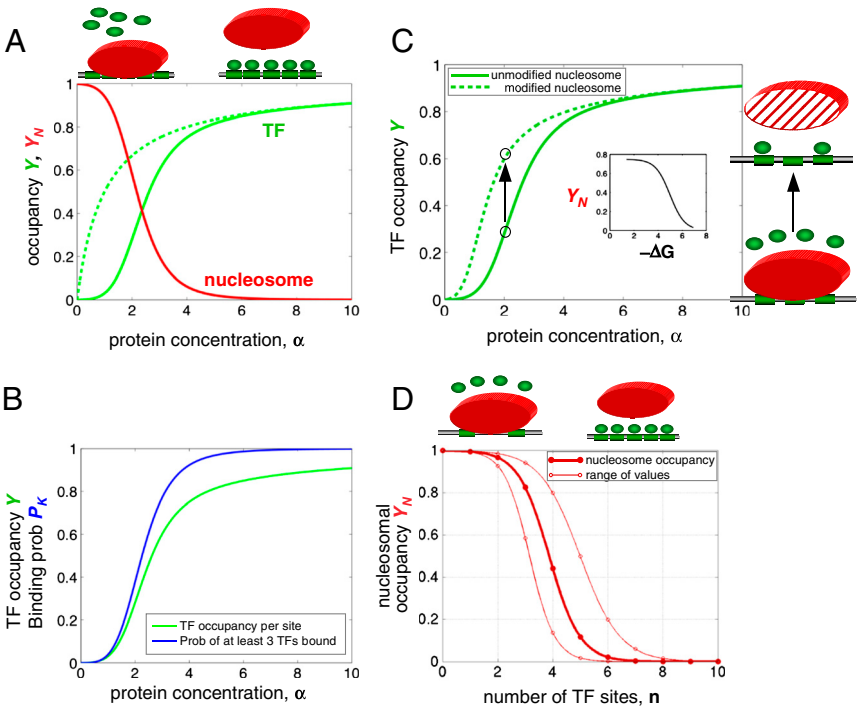

Fig. 2. Nucleosome-mediated cooperativity and its implications. (*A*) Cooperative transition in the equilibrium TFBS occupancy, Y (solid green line), and nucleosome occupancy, Y_N (red line), as a function of TF concentration α (Eqs. **1** and **2**, $n = 6$, $L = 10^3$, $c = 10^{-3}$). Notice that nucleosome-mediated cooperativity leads to suppression of TF binding at a low concentration, as compared to noncooperative binding (dashed green line). (*B*) Probability of having at least 3 TFs bound P_3 as a function of TF concentration (blue). Mean occupancy per site, Y, is shown for comparison (green). (*C*) The Bohr effect: attenuation of histone affinity for DNA, due to modifications or as a function of DNA sequence (modified—dashed line, unmodified—solid line), leads to a shift in TF-nucleosome competition and displacement of the nucleosome by TFs (arrow). This competition renders nucleosomal occupancy, Y_N, considerably responsive to small changes in nucleosome affinity (*Inset*), as demonstrated by the dependence of Y_N on $-\Delta G = k_B T \log(L/L')$ (kcal/mol). (*D*) The effect of number of TF sites, n, on nucleosome stability, obtained for three concentrations of TF: $\alpha = 3, 5, 8$ (lines from top to bottom). There is a critical number of sites (~4–5) below which TFs are unable to displace a nucleosome and above which the nucleosome is unstable even at a lower concentration of TFs.

ers the equilibrium between two states of hemoglobin: the R state, which has a higher affinity for O_2, and the low-affinity T state. In the absence of the O_2, hemoglobin is mostly in the T state. Binding of O_2 shifts the equilibrium toward the R state, making binding of successive oxygen molecules more likely and thus cooperative (Fig. 1*D*). The R and T states of hemoglobin correspond to the O and N states of the nucleosomal DNA, and the binding of O_2 to hemoglobin domains corresponds to the binding of TFs to individual sites (Fig. 1 *C* and *D* and Table S1).

The analogy to the MWC model allows us to reveal features of the nucleosome–TF system that are essential for cooperativity (Table S1). The MWC system has a strong cooperativity as long as L is sufficiently large ($L > 100$) and c is sufficiently small ($c < 0.1$), i.e., the nucleosome is stable (in the absence of TFs) and significantly attenuates TF binding. These requirements are consistent with estimated parameters (see *Materials and Methods*), as well as with high stability (32, 33) and slow exchange (34, 35) of nucleosomes in regions depleted in TFBSs. In vitro studies of TF binding to nucleosomal DNA demonstrate the required attenuation of TF binding (36).

The Bohr Effect and Chromatin Modification. We also use the analogy to hemoglobin to examine implications of sequence-specific nucleosome positioning, histone modifications, and other processes involved in gene regulation. These effects can be considered as allosteric heterotropic regulation of the nucleosome–TF system, analogous to heterotropic effectors of hemoglobin. A prototypical heterotropic allosteric regulation of hemoglobin is the Bohr effect: Lowering the pH decreases the affinity to oxygen, thus providing more oxygen to actively working muscles. The basis of the Bohr effect is the higher affinity of hydrogen ions to the T state. Thus, low pH stabilizes the T state, shifting the equilibrium away from the high-affinity R state. Other allosteric effectors of hemoglobin (e.g., DPG) act in a similar way: Binding hemoglobin in one state affects the R-T equilibrium and thus changes the affinity of hemoglobin to oxygen.

While heterotropic effectors of hemoglobin affect equilibrium between R and T states, effectors of the nucleosomes–TF system such as histone modifications and histone-binding proteins influence nucleosome stability, thus altering the balance between N and O states. Fig. 2*C* shows the manifestation of the Bohr effect in the TF-nucleosome system: Small changes in nucleosomal affinity (from L to L') due to histone modifications can shift the balance in TF-nucleosome competition toward or away from the nucleosome. For example, a modification that reduces nucleosome stability by about $\Delta G = 1$ kcal/mol ($\Delta G = k_B T \log(L/L')$) can lead to an 80% drop in nucleosome occupancy (Fig. 2*C*, *Inset*) and a concurrent rise of the TF occupancy (Fig. 2*C*). Similarly, nucleosome-positioning sequences have a similar effect: They alter nucleosome stability, thus shifting the occupancy curve. Competition between the nucleosome and TFs leads to amplification of the nucleosome-positioning sequence signal, i.e., small changes in histone affinity translate into significant changes in nucleosome occupancy (Fig. 2*C*, *Inset*).

Similarly, small changes in TF concentration can significantly reduce nucleosomal occupancy (Fig. 2*A* and Fig. S3*C*) in TFBS-rich regions. For example, activation of a tissue-specific TF can lead to selectively reduced chromatization and increased accessibility of tissue-specific regulatory regions (Fig. 3). This is in agreement with a recent genome-scale mapping of DNase I hypersensitive sites (DHS), which found that several loci exhibit tissue-specific DHS profile (37). Note that this mechanism of passive nucleosome eviction does not rely on recruitment of chromatin modification machinery, which may play a role in further destabilizing nucleosomes and expanding nucleosome-free regions.

Critical Size of the TFBS Cluster. Nucleosome-induced cooperativity, however, has some properties without counterparts in hemoglobin. For example, the number and the affinity of binding sites are constant in hemoglobin but vary in cis-regulatory regions. Fig. 2*D* presents nucleosomal occupancy as a function of the number, n, of TFBSs. As the number of TFBSs exceeds a certain critical value n_c, nucleosomal occupancy drops sharply, manifesting another allosteric effect in the system. Our calculations show that the critical number of TFBSs is given by $n_c \approx \log(L)/\log(1+\alpha)$, yielding a narrow range $n_c = 3$–6 that is not very sensitive to model parameters (see *SI Text*). This range of TFBSs per nucleosomal footprint is consistent with the recent characterization of *Drosophila* enhancers (1, 3) that contain about 20 TFBSs per 0.7–1 Kb

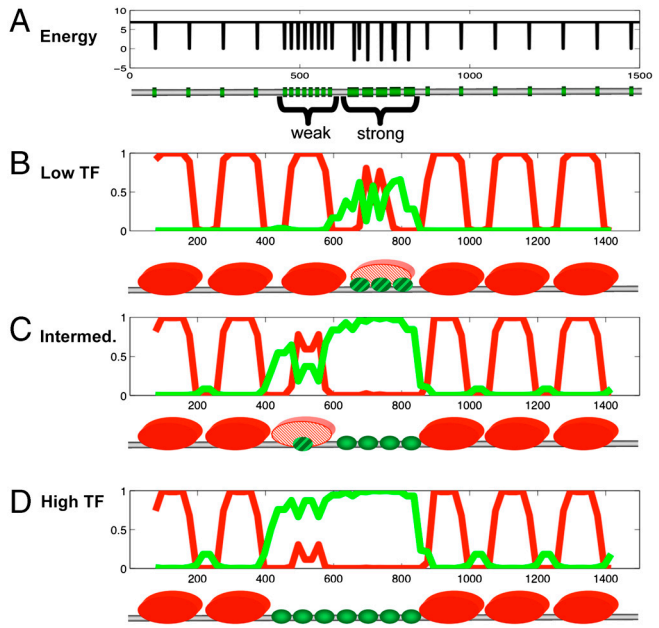

Fig. 3. Cooperative binding to high- and low-affinity sites. The nucleosomal (red) and TF (green) occupancy profiles for a regulatory region that contains clusters of high- and low-affinity sites. (*A*) The binding energy profile: a cluster of 8 low-affinity sites and a cluster of 5 high-affinity sites located over the background of spurious low-affinity sites (27). The region contains seven stable, equally spaced nucleosomes with a liner of 50 bp. (*B–D*) Diagrams show nucleosomal occupancy Y_N (red), and TF cluster occupancy P_3 for three values of TF concentration. While an intermediate TF concentration is sufficient to get high-affinity clusters nucleosome-free and TF-bound, a higher concentration is needed for low-affinity clusters. A combination of low- and high-affinity sites in a regulatory region can result in different responses to various TF concentrations. Notice that isolated low-affinity sites are unable to displace nucleosomes. Nucleosomes were assumed to be well-positioned by DNA sequence, and sliding was disregarded. The following parameters were used: $c = 0.01$; $L = 1{,}000$, $\alpha_{\text{non-site}} = 0.001$; $\alpha_{\text{high-affinity}} = 20$; $\alpha_{\text{low-affinity}} = 1$.

(i.e., 4–6 sites per nucleosome) (2). Similarly, the cis-regulatory system of *endo16* in sea urchins has about 50 sites localized within 2.3 Kb and grouped into seven clusters of five to 10 sites (4). Such passive eviction of nucleosomes could explain widespread depletion and rapid exchange of nucleosomes in TFBS-rich regions (33, 34, 38–40).

This cluster size is consistent with our recent information-theoretical estimates (27) of the minimal significant TFBS cluster. Nucleosomes-mediated cooperativity can suppress binding to widespread spurious sites that are unlikely to cluster, while providing binding to clusters of sites. An example on Fig. 3 demonstrates that clusters of five high-affinity (or eight low-affinity TFBSs) become occupied and nucleosome-free, while isolated sites remain unoccupied by TFs.

Our approach allows one to consider the contributions of low-affinity sites (9, 41) and mixtures of sites of different TFs (see *SI Text*). Fig. 3 illustrates how arrays of low-affinity and high-affinity sites in nucleosomal DNA respond differently to increasing levels of TFs.

Discussion

Below we discuss several experimental results that support the nucleosome-mediated mechanism of cooperativity, summarized in Table S2, and propose direct experimental tests of the mechanism. We believe that the proposed mechanism contributes to the complex interplay of nucleosome-positioning DNA sequences (23, 42–46), TFs interacting with each other (2) as well as with histones (47) and regulating gene expression (25, 41). The relative importance of different mechanisms of cooperativity (16) and nucleosome positioning (23) may vary for different DNA regions and different organisms (46). A recent study examined intrinsic and extrinsic nucleosome-positioning factors and demonstrated that taking into account TF-nucleosome competition improves the accuracy of nucleosome position predictions (23).

Most direct evidences of nucleosome-mediated cooperativity come from experimental studies that demonstrated cooperative binding of nonendogenous TFs without involvement of protein–protein interactions and for a range of up to 200 bps (15–17). Moreover, experiments with TFs lacking activation domains have shown that synergistic activation of gene expression is determined more by the number of TFBSs than by the interactions with general TFs, polymerase, or chromatin modification machinery (31, 48). Consistent with the nucleosome-induced mechanism, transcomplementation experiments on stripe 2 enhancers demonstrated that precise expression does not require special Bcd–Hb interactions and can be achieved by chimeric and nonendogenous (i.e., noninteracting) TFs (9). The range of nucleosome-induced cooperativity (~150–200 bps) is also consistent with the modularity of the *otd* enhancer, which contains two 180-bp TFBS clusters, each able to provide the correct expression pattern (49).

Presented mechanism also ties together several observations in genomics. Passive nucleosome eviction can explain how low nucleosomal density is maintained on cis-regulatory regions and how sharp boundaries of such nucleosome-depleted regions are achieved. The critical number of sites $n_c = 3$–5 (see above) required for the TF-induced nucleosome displacement is consistent with clustered arrangements of TFBSs and can explain why such clustering serves as a powerful criterion for bioinformatic identification of cis-regulatory regions (3, 50). Our mechanism suggests a possible role for low-affinity TFBSs, which are abundant (2, 51) and essential (9) in *Drosophila* in assisting high-affinity sites to displace nucleosomes. Cooperative nucleosome displacement by competing TF serves, along with sequence information (42–46, 52), to determine nucleosome positioning but, in contrast, can be tissue or condition specific.

The nucleosome-mediated mechanism allows significant flexibility in arrangements of TFBSs while retaining cooperativity, requiring only that a sufficient number of sites are located sufficiently deep inside a nucleosomal footprint. Widespread turnover of sites in enhancers (1–3) and the paucity of local protein–protein interactions (for example, in a crystal structure of interferon-beta enhanceosome) (53) are difficult to reconcile with the classical model of cooperativity by protein–protein interactions. The nucleosome-mediated mechanism can explain observed promiscuity of regulatory regions: Unrelated TFs can cooperate by evolving proximal TFBSs or by site duplication and divergence. A new TF can become a part of an existing assembly by acquiring TFBSs within an existing cluster, a fairly fast and widespread evolutionary process (12, 54). A classical model of cooperativity, in contrast, requires interacting TFs to evolve protein–protein interfaces used for interactions—a much slower evolutionary process.

The nucleosome-mediated mechanism, however, works only for TFBSs separated by at most 150 bps. This range can be increased by synergy of nearby nucleosomes (55) and spread much further through recruitment of chromatin modification machinery and positive feedback in this process (56).

The proposed mechanism makes concrete predictions that can be tested experimentally. First, it suggests a dependence of the cooperative effect and nucleosome occupancy on the number of TFBSs. Both the nucleosome occupancy and level of gene expression can be assayed directly for synthetic or natural promoters that contain TFBS clusters of different densities and spacings and that are also confirmed to be SWI/SNF-independent. Our model specifically predicts that cooperativity (as measured by the Hill coefficient) increases with the number of sites and is greater for sites located closer to the nucleosome center. Second,

the cooperativity of TF binding is predicted to depend on nucleosome stability in a nontrivial fashion: Stabilization of the nucleosome (e.g., by strong nucleosome-positioning DNA sequences) is expected to make binding and expression more cooperative, while destabilization of the nucleosome [e.g., by poly(dA:dT) sequences (57, 58) or nucleosome eviction by a nonendogenous TF or polymerase] is expected to decrease cooperativity. This prediction can be used to distinguish our mechanism from the effect of direct protein–protein interactions that are expected to exhibit the opposite dependence on nucleosome stability. A recent study of human nucleosome positioning (26) provided evidences consistent with our predictions. First, it reports elevated in vivo nucleosomal occupancy of human regulatory sequences. This observation is consistent with high nucleosome stability required by nucleosome-mediated cooperativity. Second, this study demonstrates that such high nucleosomal occupancy is encoded in the DNA sequences of regulatory regions, and such sequences are depleted in nucleosome-repelling rigid poly(dA:dT). Again, the depletion of poly(dA:dT) supports the requirement of our model for stable nucleosome positioning. Moreover, poly(dA:dT) that flank TFBS clusters can further stabilize TF-competing nucleosomes by preventing their sliding. Removal of such sequences, we predict, will reduce the degree of cooperativity, while increasing the binding affinity for TFs.

Nucleosome-mediated cooperativity works particularly well for clusters of several TFBSs and thus may be utilized more by higher eukaryotes with more complex regulatory regions (4). Yeast, to the contrary, has simpler promoters with individual sites, rather than clusters carrying regulatory potential (27), and may rely less on nucleosome-mediated cooperativity. Consistent with this argument is the observation that yeast has a lower nucleosome occupancy and lower intrinsic nucleosome propensity of regulatory regions (44, 59).

In summary, we have shown how competition between a nucleosome and TFs can lead to cooperative binding of TFs and cooperative eviction of nucleosomes from regulatory regions. We have established and employed the analogy between this process and cooperativity in hemoglobin according to the MWC model. This analogy has allowed us to consider chromatin modification and nucleosome-positioning sequences as heterotropic allosteric effectors, similar to the Bohr effect. Most importantly, presented mechanism explains several observations in comparative and functional genomics that cannot be reconciled with the classical model of cooperativity via direct protein–protein interactions or a simplistic view of nucleosomes as suppressors of gene expression. Our study provides a mechanistic rationale for widespread flexibility in arrangement of binding sites, allowing highly evolvable regulatory regions. Finally, the analogy between cooperativity in hemoglobin and nucleosome-mediated cooperativity of TFs emphasizes a widespread MWC mechanism of cooperativity observed in a range of biological systems from protein folding (60) to receptor coupling in bacteria (61), ligand-gated ion channels, and enzymatic phosphorylation (62).

Materials and Methods

Approach. We use a statistical mechanics approach to derive occupancy and other equilibrium properties of the system presented in Fig. 1 (see *SI Text*). The advantage of our approach is that it allows generalization for more complex cases considered in *SI Text*.

Estimation of Parameters from Experimental Data. To estimate α, we take into account TF binding to both specific and nonspecific DNA, with binding constants K and K^{NS}, respectively. Due to competition between specific and nonspecific DNA the occupancy of a single site is $y = [P]/([P]+K(1+[\mathrm{DNA}]/K^{NS})) = \alpha/1+\alpha$, where $\alpha \equiv [P]/K^{eff}$, effective binding constant defined as $K^{eff} = K(1+[\mathrm{DNA}]/K^{NS}) \approx K[\mathrm{DNA}]/K^{NS}$, and [DNA] is the concentration of TF-accessible nonspecific DNA. The dissociation constant of most eukaryotic TFs is in the range of $K \approx 1$–10 nM, while known nonspecific binding constants are $K^{NS} \approx 1$–10 μM (63, 64). Using the length of genomic DNA and the measured copy number of TFs ($[P] \approx 500$–$5{,}000$ in yeast and $[P] \approx 10^4$–10^5 protein copies per nucleus in multicellular eukaryotes) (65) (BioNumbers database, http://bionumbers.hms.harvard.edu/) with the assumption of 90% chromatinization, we obtain the range of $\alpha \approx 0.5$–5.

We estimate L using in vivo nucleosome equilibrium occupancy $f = [N]/([N]+[O])$, yielding $L = f/(1-f)$. Although for stable nucleosomes, occupancy is very close to 1, the fraction of DNA covered by nucleosomes provides a lower bound for f and has a range of 0.9–0.99 (66), yielding $L > 10$–100. This constitutes a lower bound for f and L, because most nucleosome-free regions are maintained by competition with TFs or active chromatin modification.

Parameter c of the model reflects suppression of TF binding by a nucleosome while in the N state. Acting through steric hindrance, such suppression is not permanent and nucleosomal DNA becomes transiently exposed for TF binding (Fig. 1*B*) (36, 67). This suppression is equivalent to the experimentally measured equilibrium constant of site exposure, which depends on the location of the site with respect to the center of the nucleosome and has the range $c = 2{\cdot}10^{-2}$–10^{-5}. Detailed treatment of partially unwrapped states (see *SI Text*) shows that they can be aggregated into a single N state.

Obtained estimated values of parameters are sufficient to provide nucleosome-mediated cooperativity. The mechanism of cooperativity is very robust, requiring only $L \gg 1$ and $c \ll 1$ for the onset of cooperativity (Fig. S1).

ACKNOWLEDGMENTS. I am grateful to Mehran Kardar, Stanislav Shvartsman, Paul Wiggins, Jane Kondev, Ned Wingreen, Rob Phillips, George Benedek, and Ilya Ioschikhes for insightful discussions of this study; to John Stamatoyannopoulos for carefully reading of and commenting on the manuscript as well as providing several important suggestions; and to Jason Leith, Zeba Wunderlich, and Albert Liau for discussing and editing the manuscript. Part of this work took place at the Kavli Institute for Theoretical Physics, Santa Barbara, CA, and the Aspen Center for Physics, Aspen, CO. I acknowledge the support of the National Institutes of Health-funded National Center for Biomedical Computing, *i2b2*, Informatics for Integrating Biology and the Bedside, and National Cancer Institute-funded Physical Sciences in Oncology Center at MIT.

1. Schroeder MD, et al. (2004) Transcriptional control in the segmentation gene network of *Drosophila*. *PLoS Biol* 2:e271.
2. Papatsenko DA, et al. (2002) Extraction of functional binding sites from unique regulatory regions: The *Drosophila* early developmental enhancers. *Genome Res* 12:470–481.
3. Berman BP, et al. (2004) Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in *Drosophila melanogaster* and *Drosophila pseudoobscura*. *Genome Biol* 5:R61.
4. Davidson EH (2006) *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution* (Elsevier/Academic, Amsterdam; Boston).
5. Lebrecht D, et al. (2005) Bicoid cooperative DNA binding is critical for embryonic patterning in *Drosophila*. *Proc Natl Acad Sci USA* 102:13176–13181.
6. Pettersson M, Schaffner W (1990) Synergistic activation of transcription by multiple binding sites for NF-kappa B even in absence of co-operative factor binding to DNA. *J Mol Biol* 214:373–380.
7. Ptashne M, Gann A (2002) *Genes & Signals* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).
8. Bintu L, et al. (2005) Transcriptional regulation by the numbers: Models. *Curr Opin Genet Dev* 15:116–124.
9. Arnosti DN, Barolo S, Levine M, Small S (1996) The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* 122:205–214.
10. Saiz L, Vilar JM (2006) DNA looping: The consequences and its control. *Curr Opin Struct Biol* 16:344–350.
11. Hare EE, Peterson BK, Iyer VN, Meier R, Eisen MB (2008) Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genet* 4:e1000106.
12. Moses AM, et al. (2006) Large-scale turnover of functional transcription factor binding sites in Drosophila. *PLoS Comput Biol* 2:e130.
13. Arnosti DN, Kulkarni MM (2005) Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem* 94:890–898.
14. Lin YS, Carey M, Ptashne M, Green MR (1990) How different eukaryotic transcriptional activators can cooperate promiscuously. *Nature* 345:359–361.
15. Miller JA, Widom J (2003) Collaborative competition mechanism for gene activation in vivo. *Mol Cell Biol* 23:1623–1632.
16. Vashee S, Melcher K, Ding WV, Johnston SA, Kodadek T (1998) Evidence for two modes of cooperative DNA binding in vivo that do not involve direct protein-protein interactions. *Curr Biol* 8:452–458.
17. Adams CC, Workman JL (1995) Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol Cell Biol* 15:1405–1421.
18. Morse RH (2003) Getting into chromatin: How do transcription factors get past the histones? *Biochem Cell Biol* 81:101–112.

19. Polach KJ, Widom J (1996) A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *J Mol Biol* 258:800–812.
20. Monod J, Wyman J, Changeux JP (1965) On the nature of allosteric transitions: A plausible model. *J Mol Biol* 12:88–118.
21. Hill AV (1910) The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves. *J Physiol* 40:1–7.
22. Wasson T, Hartemink AJ (2009) An ensemble model of competitive multi-factor binding of the genome. *Genome Res* 19:2101–2112.
23. Morozov AV, et al. (2009) Using DNA mechanics to predict in vitro nucleosome positions and formation energies. *Nucleic Acids Res* 37:4707–4722.
24. Raveh-Sadka T, Levo M, Segal E (2009) Incorporating nucleosomes into thermodynamic models of transcription regulation. *Genome Res* 19:1480–1496.
25. Lam FH, Steger DJ, O'Shea EK (2008) Chromatin decouples promoter threshold from dynamic range. *Nature* 453:246–250.
26. Tillo D, et al. (2010) High nucleosome occupancy is encoded at human regulatory sequences. *PLoS One* 5:e9129.
27. Wunderlich Z, Mirny LA (2009) Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet* 25:434–440.
28. Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. *Cell* 128:707–719.
29. Poirier MG, Bussiek M, Langowski J, Widom J (2008) Spontaneous access to DNA target sites in folded chromatin fibers. *J Mol Biol* 379:772–786.
30. Polach KJ, Widom J (1995) Mechanism of protein access to specific DNA sequences in chromatin: A dynamic equilibrium model for gene regulation. *J Mol Biol* 254:130–149.
31. Oliviero S, Struhl K (1991) Synergistic transcriptional enhancement does not depend on the number of acidic activation domains bound to the promoter. *Proc Natl Acad Sci USA* 88:224–228.
32. Lee W, et al. (2007) A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet* 39:1235–1244.
33. Lee CK, Shibata Y, Rao B, Strahl BD, Lieb JD (2004) Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat Genet* 36:900–905.
34. Dion MF, et al. (2007) Dynamics of replication-independent histone turnover in budding yeast. *Science* 315:1405–1408.
35. Phair RD, et al. (2004) Global nature of dynamic protein-chromatin interactions in vivo: Three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. *Mol Cell Biol* 24:6393–6402.
36. Anderson JD, Widom J (2000) Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *J Mol Biol* 296:979–987.
37. John S, et al. (2008) Interaction of the glucocorticoid receptor with the chromatin landscape. *Mol Cell* 29:611–624.
38. Sabo PJ, et al. (2004) Discovery of functional noncoding elements by digital analysis of chromatin structure. *Proc Natl Acad Sci USA* 101:16837–16842.
39. Valouev A, et al. (2008) A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res* 18:1051–1063.
40. Mavrich TN, et al. (2008) Nucleosome organization in the *Drosophila* genome. *Nature* 453:358–362.
41. Levine M (2008) A systems view of Drosophila segmentation. *Genome Biol* 9:207.1–207.3.
42. Trifonov EN, Sussman JL (1980) The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proc Natl Acad Sci USA* 77:3816–3820.
43. Ioshikhes I, Bolshoy A, Derenshteyn K, Borodovsky M, Trifonov EN (1996) Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences. *J Mol Biol* 262:129–139.
44. Ioshikhes IP, Albert I, Zanton SJ, Pugh BF (2006) Nucleosome positions predicted through comparative genomics. *Nat Genet* 38:1210–1215.
45. Segal E, et al. (2006) A genomic code for nucleosome positioning. *Nature* 442:772–778.
46. Peckham HE, et al. (2007) Nucleosome positioning signals in genomic DNA. *Genome Res* 17:1170–1177.
47. Kim HD, O'Shea EK (2008) A quantitative model of transcription factor-activated gene expression. *Nat Struct Mol Biol* 15:1192–1198.
48. Xu HE, Kodadek T, Johnston SA (1995) A single GAL4 dimer can maximally activate transcription under physiological conditions. *Proc Natl AcadSci USA* 92:7677–7680.
49. Gao Q, Finkelstein R (1998) Targeting gene expression to the head: the *Drosophila* orthodenticle gene is a direct target of the Bicoid morphogen. *Development* 125:4185–4193.
50. Sinha S, van Nimwegen E, Siggia ED (2003) A probabilistic method to detect regulatory modules. *Bioinformatics* 19(Suppl 1):292i–301i.
51. Zhang C, et al. (2006) A clustering property of highly-degenerate transcription factor binding sites in the mammalian genome. *Nucleic Acids Res* 34:2238–2246.
52. Sekinger EA, Moqtaderi Z, Struhl K (2005) Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol Cell* 735–748.
53. Panne D, Maniatis T, Harrison SC (2007) An atomic model of the interferon-beta enhanceosome. *Cell* 129:1111–1123.
54. Berg J, Willmann S, Lassig M (2004) Adaptive evolution of transcription factor binding sites. *BMC Evol Biol* 4:42.1–42.12.
55. Li G, Widom J (2004) Nucleosomes facilitate their own invasion. *Nat Struct Mol Biol* 11:763–769.
56. Sneppen K, Micheelsen MA, Dodd IB (2008) Ultrasensitive gene regulation by positive feedback loops in nucleosome modification. *Mol Syst Biol* 4:182.1–182.5.
57. Iyer V, Struhl K (1995) Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *EMBO J* 14:2570–2579.
58. Tillo D, Hughes TR (2009) $G + C$ content dominates intrinsic nucleosome occupancy. *BMC Bioinformatics* 10:442.1–442.13.
59. Kaplan N, et al. (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* 458:362–366.
60. Kay MS, Baldwin RL (1998) Alternative models for describing the acid unfolding of the apomyoglobin folding intermediate. *Biochemistry* 37:7859–7868.
61. Skoge ML, Endres RG, Wingreen NS (2006) Receptor-receptor coupling in bacterial chemotaxis: Evidence for strongly coupled clusters. *Biophys J* 90:4317–4326.
62. Changeux JP, Edelstein SJ (2005) Allosteric mechanisms of signal transduction. *Science* 308:1424–1428.
63. Maerkl SJ, Quake SR (2007) A systems approach to measuring the binding energy landscapes of transcription factors. *Science* 315:233–237.
64. Weinberg RL, Veprintsev DB, Bycroft M, Fersht AR (2005) Comparative binding of p53 to its promoter and DNA recognition elements. *J Mol Biol* 348:589–596.
65. Chen SC, Zhao T, Gordon GJ, Murphy RF (2007) Automated image analysis of protein localization in budding yeast. *Bioinformatics* 23(13):i66–71.
66. Sabo PJ, et al. (2006) Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods* 3:511–518.
67. Li G, Levitus M, Bustamante C, Widom J (2005) Rapid spontaneous accessibility of nucleosomal DNA. *Nat Struct Mol Biol* 12:46–53.

Supporting Information

Mirny10.1073/pnas.0913805107

SI Text

Derivation of the Protein and Nucleosomal Occupancies: The MWC Model. Here we use a statistical mechanics approach to derive occupancy and other equilibrium properties of the system. Alternative derivations can be found elsewhere (1). The advantage of our approach is that it allows direct generalization for sites of different strengths.

Consider a fragment of DNA containing n cognate sites for a DNA-binding protein, e.g., a transcription factor (TF), with concentration P and an affinity to each site characterized by the binding constant K. Because the sites are bound independently, the probability of each site being occupied is

$$y = \frac{P}{P+K} = \frac{\alpha}{1+\alpha}, \qquad \text{[S1]}$$

where $\alpha = P/K$ is a dimensionless protein concentration. It is easy to see that α is simply a statistical weight of the bound state.

The system of sites can be in two states, N and O, that determine the binding constants of all the sites: K_N and K_O. The statistical weights of the bound site in each state are: $\alpha_O = P/K_O \equiv \alpha$ and $\alpha_N = P/K_N = cP/K_O = \alpha c$, where $c = K_O/K_N$ is another dimensionless parameterof the system.

The system has $2n$ states: $N_0,N_1,\ldots N_n$, $O_0,O_1,\ldots O_n$ where the subscript denotes the number of occupied sites. The states N and O have different energies, and in the absence of occupied sites, the concentrations of the two states are connected by $L = N_0/O_0$.

Thus, the system is fully defined by three dimensionless parameters: α—the effective concentration of the protein, c—the suppression of the protein's affinity for its site by the nucleosome, and L—the equilibrium stability of the nucleosome in the absence of the DNA-binding protein. First, we calculate the equilibrium occupancy per site, i.e.,

$$Y = \frac{1}{n}\frac{\sum_{i=0}^{n} i[w(O_i)+w(N_i)]}{Z}, \qquad \text{[S2]}$$

with the partition function $Z = \sum_{i=0}^{n}[w(O_i)+w(N_i)]$. The function $w(\cdot)$ is a statistical weight of each state:

$$w(O_i) = C_n^i \alpha_O^i = C_n^i \alpha^i \qquad \text{[S3]}$$

$$w(N_i) = LC_n^i \alpha_N^i = LC_n^i (\alpha c)^i, \qquad \text{[S4]}$$

where C_n^k is the binomial coefficient and L takes care of the higher statistical weight of the state N.

Sums in the numerator and denominator can be easily calculated:

$$\sum_{i=0}^{n} w(O_i) = \sum_{i=0}^{n} C_n^i \alpha^i = (1+\alpha)^n, \qquad \text{[S5]}$$

$$\sum_{i=0}^{n} i w(O_i) = \sum_{i=0}^{n} C_n^i i\alpha^i = \alpha \sum_{i=0}^{n} C_n^i i\alpha^{i-1} \qquad \text{[S6]}$$

$$= \alpha\frac{\partial}{\partial\alpha}\sum_{i=0}^{n} C_n^i \alpha^i = \alpha\frac{\partial}{\partial\alpha}(1+\alpha)^n = \alpha n(1+\alpha)^{n-1}. \qquad \text{[S7]}$$

Other sums can be calculated in the same way to give a closed form solution:

$$Y = \alpha\frac{(1+\alpha)^{n-1} + Lc(1+c\alpha)^{n-1}}{(1+\alpha)^n + L(1+c\alpha)^n}. \qquad \text{[S8]}$$

Eq. **S8** is identical to the mean occupancy of hemoglobin sites obtained in the MWC model. As in the case of hemoglobin, cooperativity requires

$$L \gg 1, \qquad c \ll 1.$$

Occupancy curves do not change much with c but require a sufficiently large L to achieve high cooperativity and a large Hill coefficient (see Fig. S1 and below).

We can also calculate quantities that have no counterparts in the MWC model of hemoglobin, such as the nucleosomal occupancy:

$$Y_N = \frac{\sum_{i=0}^{n} w(N_i)}{Z} = \frac{L(1+c\alpha)^n}{(1+\alpha)^n + L(1+c\alpha)^n}, \qquad \text{[S9]}$$

and the probability of having exactly k sites occupied:

$$p_k = \frac{C_n^k \alpha^k (1+Lc^k)}{(1+\alpha)^n + L(1+c\alpha)^n}, \qquad \text{[S10]}$$

or at least k sites occupied:

$$P_k = \sum_{i=k}^{n} p_i = \frac{\sum_{i=k}^{n} C_n^i \alpha^i (1+Lc^i)}{(1+\alpha)^n + L(1+c\alpha)^n}. \qquad \text{[S11]}$$

These quantities are particularly useful for dealing with large clusters of sites, where a few bound proteins (transcription factors) can be sufficient to activate transcription. For example, the probability of having at least one site occupied in a cluster is

$$P_1 = \frac{(1+\alpha)^n - 1 + L[(1+\alpha c)^n - 1]}{(1+\alpha)^n + Lc(1+\alpha)^n} = 1 - \frac{1+L}{(1+\alpha)^n + Lc(1+\alpha)^n}.$$

The Hill Coefficient of the MWC Model. The Hill coefficient has been historically used to characterize the degree of cooperativity in a binding reaction. Although it cannot be obtained in a simple and closed form, here we derive an approximate but transparent expression for the Hill coefficient.The Hill coefficient h is defined for an all-or-none binding reaction of h independent molecules:

$$Y = \frac{[P]^h}{K_d^h + [P]^h} = \frac{\alpha^h}{1+\alpha^h}. \qquad \text{[S12]}$$

This expression is a poor approximation for MWC occupancy obtained above. Historically, the value of h was sought as a slope of $\frac{Y}{(1-Y)}$ vs α plotted in log-log scale. Using this approach, we define

$$h(\alpha) = \frac{\partial \log\left[\frac{Y}{1-Y}\right]}{\partial \log \alpha}, \quad \text{[S13]}$$

and then find

$$h = \max_{\alpha} h(\alpha). \quad \text{[S14]}$$

Using Eq. **S8** and assuming $c = 0$, we obtain

$$h(\alpha) = \frac{(1+\alpha)^n + L(1+n\alpha)}{(1+\alpha)^n + L(1+\alpha)}. \quad \text{[S15]}$$

While there is no analytical solution for

$$\frac{\partial h(\alpha^*)}{\partial \alpha} = 0, \quad \text{[S16]}$$

the equation simplifies to:

$$(1+\alpha^*)^n = L\frac{1+\alpha^*}{(n-1)\alpha^* - 1}, \quad \text{[S17]}$$

leading to:

$$h(\alpha^*) = \frac{n\alpha^*}{\alpha^* + 1}. \quad \text{[S18]}$$

The value of α^* can be obtained by solving Eq. **S17** numerically, or by approximating its right side with $L/(n-1)$, yielding

$$\alpha^* \approx \left(\frac{L}{n-1}\right)^{\frac{1}{n}} - 1, \quad \text{[S19]}$$

which provides a very good approximation for the Hill coefficient computed numerically (Table S3).

There are two important conclusions to be made from the form of Eq. **S18**. First, the Hill coefficient is linear in n, and is close to n. Thus, the more binding sites there are in the footprint, the more cooperative the binding reaction is. Second, the Hill coefficient depends, though weakly, on L, stability of the nucleosome. Nucleosome-mediated cooperativity, thus requires a nucleosome that is sufficiently stable in the absence of activated TFs poised to displace it.

Derivation of the Protein and Nucleosomal Occupancies for Distinct Sites: Generalization of the MWC Model. The model presented above can be easily generalized for cases when sites have different strengths, or when two or more types of proteins are poised to bind their respective sites in the region of interest. These cases lead to different statistical weights of different sites, i.e., α_i, $i = 1 \ldots n$. While there are no closed form solutions for such cases, we derived the following equations that can be treated numerically. The partition function is the sum of contributions by the two (N and O) states. Each state has n sites, each either occupied or unoccupied independent of the others. Thus, each site contributes $(1+\alpha_i)$ to the product in the corresponding state, leading to the partition function

$$Z = \prod_{i=1}^{n}(1+\alpha_i) + L\prod_{i=1}^{n}(1+c\alpha_i). \quad \text{[S20]}$$

The occupancy per site and the nucleosome occupancy can be computed as follows:

$$Y = \frac{1}{n}\frac{\sum_{k=1}^{n}\alpha_k\prod_{i=1,i\neq k}^{n}(1+\alpha_i) + Lc\sum_{k=1}^{n}\alpha_k\prod_{i=1,i\neq k}^{n}(1+c\alpha_i)}{Z}, \quad \text{[S21]}$$

$$Y_N = \frac{L\prod_{i=1}^{n}(1+c\alpha_i)}{Z}. \quad \text{[S22]}$$

While calculating P_k for any k becomes more cumbersome, P_1 and P_2 have simple forms:

$$P_1 = 1 - \frac{1+L}{Z}, \quad \text{[S23]}$$

$$P_2 = 1 - \frac{1+L+(1+Lc)\sum_{i=1}^{n}\alpha_i}{Z}. \quad \text{[S24]}$$

The Effect of Site Position. Interference of the histone core with protein binding depends on the location of sites with respect to the nucleosome. Sites located closer to the center of the nucleosome are affected more by the presence of the histone core than sites located closer to the ends. This effect can be taken into account by assigning different values of c_i to different sites: greater c_i values for central sites and lower for peripheral sites. Assuming otherwise identical sites, we derived the following expressions for the mean site occupancy and the nucleosomal occupancy:

$$Y = \frac{\alpha}{n}\cdot\frac{n(1+\alpha)^{n-1} + L\sum_{k=1}^{n}c_k\prod_{i=1,i\neq k}^{n}(1+c_i\alpha)}{Z}, \quad \text{[S25]}$$

$$Y_N = \frac{L\prod_{i=1}^{n}(1+c_i\alpha)}{Z}, \quad \text{[S26]}$$

$$Z = (1+\alpha)^n + L\prod_{i=1}^{n}(1+c_i\alpha). \quad \text{[S27]}$$

Derivation of the Equation for the Critical Number of Sites n_c in a Cluster. Here we study how the cluster occupancy depends on the number of sites n. We focus on the nucleosomal occupancy, looking to find n_c—a critical number of sites sufficient to have a nucleosome displaced by the DNA-binding proteins.

We seek n_c such that it provides nucleosomal occupancy $Y_N(n_c) = 0.5$. We consider a case where $c \ll 1$ and $\alpha \approx 2$–5 and thus can approximate Y_N:

$$Y_N = \frac{L(1+c\alpha)^n}{(1+\alpha)^n + L(1+c\alpha)^n} \approx \frac{L}{(1+\alpha)^n + L},$$

$$Y_N(n_c) = \frac{1}{2},$$

$$(1+\alpha)^{n_c} = L,$$

$$n_c = \frac{\log L}{\log(1+\alpha)}.$$

Using the range of values $L = 100$–$1{,}000$ and $\alpha \approx 2$–5, we obtain $n_c \approx 3$–6. Note that 3-6 sites per nucleosomal footprint corre-

sponds to about 15-30 sites in a regulatory region of 1 Kb being required to displace nucleosomes from this region.

A Generalized Model: DNA Peeling. The model presented above assumes two states for the nucleosomal DNA: fully open and accessible to TFs, or fully bound by the histones forming nucleosomes. DNA in the nucleosomal state is partially accessible to TFs through partial DNA unwrapping. This effect is taken into account implicitly by setting c to a small value, the degree of DNA accessibility in the N state.

Here we generalize our model to consider explicitly conformations of the nucleosome where the histone core is bound to DNA, but the DNA can be partially unwrapped (peeled off). The physics of DNA unwrapping has been extensively studied experimentally (2, 3) and recently reviewed in depth by Garcia et al. (4).

We replace a single N-state with a continuum of states characterized by two reaction coordinates of unwrapping x_1 and x_2 (Fig. S2). The free energy of the unwrapped state depends on the energy of the remaining DNA-histone interactions and the energy of DNA bending. According to Garcia et al. (4), this free energy can be well approximated by a model in which both energies vary linearly with the length of bound DNA, $(\ell - x)$:

$$F(x) = (\gamma_{\text{bend}} - \gamma_{\text{adh}})(\ell - x),$$

where $x = x_1 + x_2$ and $\ell = 147$ bp, with γ_{bend} and γ_{adh} as the contact energies per unit of length for DNA bending ($E_{\text{bend}} = \gamma_{\text{bend}}\ell$) and DNA-histone contacts ($E_{\text{contact}} = \gamma_{\text{adh}}\ell$) introduced in (4). This approximation allows us to have only one reaction coordinate of peeling, x.

For further analysis, instead of x, the length of the peeled DNA, we use k, the number of TFBSs located on the peeled DNA, as a reaction coordinate (Fig. S2). Assuming n TFBSs equally spaced along a nucleosomal DNA of length ℓ, the free energy of a conformation with k exposed sites is:

$$F(k) = \epsilon(n - k) + C,$$

where C is an arbitrary constant and ϵ is the peeling energy per unit of DNA length which contains a single TFBS,

$$\epsilon = (\gamma_{\text{bend}} - \gamma_{\text{adh}})\ell/n = (E_{\text{bend}} - E_{\text{contact}})/n. \quad \textbf{[S28]}$$

We replace a single N-state with an array of states each having $k = 0 \ldots n$ sites exposed. The partition function Z_N of the N-state thus becomes a sum of contributions from these states, $Z_N = \sum_{k=0}^{n} Z_N(k)$. The partition function of the state with k sites exposed can be written as

$$Z_N(k) = (1 + \alpha)^k (1 + c\alpha)^{n-k} L(k) w(k), \quad \textbf{[S29]}$$

where $L(k)$ is the configurational equilibrium constant of the partially unwrapped state: $L(k) \sim \exp(F(k))$; and $w(k)$ is the number of possible ways to peel DNA such that k states are accessible. Because the histone core can form a single contiguous contact region with DNA, $w(k) \approx k + 1$ (see Fig. S2). We set $L(0) = L$, where L is the equilibrium constant of the fully nucleosomed state used above. The equilibrium constant of the open (O) state remains 1 as above. Then the equilibrium constants of the partially unwrapped states are given by

$$L(k) = L \exp(-k\epsilon), \quad \textbf{[S30]}$$

shown in Fig. S2*C*.

Note that the partially unwrapped state with $k = n$ (all sites accessible) is distinct from the open O-state considered above: The nucleosomal histone core is attached to the DNA in the $k = n$ state and is free in the solvent in the O-state. The free energy difference between the states is due to the entropy gain from dissociation of the histone core, which later further dissociates into individual histone dimers. This is reflected in the difference between the equilibrium constants of these states: partially unwrapped with $k = n$ has $L(n) = L\exp(-n\epsilon)$, while the open state has an equilibrium constant of 1. The significant entropy gain upon dissociation of the histone core requires that $L(n) \ll 1$, i.e.

$$L\exp(-n\epsilon) \ll 1. \quad \textbf{[S31]}$$

The final expression for the partition function of the system includes contributions from the N and O states:

$$Z = Z_N + Z_O, \quad \textbf{[S32]}$$

$$Z_N = L\sum_{k=0}^{n}(1 + \alpha)^k (1 + c\alpha)^{n-k} \exp(-k\epsilon)(k + 1) \quad \textbf{[S33]}$$

$$= L(1 + c\alpha)^n \quad \textbf{[S34]}$$

$$+L\sum_{k=1}^{n}(1 + \alpha)^k (1 + c\alpha)^{n-k} \exp(-k\epsilon)(k + 1), \quad \textbf{[S35]}$$

$$Z_O = (1 + \alpha)^n \quad \textbf{[S36]}$$

The partition function of the N-states naturally splits into the term (**S34**) for the fully nucleosomed state as considered in the two-state model above, and the term (**S35**), which corresponds to the partially unwrapped states:

$$Z = Z_N + Z_O, \quad \textbf{[S37]}$$

$$Z_N = Z_{\text{nucleosomed}} + Z_{\text{unwrapped}}, \quad \textbf{[S38]}$$

$$Z_O = (1 + \alpha)^n, \quad \textbf{[S39]}$$

$$Z_{\text{nucleosomed}} = L(1 + c\alpha)^n, \quad \textbf{[S40]}$$

$$Z_{\text{unwrapped}} = L\sum_{k=1}^{n}(1 + \alpha)^k (1 + c\alpha)^{n-k} \exp(-k\epsilon)(k + 1) \quad \textbf{[S41]}$$

$$= L(1 + c\alpha)^n \sum_{k=1}^{n}(k + 1) r^k, \quad \textbf{[S42]}$$

where

$$r = \left(\frac{1 + \alpha}{1 + c\alpha}\right)\exp(-\epsilon). \quad \textbf{[S43]}$$

The relative contribution of each of the states (nucleosomed, unwrapped and open) to the total partition function determines its relevance for the system. The contribution of $Z_{\text{unwrapped}}$ is determined by the value of ϵ, which is subject to the constraint (Eq. **S31**). Assuming, as above, that $L \gg 1$ and $c \ll 1$, it can be shown that for $\alpha \ll 1$, it is the fully nucleosomed state that dominates the system, $Z_{\text{nucleosomed}} \gg Z_{\text{unwrapped}}, Z_O$, i.e., at low TF concentration the system is largely in the fully nucleosomed state. For $\alpha \gg 1$, $Z_O \gg Z_{\text{nucleosomed}}, Z_{\text{unwrapped}}$, leading to a DNA population that is largely in the open state at high TF concentration. The contribution of the partially unwrapped states at intermediate values of α can be studied numerically for a range of ϵ values.

Fig. S3 presents the values of $Z_{\text{nucleosomed}}$, $Z_{\text{unwrapped}}$ and Z_O as a function of α for $\epsilon = 3, 6, 9k_BT$. The value of ϵ can be estimated from experimentally measured equilibrium constants for partially unwrapped states. Garica et al (4) calculated the values of γ_{adh} and γ_{bend} as well as the energy contributions of DNA bending and DNA-histone binding as $E_{\text{bend}} = 35k_BT$ and $E_{\text{contact}} = 14{\cdot}6k_BT = 84k_BT$, respectively. Using definition of epsilon in Eq. **S28** and $n = 5$–8, we obtain $\epsilon = 6 - 10k_BT$.

Clearly at small α, the nucleosomed state has the partition function of highest value, while at large α the largest contribution comes from the open state. Even at intermediate values, the contribution of the partially unwrapped states is much smaller than contributions of the two other major states considered above.

Conclusions. Analysis of this generalized model brings us to several conclusions. First, nucleosomes can be found largely in either the fully assembled conformation or absent from DNA. Second, while partial unwrapping provides TFs access to nucleosomal DNA, partially unwrapped states play little role in the equilibrium of the system. This two-state behavior of nucleosomes is due to the high affinity of histones for DNA and significant entropy loss upon histone assembly and DNA binding. Third, the two-state model introduced above adequately describes interactions between nucleosomes and other DNA-binding proteins.

1. Voet D, Voet JG (2004) *Biochemistry* (Wiley, New York), 3rd Ed.
2. Polach KJ, Widom J (1995) Mechanism of protein access to specic DNA sequences in chromatin: A dynamic equilibrium model for gene regulation. *J Mol Biol* 254:130–49.
3. Li G, Levitus M, Bustamante C, Widom J (2005) Rapid spontaneous accessibility of nucleosomal DNA. *Nat Struct Mol Biol* 12:46–53.
4. Garcia HG, et al. (2007) Biological consequences of tightly bent DNA: The other life of a macromolecular celebrity. *Biopolymers* 85:115–30.

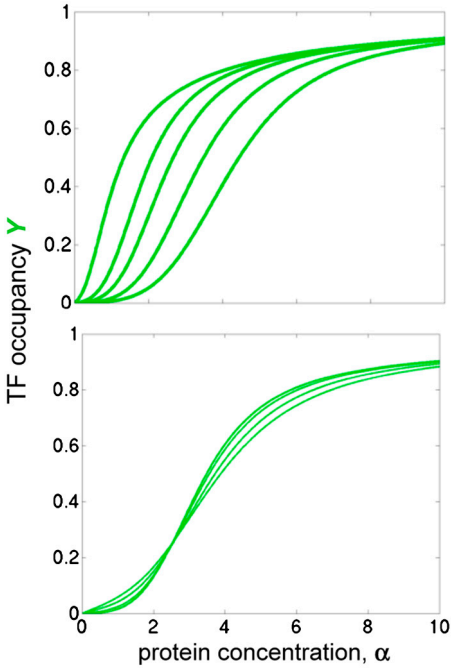

Fig. S1. Robustness of the cooperativity to 300-fold variation in parameters *L* (top, $L = 10$–3000, lines from left to right) and *c* (bottom, $c = 0.03 - 10^{-4}$, more sigmoidal for smaller *c*).

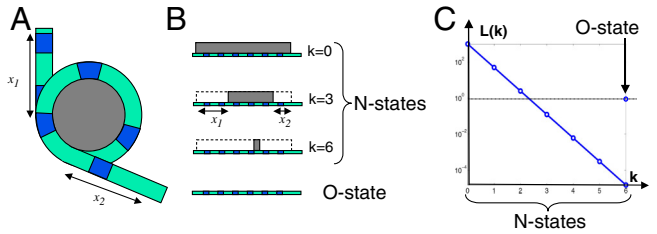

Fig. S2. (*A*) The partially unwrapped state of a nucleosome, with DNA peeling off at both ends and exposing some DNA-binding sites. (*B*) Schematic representation of states in the system: formerly a single *N*-state is now split into $n + 1$ *N*-states, each having a different number $k = 0 \ldots n$ of TFBSs exposed. Each of the new states has some degeneracy, which is $\sim k$, because there are approximately k ways for k sites to be exposed on two peeled arms of nucleosomal DNA (see Eq. **S29**). (*C*) The equilibrium constant of the partially unwrapped *N* states and the *O* state as given by Eq. **S30** ($n = 6$, $L = 1{,}000$, $\epsilon = 3k_BT$). Note that a state with all $k = n$ sites exposed is different from the *O* state: The histone core is still attached to DNA in the former but not in the latter. The free energies of the two states differ by the free energy of histone core assembly and DNA binding.

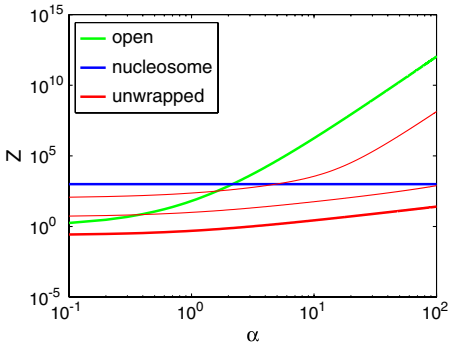

Fig. S3. Partition sums of the three main states of the system: open, fully nucleosomed, and partially unwrapped, each as a function of TF concentration α. The partition function for the unwrapped states has been calculated using $\epsilon = 3$, 6, and $9k_BT$, shown by three red lines from top down correspondingly. The sums have been computed using Eqs. **S38**, **S39**, **S40**, **S41**, and **S42** with values $n = 6$, $L = 1{,}000$, $c = 0$. At low TF concentration, the nucleosomed state has the greatest statistical weight, and at high TF concentration, the open state has the greatest weight. The contribution of the partially unwrapped state is nonnegligible at $\epsilon = 3k_BT$ but is always overshadowed by the two dominating states.

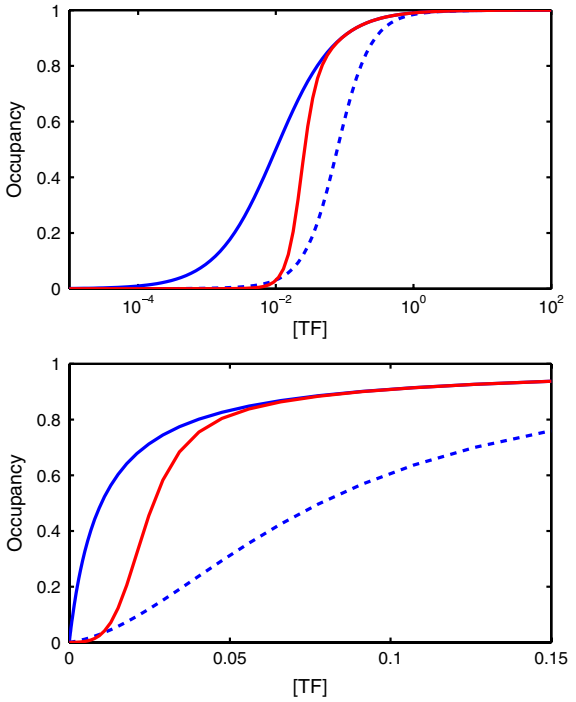

Fig. S4. Three saturation curves as functions of protein concentration for $K_d = 0.01$. Blue, solid—noncooperative binding (Eq. **S1**). Blue, dotted—cooperative binding of TFs to two sites considered by ref. 1) ($n = 2$, $L = 60$ as in ref. 1). Red—cooperative binding considered here ($n = 6$, $L = 10^3$). (*A*) Log scale of the X-axis as used in (1). (*B*) Linear scale of the X-axis as used here.

1 Raveh-Sadka T, Levo M, Segal E (2009) Incorporating nucleosomes into thermodynamic models of transcription regulation. *Genome Res* 19:1480–96.

Table S1. Comparison of the nucleosome-mediated cooperativity and the cooperative transition in hemoglobin

Cooperative transition in hemoglobin, the Monod–Wyman–Changeux model	Nucleosome-mediated cooperativity of transcription factors
Components of the system	
Oxygen pressure, pO_2	Concentration of TF, [*P*]
Two states of the hemoglobin monomer: high-affinity *R* and low-affinity *T* states	Two states of DNA: high-affinity *O* and low-affinity *N* (open or nucleosome) states
Prevalence of the *T* state at low pO_2 ($L \gg 1$)	Prevalence of the *N* state (stable nucleosome) at low TF concentration ($L \gg 1$)
Four oxygen-binding hemes ($n = 4$)	*n* TF binding sites
Allosteric transition in hemoglobin	Nucleosome assembly and displacement
Phenomena	
Cooperative binding of oxygen	Cooperative binding of TFs
The Bohr effect	Concentration of histones and their affinity for DNA
Other heterotropic regulation	Histone modifications, sequence-dependent nucleosome stability, histone-binding proteins
Homotropic regulation	TF-dependent nucleosome depletion, interaction between different TFs
Energy stored in protein/heme deformation	Energy stored in DNA deformation and histone-DNA interactions

Table S2. Comparison of the model's predictions with experiments

Model of the nucleosome-mediated cooperativity	Experimental data
Cooperative binding by noninteracting TFs	Gal4, USF, and NF-kappa B bind cooperatively to reconstituted nucleosomes in vitro.
Cooperative action of TFs that bind within a footprint of a nucleosome <150 bps, independent of site orientation.	Cooperativity of Gal4 and USF independent of site orientation (17). LexA and Gal4, with one TF lacking the activation domain, cooperate up to a range of 200 bps.
Lack of cooperative binding in the absence of the nucleosomes.	NF-kappa B acts synergistically at a promoter containing four sites. Binding is not cooperative in vitro (6)
Cooperative binding of TFs does not require direct interactions among them.	Structure of interferon-β enhanceosome demonstrating very few direct contacts among bound TFs (53).
Displacement of a nucleosome that occupies a TFBS-rich region	Enrichment of TFBSs in nucleosome- depleted regions (37, 69), nucleosome depletion in yeast regulatory regions (promoters) (33, 34).
Critical number ($n_c = 4$–6) of TF binding sites required for cooperative binding.	Clustering of TF binding sites in Drosophila enhancers, exceeding 20 sites per kb (2, 3). Importance of clustering in eve2 (9).
High concentration of TFs is required to displace a stable nucleosome, lower for modified nucleosomes (the Bohr effect).	Recruitment of chromatin remodeling is required for activation through low-affinity sites. Overexpression compensates for the lack of remodeling (70, 71). Overexpression of TFs compensates for mutation of a high-affinity site, leading to nucleosome eviction through binding of two low-affinity sites (72).
Nucleosomal occupancy depends on the presence of TFBSs.	Nucleosomal occupancy is restored by mutations eliminating TFBSs (69). Mutation in the high-affinity site of HSF reduces nucleosome eviction in vivo (72).

Table S3. The Hill coefficient as a function of the number of sites *n*, for $L = 10^3$, $c = 0$

n	*h*, numeric	*h*, Eq (18)
1	1.0	1.0
2	1.9	1.9
3	2.7	2.6
4	3.1	3.1
5	3.5	3.5
6	3.8	3.7
7	4.0	3.9
8	4.1	4.0