Classification: BIOLOGICAL SCIENCES: Biophysics

Nucleosome-mediated cooperativity between transcription factors

Leonid A. Mirny

Harvard-MIT Division of Health Sciences and Technology, and Dept. of Physics, Massachusetts Institute of Technology, Cambridge, MA
E-mail: leonid@mit.edu Tel: ++1-617-452-4862

Character count: 36,992 characters with spaces, 2 Tables, 3 Figures
Abbreviations: MWC: Monod-Wyman-Changeux

Abstract

Cooperative binding of transcription factors (TFs) to promoter 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 TFs 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 the involved TFs. Such flexibility is inconsistent with cooperativity by direct interactions between TFs. Here we demonstrate that strong cooperativity among non-interacting 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 of hemoglobin. This surprising analogy provides deep insights, with parallels between heterotropic regulation of hemoglobin (e.g. Bohr effect) and roles of nucleosome-positioning sequences and chromatin modifications in gene expression. Nucleosome-mediated cooperativity is consistent with several experimental studies, allows substantial evolutionary flexibility in and modularity of regulatory regions, and provides a rationale for a broad range of genomic and evolutionary observations. Striking parallel between cooperativity in hemoglobin and in transcription regulation points to a general mechanism that may be used in various biological systems.

Introduction

In higher eukaryotes, cis-regulatory regions (CRRs) are 200-1000 bps long and may contain clusters of 10-30 TF binding sites (TFBSs) (1-3). The arrangement, identity and affinity of the sites determine the function of the CRR. Cooperative binding of TFs to CRRs leads to highly cooperative gene activation and is essential for development (4) and other vital processes (5).

Cooperative binding is traditionally explained by protein-protein interactions among participating TFs (6, 7). While this mechanism finds support in bacterial and eukaryotic systems (8), several functional and genomic observations are inconsistent with it. The mechanism of cooperativity due to protein-protein interactions (directly or via DNA looping (6, 7, 9)) should significantly constrain arrangements of TFBS, allowing only those that provide the correct orientation, order and distance between TFs. On the contrary, recent evolutionary analysis of fly enhancers revealed massive rearrangements of TFBS (10) (11), and is further supported by experimental studies, which demonstrated that CRRs could tolerate incorporation of new binding sites (promiscuity) and significant alterations in TFBS placement while retaining *in vivo* functionality (10) (12, 13). A few mechanisms proposed to explain flexible arrangements of TFBS and promiscuity are based on the idea of transcriptional synergy, i.e. cooperative recognition or simultaneously touching of TFs by some flexible part of the transcription machinery (12, 13), rather than on cooperative binding of TFs to DNA.

An alternative mechanism of cooperativity considered in this study is based on synergistic binding of non-interacting TFs, mediated by a nucleosome. The phenomenon of nucleosome-mediated cooperativity has been documented by a series of *in vivo* and *in vitro* experiments (5, 14-16), which demonstrated synergistic binding and gene activation by non-endogenous 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 between TFs (17). Experimental studies (15) and a model (18) of synergistic binding to nucleosomal DNA considered two close-by (~30bps) sites, located at the same arm of the nucleosome (**Fig 1B**) and interacting through *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 (18).

We suggest that nucleosome-mediated cooperativity can proceed through a different and much more general mechanism, and in contrast to *assisted unwrapping* can (i) involves a cluster of several sites, (ii) can lead to nucleosome eviction, and (iii) cooperative binding among sites separated by up to 150bps. Proposed mechanism explain a broad range of phenomena that can not be explained by assisted unwrapping that works only for close-by sites and as long as the DNA remains bound to a histone core, and thus is hardly applicable to CRRs that were shown to be mostly nucleosome-free in living cells (19). Although chromatin modification plays an important role in formation and maintenance of nucleosome-free regions, binding of TFs to largely nucleosomal DNA precedes the modification and is central to our mechanism.

Here we present a mechanism and quantitative model of nucleosome-mediated cooperativity. We integrate binding of TFs to nucleosomal as well as naked DNA and include possible nucleosome displacement/eviction and emphasize the role of TFs in establishing nucleosome-free regulatory regions. We demonstrate that competition for a region of DNA that bears an array of TFBSs between histone proteins and TFs can induce strongly cooperative binding among the TFs. Strikingly, we find that the model of nucleosome-induced cooperativity is identical to the Monod-Wyman-Chaneux model of cooperativity in hemoglobin (20). Using this analogy we gain deeper insights into a range of phenomena such as nucleosome positioning sequences and nucleosome modifications, the role of low-affinity TFBSs and TFBS clustering, and the origin of DNase hypersensitive sites (see **Table 1**). We demonstrate that the mechanism of nucleosome-mediated cooperativity is sufficiently general and, in contrast to assisted unwrapping, does not rely on specific details of DNA unwinding and bending (18). Next we show how the presented mechanism allows significant flexibility of TFBS arrangements, easy substitution of new TFs, and CRR modularity. Finally we present a collection of experimental evidence in support of our nucleosome-mediated mechanism.

Results

The model of nucleosome-mediated cooperativity

We consider interactions of TFs with a stable nucleosome (Fig. 1B), containing an array of *n* TFBSs located within its footprint (147bps) (**Fig.1**). This region of DNA can be in one of two states: the nucleosome (*N*) state and the open (*O*) state, in which the 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 back at very high rate, thus making nucleosomal DNA at least partially accessible to TFs (21, 22). TFBSs can be occupied by TFs in either state, with N_i and O_i denoting which state the DNA is in, and with *i* occupied sites. The affinity of TFs for the sites depends on the state with binding constants K_N and K_O, respectively, and higher affinity in the open state: $K_O/K_N<<1$. For simplicity of the presentation we assume all TFBSs to have the same affinity and experience the same suppression, *c,* defined as K_O/K_N, by the nucleosome. Removing these assumptions leads to more complicated equations, presented in the Supplement. We consider equilibrium, assuming that rapid exchange of TFs and histones lead to fast equilibration of the system.

The equilibrium between the *N* and *O* states in the absence of bound TFs is characterized by $L=[N_0]/[O_0]$, with $L>>1$ for a stable nucleosome. In equilibrium, the system is fully defined by the three dimensionless parameters *c, L,* and the aggregate concentration of TFs $\alpha=[TF]/K_O$ **(Fig 1C).**

Estimation of parameters from experimental data

To estimate α we take into account both specific, (K_d) and non-specific (K_d^{NS}) affinity of a TF to DNA and obtain $\alpha = [P]/K_d^{eff} = [P]K_d^{NS}/[DNA]K_d$, where [*DNA*] is the concentration of TF-accessible non-specific DNA. The dissociation constant of most eukaryotic TFs is in the range of $K_D \approx 1-10$ nM, while available non-specific binding constants are about 1000 fold

greater $K_D \approx 1-10\mu\mathrm{M}$ (23, 24) }. Using length of genomic DNA, assuming 90% chromatization and measured copy-number of TFs ($[P]\approx 500\text{-}5000$ in yeast and $[P]\approx 10^4\text{-}10^5$ protein copies per nucleus in multicellular eukaryotes, (25, 26)) we obtain the range of $\alpha \approx 0.5\text{-}5$.

We estimate *L* using *in vivo* nucleosome equilibrium occupancy $f = [N]/([N]+[O])$, $L = f/(1-f)$. Although for stable nucleosomes occupancy is very close to 1 and hard to measure, the fraction of DNA covered by nucleosomes provides a lower bound for f and has a range of $0.9\text{-}0.99$ (27), yielding $L > 10\text{-}100$. This is a lower bound for f and L, because most of 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. Acting through steric hindrance, such suppression is not permanent and nucleosomal DNA get exposed for TF binding (**Fig 1B**). Widom et al (28, 29) quantitatively characterized this mechanism *in vitro*, demonstrating that such exposure of nucleosomal DNA does not require nucleosome disassembly, i.e. acting in the *N* state of our model. Thus suppression of binding in the *N* state, parameter *c*, is equivalent to the experimentally measured equilibrium constant of site exposure. This constant depends on the location of the site with respect to the center of the nucleosome and has the range $c = 2\cdot 10^{-2}\text{-}10^{-5}$. Although imprecise these estimates dare sufficient to provide cooperativity. The mechanism is very robust, requiring $L \gg 1$ and $c \ll 1$ for the onset of cooperativity (**Fig 2B**).

Analogy to hemoglobin

Strikingly, the system of TFs and a nucleosome is identical to the scheme of cooperativity in hemoglobin as described by the classical Monod-Wyman-Changeux (MWC) model (20). **Table 1** summarizes equivalent parameters and analogous phenomena between the two systems. The MWC model considers the equilibrium between two states of the hemoglobin tetramer: the R state, which has a higher affinity for O_2, and the low-affinity T state. In the absence of the oxygen, the 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 1D**). The R and T states of hemoglobin correspond to O- and N-states of the nucleosome, and the oxygen binding to individual hemoglobin domains corresponds to TFs binding individual sites (**Fig 1CD, Table 1**). This analogy helps us to gain deep understanding of cooperativity and regulation in the TF-nucleosome system.

The two quantities of primary biological interest are the occupancy per TFBS of the CRR by the TFs (*Y*) and the occupancy by the nucleosome (Y_N). By analogy to MWC, and using the tools of statistical mechanics we obtain expressions for these quantities:

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

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

Figure 2A presents equilibrium TF and nucleosome occupancies Y and Y_N as a function of TF concentration, computed using experimentally determined parameters (see Supplement for parameter estimation). Strikingly, nucleosome-mediated cooperativity can provide a sharp transition with a two-fold increase in TF concentration leading to a more than eight-fold increase in the occupancy. The nucleosome occupancy also changes cooperatively, dropping from about 65% to less than 10% due to a two-fold change in TF concentration.

Analogy to the MWC model allows us to reveal features of the nucleosome-TF system that are essential for cooperativity (**Table 1**). The MWC system has a strong cooperativity as long as L is sufficiently large ($L>10\text{-}100$) and c is sufficiently small ($c<0.1$). In our system (**Fig.2B**), this corresponds to requirements for nucleosome stability (i.e. forming a nucleosome more than 90% of the time, in the absence of TFs) and sufficient attenuation of TF binding to nucleosomal rather than naked DNA (at least ten-fold in K_d). These requirements are consistent with high stability (19, 30) and slow exchange (31, 32) of nucleosomes in regions depleted in TFBSs and unaffected by polymerase passage (e.g. inactive ORFs). *In vitro* studies of TF binding to nucleosomal DNA demonstrate the required attenuation of TF binding. **Figure 2B** shows that the cooperative transition is robust to variation in L and c, becoming, as expected from the MWC model, sharper at larger L and smaller c.

We also exploit the analogy to MWC to examine the implications of sequence-specific nucleosome positioning, histone modifications and other processes involved in gene regulation. In fact, 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 in hemoglobin is the Bohr effect: lowering the pH causes the oxygen affinity to decrease, thus providing more oxygen to actively working muscles. The basis of the Bohr effect is in 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 to one of the states of hemoglobin affects the R-T equilibrium and thus affinity of the hemoglobin to the oxygen.

Bohr effect and chromatin modification

Several processes in gene regulation are counterparts on the Bohr effect in hemoglobin: Histone modifications and histone-binding proteins affect nucleosome stability, thus altering the N-O equilibrium. To study this effect we used equations (1) and (2) to examine how the TF and nucleosomal occupancy curves change in response to reduced nucleosome stability L. **Figure 2C** shows manifestation of the "Bohr effect" in 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 ΔG=1 kcal/mol ($\Delta G=k_BT\log(L/L')$) can lead to an 80% drop in nucleosome occupancy and a concurrent rise of the TF occupancy (**Fig. 2C, inset**). Nucleosome-positioning sequences have a similar effect: they alter nucleosome stability thus shifting the occupancy curve (**Fig 2C**). Importantly, cooperativity in nucleosome binding induced by competition with TFs leads to amplification of the sequence signal in CRRs, i.e. small

changes in histone affinity translate into significant changes in nucleosome occupancy (**Fig. 2C, inset**). This effect helps to explain how weak nucleosome positioning sequence signals can lead to significant differences in nucleosome occupancy of different regions and sharp borders between the regions along the genome (31, 33).

Heterotropic regulation can also work in the opposite way: factors influencing binding of TFs can impact chromatin structure in TFBS-rich regions. For example, activation of a tissue-specific TF can lead to selective reduced chromatization and increased accessibility of tissue-specific CRRs (see **Fig.3**). This result is in agreement with a recent genome-scale mapping of DNaseI hypersensitive sites (DHS) that reported a highly tissue-specific DHS profile (34). As discussed above, cooperativity of this allosteric regulation allows small changes in the concentrations of active TFs to significantly reduce nucleosomal occupancy (**Fig 2A and 3C**). Note that this mechanism of nucleosome displacement by competition with activated TFs does not rely on recruitment of chromatin modification machinery, which may play a role in further destabilizing nucleosomes (35) and expanding nucleosome-free regions initially formed by the competition. This effect will be considered in detail elsewhere.

Critical size of the TFBS cluster

Nucleosome-induced cooperativity, however, has some properties that do not have counterparts in hemoglobin. For example, the number and the strength of biding sites is constant in hemoglobin, but vary in CRRs. **Figure 2D** presents nucleosomal occupancy as a function on the number of TFBS calculated using equations (1) and (2). As the number of TFBS 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 TFBS 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 Supplement). **Figure 3** demonstrates that clusters of 5 high-affinity and 8 low-affinity TFBS become occupied and nucleosome-free, while isolated sites remain TF-free. Having the number of sites in this range per nucleosomal footprint is both sufficient for cooperativity and consistent with the recent genomic characterization of *Drosophila* CRRs (1, 3) that contain several clusters of 4-6 sites in a region of 100-150 bps (2)**.** This effect can also explain widespread depletion and rapid exchange of nucleosomes in TFBS-rich CRRs (30, 31, 36-38).

Our approach allows generalization of MWC cooperativity to consider contributions of low-affinity sites and mixtures of sites of different TFs etc (see Supplement for equations and derivations). For example, assuming that only a few TFs are sufficient for activation, as was demonstrated experimentally (39), we obtain the expression for the probability of having at least k=2 of n sites being occupied, P_k (**Fig 3**). We also considered the contribution of low-affinity sites that were shown to play an important role in fly enhancers (8, 40). **Figure 3** illustrates how arrays of low-affinity and high-affinity sites in nucleosomal DNA respond differently to increasing level of TFs. Our formalism allows one to calculate experimentally observable nucleosomal and TF occupancy of a DNA region with the number and strength of sites that can be altered experimentally.

Discussion

Below we discuss several experimental results that support nucleosome-mediated mechanism of cooperativity, summarized in **Table 2**, and propose direct experimental tests of the mechanism. We believe that proposed mechanism contributes to the complex interplay of nucleosome-positioning DNA sequences (41-44) (45) and TFs interacting with each other (2) and with the histones (46). Such mechanisms have been actively studied and certainly play important role in gene regulation (40, 47). The relative importance of different mechanisms of cooperativity (15) and nucleosome positioning (44) may vary for different DNA regions and different organisms (45). A recent study that examined the role of intrinsic and extrinsic factors in nucleosome positioning has demonstrated that taking into account TF-nucleosome competition improves the accuracy of predicted nucleosome positions (44).

Most direct evidence of nucleosome-mediated cooperativity comes from experimental studies that have demonstrated cooperative binding of TFs without involvement of direct protein-protein interactions for a range of up to 200 bps (15). Moreover, experiments with TFs lacking activation domains have shown that synergetic activation of gene expression is determined more by the number of TFBSs than by the interactions with general TFs, polymerase, or chromatin modification machinery (39, 48). Consistent with the nucleosome-induced mechanism, trans-complementation experiments on stripe 2 enhancers showed that precise expression does not require special Bcd-Hb interactions and can be achieved by chimeric and non-endogenous (i.e. non-interacting) TFs (8). The range (~150-200 bps) of nucleosome-induced cooperativity is also consistent with the reported modularity of the *otd* enhancer, which contains two 180 bp TFBS clusters, each able to provide the correct expression pattern (49).

The presented mechanism also helps to tie together several observations in functional and comparative genomics. Cooperative nucleosome displacement can explain how low nucleosomal density is maintained on CRRs and promoters and how sharp boundaries of such nucleosome-depleted regions are achieved. The critical number of sites n_c, calculated above, required for the TF-induced nucleosome displacement is consistent with clustered arrangements of TFBSs and helps to explain why such clustering serves as a powerful criterion for bioinformatic identification of CRRs (50). The nucleosome-mediated mechanism suggests a role for low-affinity TFBS, which in fly enhancers are abundant (2, 51) and essential (8) in assisting high-affinity sites to displace nucleosomes and provide cooperative binding. This mechanism serves, along with sequence information (43, 52) (41, 42, 45), in providing nucleosome positioning, but, in contrast, can be tissue or condition specific, as recently reported (34).

The nucleosome-mediated mechanism provides significant evolutionary flexibility to CRRs, allowing considerable rearrangements of TFBSs while retaining cooperativity. Such widespread flexibility and rapid evolution of CRRs has been reported (1-3) and was hard to reconcile with the classical model of cooperativity by direct protein-protein interactions. Moreover, the described mechanism explains the observed promiscuity of CRRs: it allows unrelated TFs to cooperate in gene regulation, simply by evolving TFBSs close to each other in CRR. Similarly, TFs can become a part of an existing assembly by acquiring TFBSs within an existing cluster, a fairly fast and widespread evolutionary process (11, 53). A classical model of cooperativity would also require interacting TFs to evolve

interfaces that facilitate protein-protein interactions – a process that requires many more mutations.

Competition with a single nucleosome can provide cooperative binding for TFBSs separated by at most 150-200bps. This range can be further increased by the demonstrated synergy of nearby nucleosomes (54) and can spread much further through recruitment of chromatin modification machinery and due to positive feedback in this process (55).

Proposed mechanism makes concrete predictions that can be tested experimentally. First, it suggests the dependence of the cooperative effect and nucleosome occupancy on the number of TFBS. Both the nucleosome occupancy and gene expression can be assayed directly for synthetic or natural promoters contain TFBS clusters of different density and spacing and tested to be SWI/SNF-independent. Second, the cooperativity of binding is binding is predicted to depend on nucleosome stability in a non-trivial fashion: stabilization of the nucleosome (e.g. by strong nucleosome-positioning DNA sequence) is expected to make binding and expression *more cooperative*, while destabilization of the nucleosome (through the DNA sequence or nucleosome eviction by a non-indigenous TF) is expected to *decrease cooperativity*. This prediction can help to distinguish proposed mechanism from the effect of direct protein-protein interactions that are expected to exhibit the opposite dependence on nucleosome stability.

In summary, we have shown how competition between a nucleosome and TFs can lead to cooperative binding of TFs and cooperative displacement of nucleosomes from regulatory regions. We have established and employed the analogy between this process and cooperativity in hemoglobin according to 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, the presented mechanism explains a wealth of genomic and evolutionary observations that cannot be reconciled with the classical model of cooperativity among TFs. Our study provides strong support to the view that regulatory regions are flexible and highly evolvable regions of the genome. Finally, the analogy between cooperativity in hemoglobin and nucleosome-mediated cooperativity of TFs hints at the possible universality of the MWC mechanism of cooperativity in seemingly unrelated biological processes.

Materials and Methods

Derivation of the TF and nucleosomal occupancy: 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 (56). The advantage of our approach is that it allows direct generalization for sites of different strength.

Consider *n* sites, a protein having a concentration [*P*], and an affinity to the site characterized by the binding constant K. Since the sites are bound independently, the probability of each site being occupied is

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

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 cluster of sites located in a footprint of a single nucleosome 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: $\alpha_O = [P]/K_O = \alpha$ and $\alpha_O = [P]/K_N = c[P]/K_O = c\alpha$, where $c = K_O/K_N$ is the second dimensionless parameter of the system. Estimation of these parameters from experimental data is presented in the Supplement.

The system has $2n$ states: N_0, N_1, ..., N_n, O_0, O_1, ..., 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 related by the equilibrium constant $L = [N_0]/[O_0]$.

The system is thus fully defined by three dimensionless parameters: α, c, and L. First we calculate the equilibrium occupancy *per site*, which is

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

where Z is the partition function and $w(\cdot)$ is a statistical weight of each state:

$$w(O_i) = C_n^i \alpha_O^i = C_n^i \alpha^i$$
$$w(N_i) = LC_n^i \alpha_N^i = LC_n^i (\alpha c)^i$$

where C_n^k is the binomial coefficient and parameter L takes care of the higher statistical weight of the *N* state. Sums in the numerator and denominator can be easily calculated giving a closed-form solution

$$Y = \alpha \frac{(1+\alpha)^{n-1} + Lc(1+c\alpha)^{n-1}}{(1+\alpha)^n + L(1+c\alpha)^n} \tag{3}$$

Note that Eq (3) is identical to the mean occupancy of hemoglobin sites obtained in the MWC model.

We also calculate quantities that were not obtained in MWC, such as the nucleosomal occupancy as

$$Y_N = \frac{\sum_{i=1}^{n} w(N_i)}{Z} = \frac{L(1+c\alpha)^n}{(1+\alpha)^n + L(1+c\alpha)^n} \tag{4}$$

and the probability to have exactly *k* sites occupied

$$p_k = \frac{kC_n^k \alpha^k (1+Lc^k)}{(1+\alpha)^n + L(1+c\alpha)^n} \tag{5}$$

or *at least k* sites to be occupied:

$$P_k = \sum_{i=k}^{n} p_i = \frac{\sum_{i=k}^{n} C_n^i \alpha^i (1+Lc^i)}{(1+\alpha)^n + L(1+c\alpha)^n} \tag{6}$$

These quantities are particularly useful for large clusters of sites in eukaryotic enhancers, where a few bound activators can be sufficient to activate transcription. For example, the probability of having at least one site occupied in a cluster has a simple form**:**

$$P_1 = \frac{(1+\alpha)^n - 1 + L\left[(1+\alpha c)^n - 1\right]}{(1+\alpha)^n + Lc(1+\alpha)^n} = 1 - \frac{1+L}{(1+\alpha)^n + Lc(1+\alpha)^n} \quad (7)$$

Equations for the combinatorial effect of several distinct TFs binding to their corresponding sites are provided in the Supplement.

Acknowledgments

I am grateful to Mehran Kardar, Stanislav Shvartsman, Paul Wiggins, Jane Kondev, Ned Wingreen, and particularly to George Benedek for insightful discussions of this study, to John Stamatoyannopoulos and Ilya Ioschikhes for carefully reading and commenting the manuscript, and to Jason Leith and Zeba Wunderlich for discussing and editing the manuscript. Part of this work took place at the Kavli Institute for Theoretical Physics, Santa Barbara, California, and the Aspen Center for Physics, Aspen, Colorado. I was supported by the NIH-funded National Center for Biomedical Computing, *i2b2,* Informatics for Integrating Biology and the Bedside.

References

1. Schroeder, M. D., Pearce, M., Fak, J., Fan, H., Unnerstall, U., Emberly, E., Rajewsky, N., Siggia, E. D., & Gaul, U. (2004) *PLoS Biol* **2,** E271.
2. Papatsenko, D. A., Makeev, V. J., Lifanov, A. P., Regnier, M., Nazina, A. G., & Desplan, C. (2002) *Genome Res* **12,** 470-481.
3. Berman, B. P., Pfeiffer, B. D., Laverty, T. R., Salzberg, S. L., Rubin, G. M., Eisen, M. B., & Celniker, S. E. (2004) *Genome Biol* **5,** R61.
4. Lebrecht, D., Foehr, M., Smith, E., Lopes, F. J., Vanario-Alonso, C. E., Reinitz, J., Burz, D. S., & Hanes, S. D. (2005) *Proc Natl Acad Sci U S A* **102,** 13176-13181.
5. Pettersson, M. & Schaffner, W. (1990) *J Mol Biol* **214,** 373-380.
6. Ptashne, M. & Gann, A. (2002) *Genes & signals* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.).
7. Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., & Phillips, R. (2005) *Curr Opin Genet Dev* **15,** 116-124.
8. Arnosti, D. N., Barolo, S., Levine, M., & Small, S. (1996) *Development* **122,** 205-214.
9. Saiz, L. & Vilar, J. M. (2006) *Curr Opin Struct Biol* **16,** 344-350.
10. Hare, E. E., Peterson, B. K., Iyer, V. N., Meier, R., & Eisen, M. B. (2008) *PLoS Genet* **4,** e1000106.
11. Moses, A. M., Pollard, D. A., Nix, D. A., Iyer, V. N., Li, X. Y., Biggin, M. D., & Eisen, M. B. (2006) *PLoS Comput Biol* **2,** e130.
12. Arnosti, D. N. & Kulkarni, M. M. (2005) *J Cell Biochem* **94,** 890-898.
13. Lin, Y. S., Carey, M., Ptashne, M., & Green, M. R. (1990) *Nature* **345,** 359-361.
14. Miller, J. A. & Widom, J. (2003) *Mol Cell Biol* **23,** 1623-1632.
15. Vashee, S., Melcher, K., Ding, W. V., Johnston, S. A., & Kodadek, T. (1998) *Curr Biol* **8,** 452-458.
16. Adams, C. C. & Workman, J. L. (1995) *Mol Cell Biol* **15,** 1405-1421.
17. Morse, R. H. (2003) *Biochem Cell Biol* **81,** 101-112.
18. Polach, K. J. & Widom, J. (1996) *J Mol Biol* **258,** 800-812.
19. Lee, W., Tillo, D., Bray, N., Morse, R. H., Davis, R. W., Hughes, T. R., & Nislow, C. (2007) *Nat Genet* **39,** 1235-1244.
20. Monod, J., Wyman, J., & Changeux, J. P. (1965) *J Mol Biol* **12,** 88-118.
21. Li, B., Carey, M., & Workman, J. L. (2007) *Cell* **128,** 707-719.
22. Poirier, M. G., Bussiek, M., Langowski, J., & Widom, J. (2008) *J Mol Biol* **379,** 772-786.
23. Maerkl, S. J. & Quake, S. R. (2007) *Science* **315,** 233-237.
24. Weinberg, R. L., Veprintsev, D. B., Bycroft, M., & Fersht, A. R. (2005) *J Mol Biol* **348,** 589-596.
25. Chen, S. C., Zhao, T., Gordon, G. J., & Murphy, R. F. (2007) *Bioinformatics* **23,** i66-71.
26. Milo, R. (2008).
27. Sabo, P. J., Kuehn, M. S., Thurman, R., Johnson, B. E., Johnson, E. M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A.*, et al.* (2006) *Nat Methods* **3,** 511-518.
28. Anderson, J. D. & Widom, J. (2000) *J Mol Biol* **296,** 979-987.

29. Li, G., Levitus, M., Bustamante, C., & Widom, J. (2005) *Nat Struct Mol Biol* **12,** 46-53.
30. Lee, C. K., Shibata, Y., Rao, B., Strahl, B. D., & Lieb, J. D. (2004) *Nat Genet* **36,** 900-905.
31. Dion, M. F., Kaplan, T., Kim, M., Buratowski, S., Friedman, N., & Rando, O. J. (2007) *Science* **315,** 1405-1408.
32. Phair, R. D., Scaffidi, P., Elbi, C., Vecerova, J., Dey, A., Ozato, K., Brown, D. T., Hager, G., Bustin, M., & Misteli, T. (2004) *Mol Cell Biol* **24,** 6393-6402.
33. Mito, Y., Henikoff, J. G., & Henikoff, S. (2007) *Science* **315,** 1408-1411.
34. John, S., Sabo, P. J., Johnson, T. A., Sung, M. H., Biddie, S. C., Lightman, S. L., Voss, T. C., Davis, S. R., Meltzer, P. S., Stamatoyannopoulos, J. A.*, et al.* (2008) *Mol Cell* **29,** 611-624.
35. Morse, R. H. (2007) *J Cell Biochem* **102,** 560-570.
36. Sabo, P. J., Hawrylycz, M., Wallace, J. C., Humbert, R., Yu, M., Shafer, A., Kawamoto, J., Hall, R., Mack, J., Dorschner, M. O.*, et al.* (2004) *Proc Natl Acad Sci U S A* **101,** 16837-16842.
37. Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J. A., Costa, G., McKernan, K.*, et al.* (2008) *Genome Res* **18,** 1051-1063.
38. Mavrich, T. N., Jiang, C., Ioshikhes, I. P., Li, X., Venters, B. J., Zanton, S. J., Tomsho, L. P., Qi, J., Glaser, R. L., Schuster, S. C.*, et al.* (2008) *Nature* **453,** 358-362.
39. Oliviero, S. & Struhl, K. (1991) *Proc Natl Acad Sci U S A* **88,** 224-228.
40. Levine, M. (2008) *Genome Biol* **9,** 207.
41. Ioshikhes, I., Bolshoy, A., Derenshteyn, K., Borodovsky, M., & Trifonov, E. N. (1996) *J Mol Biol* **262,** 129-139.
42. Ioshikhes, I. P., Albert, I., Zanton, S. J., & Pugh, B. F. (2006) *Nat Genet* **38,** 1210-1215.
43. Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I. K., Wang, J. P., & Widom, J. (2006) *Nature* **442,** 772-778.
44. Morozov, A. V., Fortney, K., Gaykalova, D. A., Studitsky, V. M., Widom, J., & Siggia, E. D. (2008).
45. Peckham, H. E., Thurman, R. E., Fu, Y., Stamatoyannopoulos, J. A., Noble, W. S., Struhl, K., & Weng, Z. (2007) *Genome Res* **17,** 1170-1177.
46. Kim, H. D. & O'Shea, E. K. (2008) *Nat Struct Mol Biol* **15,** 1192-1198.
47. Lam, F. H., Steger, D. J., & O'Shea, E. K. (2008) *Nature* **453,** 246-250.
48. Xu, H. E., Kodadek, T., & Johnston, S. A. (1995) *Proc Natl Acad Sci U S A* **92,** 7677-7680.
49. Gao, Q. & Finkelstein, R. (1998) *Development* **125,** 4185-4193.
50. Sinha, S., van Nimwegen, E., & Siggia, E. D. (2003) *Bioinformatics* **19 Suppl 1,** i292-301.
51. Zhang, C., Xuan, Z., Otto, S., Hover, J. R., McCorkle, S. R., Mandel, G., & Zhang, M. Q. (2006) *Nucleic Acids Res* **34,** 2238-2246.
52. Sekinger, E. A., Moqtaderi, Z., & Struhl, K. (2005) in *Molecular Cell*, pp. 735-748.
53. Berg, J., Willmann, S., & Lassig, M. (2004) *BMC Evol Biol* **4,** 42.
54. Li, G. & Widom, J. (2004) *Nat Struct Mol Biol* **11,** 763-769.
55. Sneppen, K., Micheelsen, M. A., & Dodd, I. B. (2008) *Mol Syst Biol* **4,** 182.
56. Voet, D. & Voet, J. G. (2004) *Biochemistry* (J. Wiley & Sons, Hoboken, NJ).
57. Panne, D., Maniatis, T., & Harrison, S. C. (2007) *Cell* **129,** 1111-1123.

58. Bernstein, B. E., Liu, C. L., Humphrey, E. L., Perlstein, E. O., & Schreiber, S. L. (2004) *Genome Biol* **5,** R62.
59. Dhasarathy, A. & Kladde, M. P. (2005) *Mol Cell Biol* **25,** 2698-2707.
60. Erkina, T. Y., Tschetter, P. A., & Erkine, A. M. (2008) *Mol Cell Biol* **28,** 1207-1217.
61. Gross, D. S., Adams, C. C., Lee, S., & Stentz, B. (1993) *EMBO J* **12,** 3931-3945.

Figure legends

Figure 1. The model of nucleosome-mediated cooperativity. A. A region of DNA that contains an array of ***n*** sites (green boxes) can be bound by a histone core (red oval), thus becoming a nucleosomal DNA, or can 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 the naked DNA ($K_O << K_N$), but is possible due to transient, partial unwinding of the DNA (shown in **B**) (28). The system is in thermal equilibrium and is fully characterized by the scheme in **C.** Each state of the system is determined by the form of the DNA: 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 T and R states of the hemoglobin; binding of TFs is equivalent to binding of the oxygen molecule; the attenuation of TF affinity to their sites in the nucleosome corresponds to weaker affinity for the oxygen in the T state. At a low concentration of TFs (low oxygen pressure), the DNA is mostly in the nucleosomal state (the hemoglobin is in the T state). At a high concentration of TFs (oxygen), binding of ligand pulls the equilibrium toward the open form (R state), displacing the nucleosome. The key feature of the MWC model is the coordinated transition between R and T states of all four domains, corresponding to binding of the histone octamer to the whole stretch of 147bps of DNA. Like MWC, the nucleosome-TF system is determined by three dimensionless parameters: *L, c* and α that respectively control the preference of the N (T) state in the absence of the ligand, attenuated affinity for the ligand in this state, and the ligand concentration normalized to affinity for a site in the O (R) state. The range of values is estimated using experimentally measured quantities (see Supplement). **B.** Assisted unwrapping Polach-Widom (18) mechanism of synergistic binding by two TFs to nearby sites through partial unwinding of nucleosomal DNA. The mechanism requires presence of stable nucleosome and does not lead to nucleosome eviction. In our model binding of multiple TFs can evict nucleosome completely thus allowing much more distant sites to interact, more sites to be involved and hence a much higher Hill coefficient.

Figure 2. Nucleosome-mediated cooperativity and its implications. A. Cooperative transition in the equilibrium occupancy of TFs, *Y* (green line), and nucleosome, Y_N (red line), as a function of TF concentration (eq. 1 and 2). Here and below, $n=5$, $L=10^3$, and $c=10^{-3}$ unless stated otherwise. **B.** 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). **C.** Bohr effect: attenuation of nucleosomal core affinity for DNA, due to modifications or as a function of DNA sequence (modified - dashed line, unmodified - solid line), leads to a shift of balance in TF-nucleosome competition and displacement of the nucleosome by TFs (arrow). This competition makes nucleosomal occupancy, Y_N, respond considerably to small changes in nucleosome affinity **(inset),** as demonstrated by the dependence of Y_N on $-\Delta G = k_B T \log(L)$ *(kcal/mol)* (see text for details). **D.** Effect of the 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 clear critical number of sites (~4-5) below which TFs are unable to displace a nucleosome and above which the nucleosome in unstable even at a low concentration of TFs. This effect explains the clustering of sites in regulatory regions and demonstrates the origin of chromatin hypersensitive sites among such regions.

Figure 3. Cooperative binding to high- and low-affinity sites. The nucleosomal (red) and TF (green) occupancy profiles at a CRR that contains high-affinity (strong) and low-affinity (weak) sites. The top diagram shows the binding energy profile: a cluster of 8 low-affinity sites and a cluster of 5 high-affinity sites located over the background of scattered low-affinity sites. The CRR is packed by nucleosomes (nucleosomal binding profile is not shown). Three diagrams show

nucleosomal occupancy Y_N (red), and TF cluster occupancy P_3 (probability of having at least three sites in a cluster occupied, green) for three values of TF concentration (see Supplement). These profiles show that at TFBSs, rich clusters can become nucleosome-free at the intermediate concentrations of TFs (see text for comparison to experiments). While intermediate TF concentration is sufficient to get high-affinity clusters nucleosome-free and TF-bound, a higher concentration may be needed for clusters of low-affinity site. Combination of low- and high-affinity sites in a CRR can make it respond differently to different TF concentrations. Notice that a cluster of several low-affinity TF is required to destabilize a nucleosome; an isolated low-affinity site is unable to do this. Nucleosomes were considered to be well positioned by the DNA sequence and thus considered independently. The following parameters were used: *c=0.01; L=1000* (for positioned nucleosomes), $\alpha_{non\text{-}site} = 0.001$; $\alpha_{high\text{-}affinity} = 20$; $\alpha_{low\text{-}affinity} = 1$.

Tables

Table 1. Comparison of the nucleosome-mediated cooperativity and cooperative transition in hemoglobin.

Cooperative transition in hemoglobin, 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 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>>1)	Prevalence of the **N** state (stable nucleosome) at low TF concentration (L>>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 TF
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 2. Comparison of the model's predictions with experiments.

Model of the nucleosome-mediated cooperativity	Experimental data
Cooperative binding by non-interacting TFs	Gal4, USF and NF-kappa B bind cooperatively to reconstituted nucleosome *in vitro*.
Cooperative action of TF that bind within a footprint of a nucleosome <150bps, and independent of site orientations.	Cooperativity of Gal4 and USF independent of site orientations (16). LexA and Gal4, with one TF lacking the activation domain, cooperate up to a range of 200bps.
Lack of cooperative binding in the absence of the nucleosomes.	NF-kappa B act synergistically at a promoter containing four sites. Binding is not cooperative in vitro (5)
Cooperative binding of TFs does not require direct interactions between them.	Structure of interferon-β enhanceosome demonstrating very few direct contacts between bound TFs (57).
Displacement of a nucleosome that occupies a TFBS-rich region	Enrichment of TFBSs in nucleosome-depleted regions (34, 58), nucleosome depletion in yeast regulatory regions (promoters) (30, 31).
Critical number (n_c=4-6) of TF binding sites required for cooperative binding.	Clustering of TF binding sites in Drosophila enhancers, exceding 20 sites per kb (2, 3). Importance of clustering in eve2 (8).
High concentration of TFs is required to displace a stable nucleosome, lower for modified nucleosome (Bohr effect).	Recruitment of chromatin remodeling is required for activation through low-affinity sites. Overexpression compensates for the lack of remodeling (59, 60). Overexpression of TF compensates for mutation in high-affinity site, leading to nucleosome eviction through binding to two low-affinity sites (61).
Nucleosomal occupancy depends on the presence of TFBSs.	Nucleosomal occupancy is restored by mutations eliminating TFBSs (58). Mutation in the high-affinity site of HSF reduces nucleosome eviction in vivo (61).

Fig.1

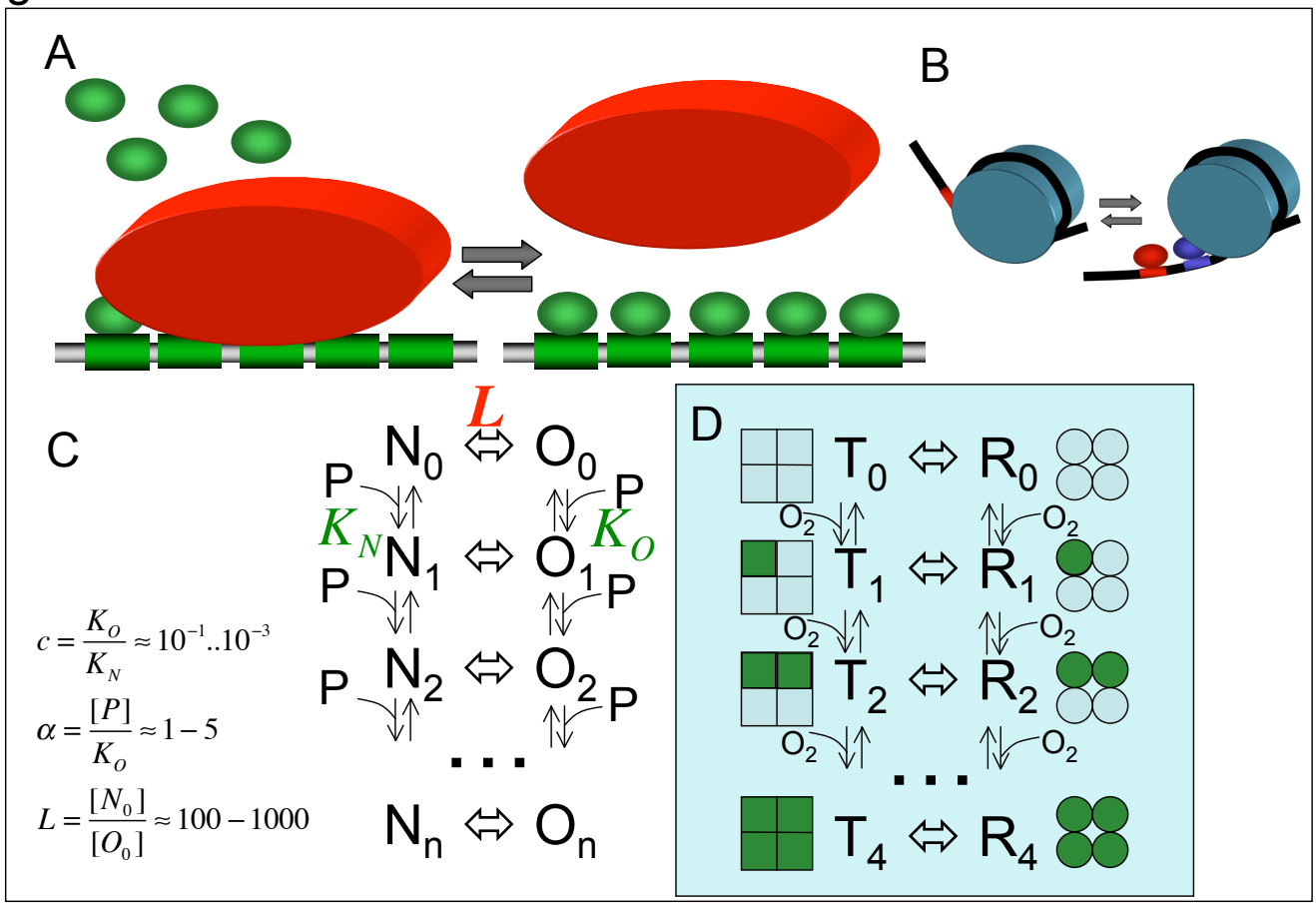

A
B
C
D
L
N_0 ⇔ O_0
P
K_N
K_O
N_1 ⇔ O_1
N_2 ⇔ O_2
N_n ⇔ O_n
$c = \frac{K_O}{K_N} \approx 10^{-1}..10^{-3}$
$\alpha = \frac{[P]}{K_O} \approx 1-5$
$L = \frac{[N_0]}{[O_0]} \approx 100-1000$
T_0 ⇔ R_0
T_1 ⇔ R_1
T_2 ⇔ R_2
T_4 ⇔ R_4
O_2

Fig.2

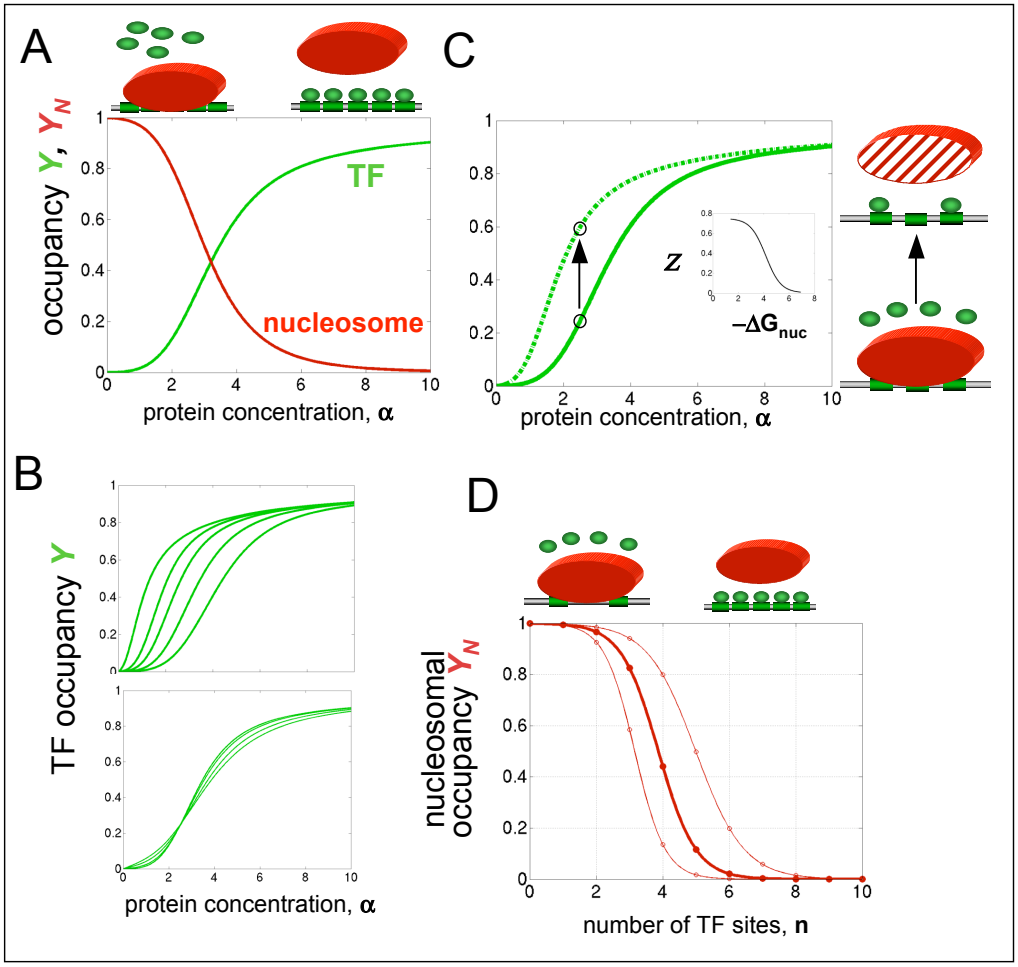
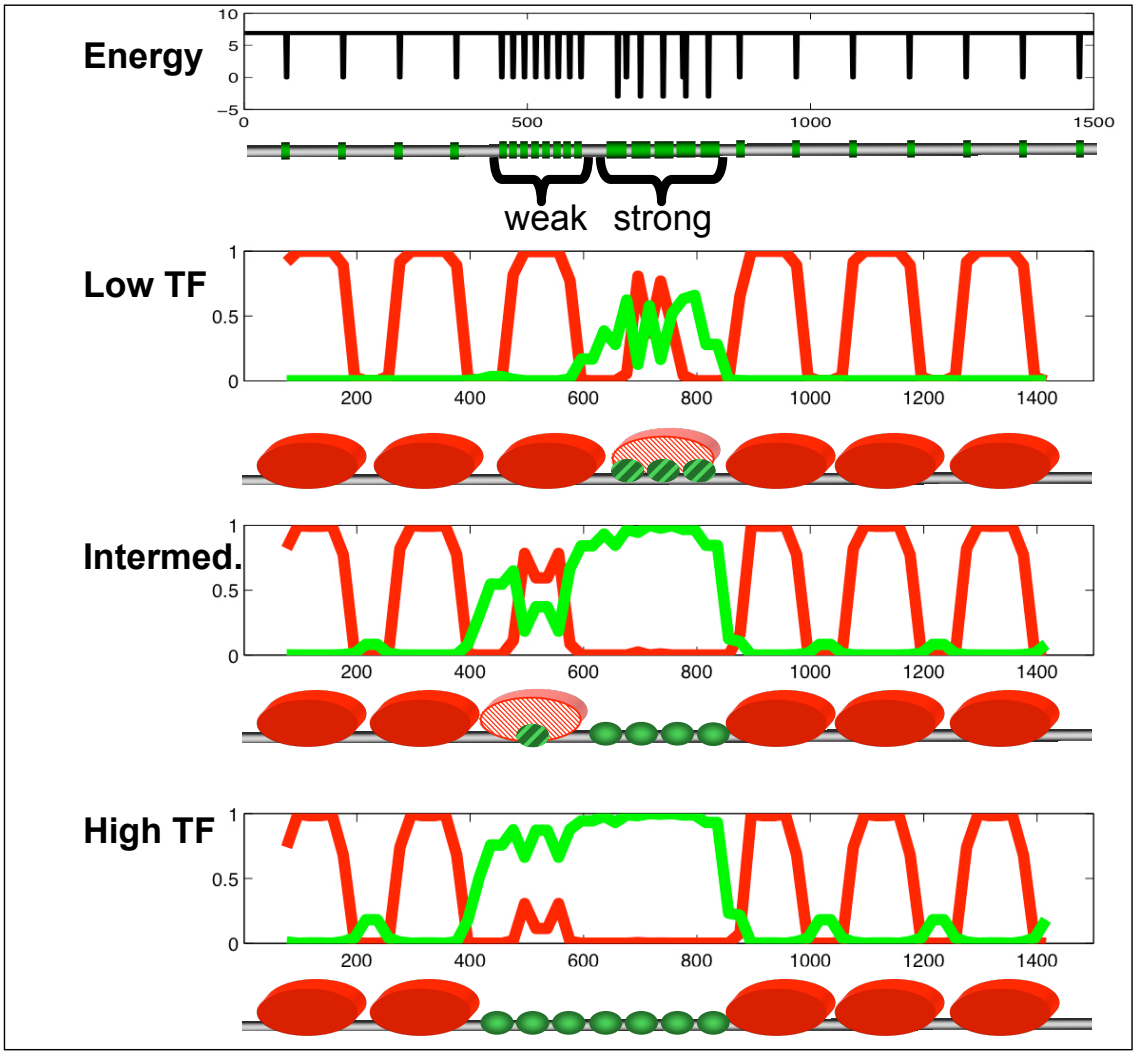

A
occupancy Y, Y_N
TF
nucleosome
protein concentration, α
C
Z
−ΔG_nuc
protein concentration, α
B
TF occupancy Y
protein concentration, α
D
nucleosomal occupancy Y_N
number of TF sites, n

Fig.3


Energy
10
5
0
−5
0
500
1000
1500
weak
strong
Low TF
1
0.5
0
200
400
600
800
1000
1200
1400
Intermed.
1
0.5
0
200
400
600
800
1000
1200
1400
High TF
1
0.5
0
200
400
600
800
1000
1200
1400