



Incorporating nucleosomes into thermodynamic models of transcription regulation

Tali Raveh-Sadka, Michal Levo and Eran Segal

Genome Res. 2009 19: 1480-1496 originally published online May 18, 2009

Access the most recent version at doi:[10.1101/gr.088260.108](https://doi.org/10.1101/gr.088260.108)

References This article cites 49 articles, 19 of which can be accessed free at:
<http://genome.cshlp.org/content/19/8/1480.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Copyright © 2009 by Cold Spring Harbor Laboratory Press

Methods

Incorporating nucleosomes into thermodynamic models of transcription regulation

Tali Raveh-Sadka,¹ Michal Levo,¹ and Eran Segal²

Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel

Transcriptional control is central to many cellular processes, and, consequently, much effort has been devoted to understanding its underlying mechanisms. The organization of nucleosomes along promoter regions is important for this process, since most transcription factors cannot bind nucleosomal sequences and thus compete with nucleosomes for DNA access. This competition is governed by the relative concentrations of nucleosomes and transcription factors and by their respective sequence binding preferences. However, despite its importance, a mechanistic understanding of the quantitative effects that the competition between nucleosomes and factors has on transcription is still missing. Here we use a thermodynamic framework based on fundamental principles of statistical mechanics to explore theoretically the effect that different nucleosome organizations along promoters have on the activation dynamics of promoters in response to varying concentrations of the regulating factors. We show that even simple landscapes of nucleosome organization reproduce experimental results regarding the effect of nucleosomes as general repressors and as generators of obligate binding cooperativity between factors. Our modeling framework also allows us to characterize the effects that various sequence elements of promoters have on the induction threshold and on the shape of the promoter activation curves. Finally, we show that using only sequence preferences for nucleosomes and transcription factors, our model can also predict expression behavior of real promoter sequences, thereby underscoring the importance of the interplay between nucleosomes and factors in determining expression kinetics.

[Supplemental material is available online at www.genome.org. The computational model described in this paper is available at http://genie.weizmann.ac.il/pubs/tf_nuc_model09/.]

The control over when and where each gene is expressed and to what extent is of fundamental importance in nearly all biological processes. Since the discovery of RNA polymerase by Weiss and Gladstone in 1959 (Weiss and Gladstone 1959), scientists have been trying to understand the mechanisms that underlie this regulation. In the following years, proteins termed “transcription factors,” which bind to specific sites on the DNA and affect the transcription of neighboring genes, were identified (Jacob and Monod 1961; Ptashne 1967). Subsequent studies of transcription focused on these proteins, and much progress has been made in identifying the targets and binding specificities of many transcription factors. Recently, high-throughput experimental methods such as ChIP-chip (Ren et al. 2000; Iyer et al. 2001; MacIsaac et al. 2006), ChIP-seq (Johnson et al. 2007), protein binding microarrays (Bulyk et al. 2001), and specialized microfluidics platforms (Maerkl and Quake 2007) allowed a more global characterization of transcription factors.

In addition to the role of transcription factors in transcriptional regulation, the role of chromatin in this regulation process has also been the subject of much research. Since nucleosomes compact $\approx 75\%$ – 90% of the total genomic DNA (van Holde 1989), it was speculated that nucleosomes will be of importance in transcriptional control. However, when this nucleosome hypothesis was first proposed by Roger Kornberg in 1974 (Kornberg 1974), many researchers believed that nucleosomes were transparent to the transcriptional machinery in the sense that they could be easily removed when DNA-binding proteins required access to the underlying DNA. Subsequent experiments falsified this view, and it is now widely accepted that the organization of nucleosomes

can significantly affect transcription (Han and Grunstein 1988; Miller and Widom 2003; Lam et al. 2008).

Recent experiments established that the histone octamer has differing binding affinities to different DNA sequences (Thastrom et al. 1999; Anderson and Widom 2001; Sekinger et al. 2005; Segal et al. 2006; Kaplan et al. 2009), most likely because DNA sequences differ greatly in their ability to sharply bend and conform to the nucleosome structure (Richmond and Davey 2003). A consequence of these nucleosome sequence preferences is that different genomic regions encode different nucleosome affinity landscapes, and thus direct different patterns of nucleosome organization. Since most transcription factors cannot bind sequences that are already occluded by nucleosomes, transcription factors seeking access to specific genomic locations need to compete with nucleosomes for access to the DNA, where the competition at specific locations depends on the binding affinity landscapes and concentrations of both the nucleosomes and the transcription factors. For transcription factor binding sites located at genomic regions that are also energetically favorable for nucleosome formation, this competition may result in significantly reduced binding of the cognate transcription factor and thus have major consequences for gene expression. However, a comprehensive and quantitative understanding of the possible effects that different nucleosome affinity landscapes may have on gene expression is still missing.

To attain such an understanding, a quantitative model that combines the various components of the transcriptional system is much needed. Thermodynamic models are one such appealing approach, since these models naturally arise from fundamental principles of statistical mechanics. In its most general form, this approach can model the binding of various molecules that are involved in the transcriptional process, such as transcription factors, nucleosomes, and RNA polymerase. In principle, all molecules can bind everywhere along the sequence, but the probability of binding at each sequence location depends on the molecule's concentration

¹These authors contributed equally to this work.

²Corresponding author.

E-mail eran@weizmann.ac.il; fax +972-8-9346023.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.088260.108>.

and binding affinity. Additionally, since steric hindrance effects usually do not permit the overlapping binding of two molecules to the same sequence locations, the probability of binding also depends on the presence of additional molecules competing for binding at the same locations. Under the assumption of thermodynamic equilibrium, all of these elements naturally combine using the Boltzmann distribution to produce a distribution over all possible organizations (or configurations) of bound molecules on each DNA sequence. This distribution can later be used to compute various quantities such as the binding probability of a transcription factor at a specific sequence location, and the probability that a certain base pair is covered by nucleosomes. Importantly, this distribution can also be used to predict the resulting expression level for a given DNA sequence (Fig. 1).

Several papers have used a thermodynamic approach (Shea and Ackers 1985; Rajewsky et al. 2002; Buchler et al. 2003; Sinha et al. 2003; Bintu et al. 2005; Segal et al. 2006, 2008), significantly advancing the ability to predict DNA binding events and expression patterns from sequence. However, these models focused solely on the contribution of transcription factors (Shea and Ackers 1985; Rajewsky et al. 2002; Buchler et al. 2003; Sinha et al. 2003; Bintu et al. 2005; Segal et al. 2008) and did not model other transcriptional components such as nucleosomes, or they only modeled binding of histone octamers and disregarded the binding of transcription factors (Segal et al. 2006).

Here we extend the thermodynamic framework and model the binding of both histone octamers and transcription factors. Using this extension, we theoretically explore the interactions between transcription factors and nucleosomes and the possible effects that various nucleosome affinity landscapes have on transcription. We show that our analysis reproduces experimental results that have been made regarding the effect of nucleosomes as general repressors and as generators of obligate binding cooperativity between transcription factors. Our modeling approach also offers simple mechanistic explanations to complex kinetic behaviors (Lam et al. 2008) and generates new testable hypotheses regarding the diverse range of transcriptional responses that can be achieved using different nucleosome affinity landscapes. We thoroughly examine the role of different promoter elements in determining nucleosome and transcription factor affinity landscapes and their resulting effect on the promoter activation dynamics curves, the curves describing the induction levels of a promoter in response to increasing concentrations of the regulating transcription factor. We also characterize these effects analytically, observing that while some changes to the affinity landscapes of transcription factors and nucleosomes result in a simple effect of a shift (in log scale) of the activation curves, other changes affect the activation curve in a more complex way. Such effects on activation curves have distinct consequences in terms of cell-to-cell expression variability, defining the location and extent of the range of transcription factor concentrations in which promoters are “noisy.” By exploring the relations between nucleosomes and transcription, our study represents a step toward a unified and quantitative understanding of transcriptional control and toward the ability to predict expression patterns from sequence.

Methods

Modeling framework

We now formally describe how we compute the distribution over the possible organizations or configurations of molecules along the regulatory sequence. As mentioned above, assuming a state of

thermodynamic equilibrium, this distribution is dictated by the concentrations and binding affinities of the various molecules.

To compute this distribution $P(c)$ over all configurations c on a given sequence, we consider all molecules simultaneously and evaluate all possible “legal” configurations of molecules binding to the sequence. A configuration is legal if there is no overlap between two molecules on the DNA. Thus, we explicitly model the competition between molecules that results from their mutual steric hindrance constraints, which for simplicity we take as the lack of overlap in binding to the same base pair. Note, however, that molecules can occupy the same sequence in different configurations. In accordance with the Boltzmann distribution, it follows that the probability of each configuration is given by:

$$P(c) = \frac{W(c)}{\sum_{c' \in C} W(c')} \quad (1)$$

where $W(c)$ represents the statistical weight associated with configuration c . We are thus left with defining this statistical weight. Intuitively, if we assume that molecules bind independently (except for mutually exclusive binding due to steric hindrance constraints), then the statistical weight of a legal configuration c should be the product of the contribution of each of the molecules bound in the configuration. The contribution of a molecule bound at a given position within the configuration is, in turn, determined by its concentration, and by the affinity, or energetic gain, from its binding to the underlying sequence at the bound position. Thus, for a configuration with k molecules m_1, \dots, m_k bound at positions p_1, \dots, p_k , the statistical weight $W(c)$ of the configuration is given by:

$$W(c) = F_0 \prod_{i=1}^k \tau_{m_i} \cdot F_{(m_i, p_i, p_i + L_i)} \quad (2)$$

where τ_{m_i} is the concentration of the molecule bound at position p_i , $F_{(m_i, p_i, p_i + L_i)} = e^{-E_i/K_B \cdot T}$ is the energetic contribution from the binding of molecule m_i from position p_i to position $p_i + L_i$, with L_i being the length of the binding site for molecule m_i , and F_0 is the energetic contribution from the empty configuration without any bound molecules. Since F_0 appears in all configurations, it cancels out in the computation of $P(c)$. Therefore, its value can be arbitrarily chosen, and we set its value to 1 in order to simplify our calculations. Although the computation of $\sum_{c' \in C} W(c')$ seems complex, as it requires traversing a huge space of legal configurations, we can use dynamic programming to compute it efficiently, in time linear in the length of the sequence (Segal et al. 2008).

Modeling assumptions

As with all models, our model makes several simplifying assumptions, which we list below, allowing us to focus our attention on the effect of nucleosome organization on transcription factor binding. As we note in the Discussion section, most of these assumptions can easily be relaxed in more advanced models.

First, we assume that the system is in a state of thermodynamic equilibrium, such that each configuration is achieved with a probability proportional to its probability under the Boltzmann distribution. This assumption is justified in cases in which there is separation of time scales, for example, if the transcription rate is much slower than the rate at which transcription factors and histones bind and unbind DNA.

Second, we assume that the histone octamer binds DNA according to the displacement model, that is, a nucleosome wraps

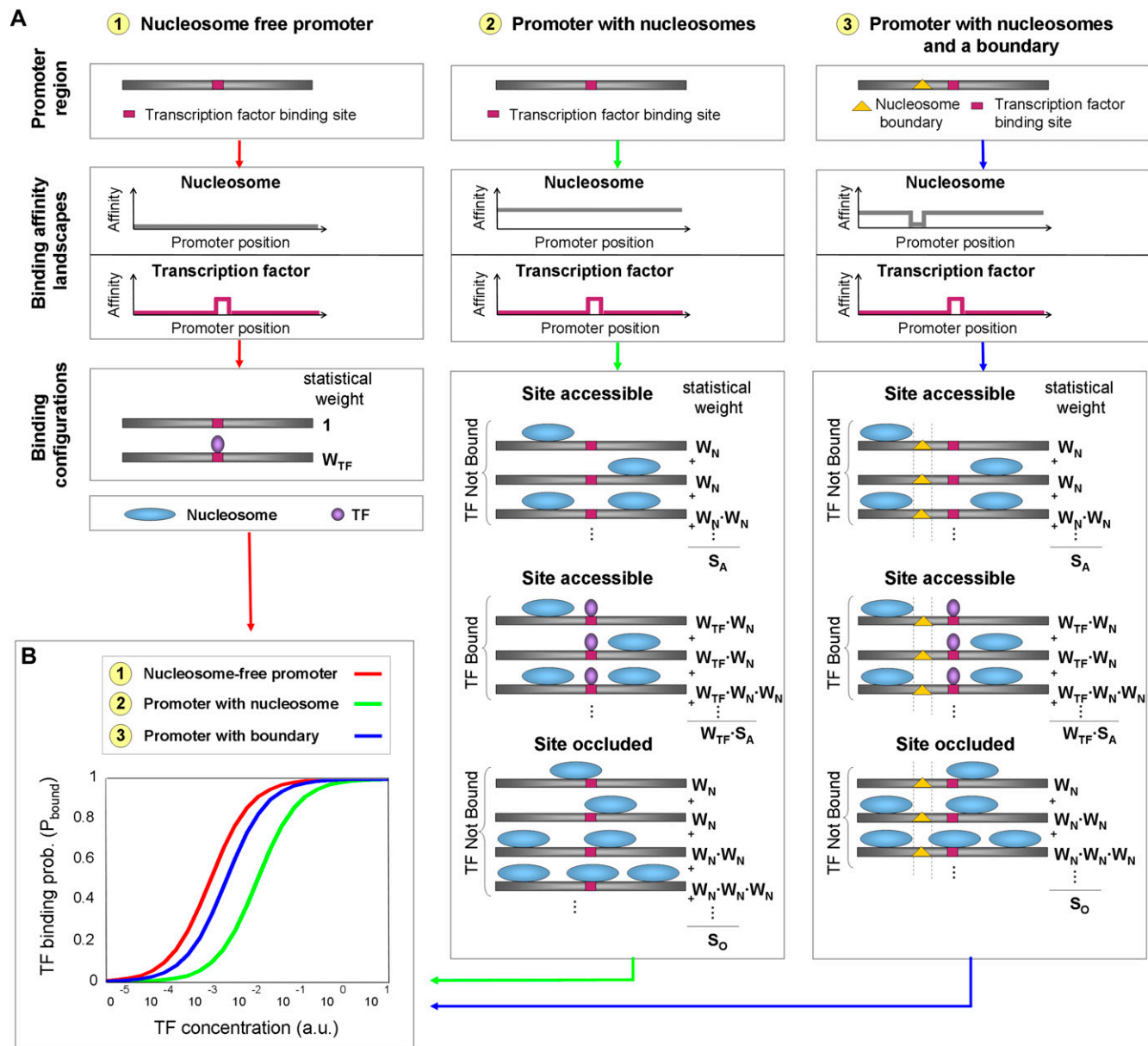


Figure 1. Illustration of our thermodynamic framework. Each promoter sequence encodes particular binding affinity landscapes for both transcription factors and nucleosomes. Given these landscapes as input as well as the concentrations of transcription factors and nucleosomes, our framework can then compute the distribution over all possible configurations of molecules bound to the promoter (see “Modeling Framework”). Applying these computations to different promoters (represented by different affinity landscapes) and over a range of transcription factor concentrations thus allows us to compute the activation curve of various promoters as a function of transcription factor concentrations. (A) Here we examine three promoters (discussed in detail in the Results section): (1) A nucleosome-free promoter, represented by a binding affinity landscape for nucleosomes that is zero at every promoter location. Shown are the two possible configurations (bound/unbound) with their respective statistical weights. (2) A promoter with a uniform binding affinity landscape for nucleosomes. Shown is a subset of the possible binding configurations, each with its respective statistical weight. (3) A promoter with a boundary element for nucleosome formation—resulting in a trough in the otherwise uniform landscape for nucleosomes. The transcription factor binding site is located 10 bp from the boundary element. A subset of the possible configurations is shown. Note that in all configurations the boundary element cannot be occluded by a nucleosome. (B) Promoter activation curves showing the probability of transcription factor binding at the site, P_{bound} , at various transcription factor concentrations (log scale, arbitrary units), for the three promoters in A. The activation curves of these promoters are identical in shape but shifted (in log scale) relative to one another.

the DNA along the entire length of the nucleosome (147 bp), with partial wrapping around the histone core disallowed (Polach and Widom 1995). Although experimental evidence supports both nucleosome displacement and partial unwrapping, as we explain below, nucleosome displacement arises as a natural consequence

of our framework, while partial unwrapping requires explicit modeling and additional parameters. Thus, for simplicity, here we only use the nucleosome displacement model.

Third, although some transcription factors are able to bind nucleosomal sequences (Perlmann and Wrangé 1988), most

transcription factors cannot, due to steric hindrance constraints. Our model thus does not allow binding locations of transcription factors and nucleosomes to overlap in the same binding configuration, thereby forcing transcription factors and nucleosomes to compete for DNA access.

Another simplifying decision is our choice to model only binding of transcription factors and histone octamers to DNA and leave out the transition from transcription factor binding to expression output. We assume that for transcriptional activators, the probability of binding is proportional to the binding probability of RNA polymerase, which is, in turn, proportional to the resulting transcription rate. Thus, we use the binding probabilities of transcription factors as a proxy to the resulting transcriptional output. Clearly, since more complex relations may exist between transcription factor binding and RNA polymerase binding, and between RNA polymerase binding and transcription output, there may be cases in which our proportionality assumption is violated. However, such simplifying assumptions were used by several models (Shea and Ackers 1985; Buchler et al. 2003; Bintu et al. 2005; Segal et al. 2008) and have led to predictive and informative models. Our model also ignores the potential effect of the distance between transcription factor binding sites and transcription start sites and the orientation of binding sites.

An additional noteworthy fact is that our model considers transcription factors and nucleosomes only, whereas other molecules, such as ATP-dependent chromatin remodelers (Cairns 2005; Whitehouse et al. 2007), are clearly important for a full understanding of transcription factor–nucleosome interactions. However, the precise functions of these molecules are not well understood, and we thus disregarded them in our model.

Regarding binding cooperativity, we do not explicitly model cooperative effects in transcription factor binding that may result, for example, from protein–protein interactions. However, a consequence of our modeling framework, which we discuss in detail in the Results section, is obligate binding cooperativity of transcription factors that results from the implicit modeling of competition between transcription factors and nucleosomes (Polach and Widom 1996; Miller and Widom 2003).

When examining activation dynamics curves for promoters, we will generally assume that upon induction, the increase in the concentration of the activating transcription factor is multiplicative. Therefore, activation curves will be presented in log-concentration scale. However, in several cases, we will also distinguish between the predictions made by our model under this assumption and under an alternative assumption according to which the increase in transcription factor concentrations is additive.

Finally, in order to explore the most fundamental effects of nucleosomes on transcription factor binding, we consider mostly simplified promoters. These include a small number of well-defined binding sites for transcription factors. This can be achieved in real promoters in which the transcription factor landscape is peaked, meaning that although transcription factors can potentially bind at any location along the sequence, their probability of binding in all locations along the sequence except at the well-defined sites is negligible. Similarly, we mostly consider simple binding affinity landscapes for nucleosomes, which are either uniform, in which nucleosomes can form with equal probability at each sequence location; peaked, in which nucleosomes can form only at specific locations; or boundary-like landscapes, in which the histone octamer has extremely high or low binding probability at a few sequence locations, and uniform binding probability at all other sequence locations. These simplifying assumptions allow us to analytically

study general phenomena that are applicable to many promoters, without the need to know the detailed binding affinity landscapes of specific promoters. Still, our framework can be easily and successfully adapted to real sequences using realistic models for nucleosome (Segal et al. 2006) and transcription factor binding (Bulyk et al. 2001; MacIsaac et al. 2006; Maerkl and Quake 2007), as we demonstrate below.

Toy example

In the sections below, we examine simple promoter architectures, represented by binding affinity landscapes of transcription factors and nucleosomes, and the activation curves that they produce. This is done by examining the effect of each architecture on the probability, P_{bound} , that a transcription factor occupies one or more of its sites on a given sequence. As explained above, we compute this term according to the rules of statistical mechanics, that is, as the ratio between the sum of the statistical weights of all configurations in which the transcription factor binds the site, and the sum of the statistical weights of all configurations.

As a simple illustration of such computations, assume that we have a single binding site for a transcription factor on a certain sequence. This can be achieved by a highly specific binding affinity landscape that is peaked around one sequence location. With no nucleosomes, this case has only two possible configurations (Fig. 1A, promoter 1), since the site can either be occupied by the transcription factor or not. Recall that the weight of the unbound configuration F_0 is arbitrarily set to 1. The weight of the bound configuration is then $W(c) = W_{TF} = \tau_{TF} \cdot F_{(TF:P_{site}:P_{site}+L_{TF})}$, where W_{TF} is introduced for notational simplicity, τ_{TF} is the transcription factor concentration, and $F_{(TF:P_{site}:P_{site}+L_{TF})}$ is the energetic contribution from the binding of the transcription factor to its site (see “Modeling Framework” above). Together, we get that the binding probability is $W_{TF}/(1+W_{TF})$.

Adding nucleosomes with a uniform binding affinity landscape (in which nucleosomes can form at any location with equal probability), we have many more configurations, both where the site is occupied by the transcription factor and where it is not bound by the transcription factor (Fig. 1A, promoter 2). In each configuration, in addition to the weight contribution $W_{TF} = \tau_{TF} \cdot F_{(TF:P_{site}:P_{site}+L_{TF})}$ made by the transcription factor (if bound), each nucleosome N_i contributes a weight denoted as $W_{N_i} = \tau_{N_i} \cdot F_{(N_i:P_{N_i}:P_{N_i}+L_{N_i})}$, which corresponds to the energetic contribution from the formation of the nucleosome at the bound location, multiplied by its concentration τ_{N_i} . Under this uniform binding affinity landscape for nucleosomes, this term is the same for all sequence locations and is denoted as W_N . To compute P_{bound} , we use a dynamic programming algorithm, which allows us to efficiently traverse all of the possible configurations where the transcription factor is bound, and sum their weights. This sum is then divided by the sum of weights of all configurations, computed by the same algorithm.

Parameters and definitions

In addition to W_{TF} and W_N above, our computations also depend on L_p , the length of the promoter, L_N , the length of the sequence wrapped around the histone octamer, and L_{TF} , the sequence length bound by the transcription factor. These parameters are assigned plausible although arbitrary values. For example, we use 60 as the value for W_N , since this value results in an average nucleosome coverage of ~90% of the DNA under a uniform affinity landscape,

in accordance with the literature. Our observations in the following text are general and do not depend qualitatively on the actual parameter values used. For a list of all parameters and values, see Table 1.

Results

Nucleosomes determine activation dynamics at the *PHO5* promoter

In this section, we apply our framework to the yeast *PHO5* promoter in order to demonstrate the ability of our framework to explore the interactions between transcription factors and nucleosomes and to offer mechanistic explanations for observed promoter kinetics.

In 2008, a paper by Lam et al. (2008) studied the activation of genes in the phosphate response system of yeast. In this system, low inorganic phosphate (P_i) levels activate the transcription factor Pho4p. Active Pho4p binds to promoters of several genes in the phosphate response pathway responsible for uptake and scavenging of P_i , including the *PHO5* gene, and promotes their expression. Lam and colleagues experimentally mapped the locations of nucleosomes in the *PHO5* promoter and observed that the strength of Pho4p sites not covered by nucleosomes in the uninduced state was predictive of the time of activation. To reach this conclusion, Lam et al. constructed a set of variants for the *PHO5* promoter, where each variant had a different combination of low- or high-affinity sites for Pho4p, that were either covered or not covered by nucleosomes in the uninduced state (Fig. 2A). The expression levels of each of these variants were measured under varying P_i levels, and showed a clear difference between variants with an exposed high-affinity Pho4p site, and variants with an exposed low-affinity Pho4p site in promoter activation times, where variants with an exposed high-affinity Pho4p site exhibited a significantly earlier activation time. Lam et al. proposed competition between transcription factors and nucleosomes as a possible mechanism for their observations. We show that this proposal also arises as a plausible hypothesis from our framework.

Using the measurements of Lam et al. (2008) regarding the location of Pho4p sites and the locations of nucleosomes in the uninduced *PHO5* promoter, we set the binding affinity landscapes of nucleosomes and transcription factors to be “peaked” around these measured locations. As expected from this construction, by applying our thermodynamic framework to these landscapes, we reproduce the partition of variants into two groups according to the strength of the exposed site (Fig. 2B,C). Recall that our mod-

eling framework does not explicitly model the transition from binding to expression. Thus, differences between the measured expression curves and predicted activation curves may stem from this transition.

Intriguingly, within the group with exposed low-affinity sites, our model is also able to correctly order the variants according to their measured activation times. The order that we predict (and that is observed in the data) is intuitive and is based on the total strength of sites in the promoter, where sites covered by nucleosomes contribute far less to fast activation time compared to exposed ones, but even for covered sites, high-affinity sites provide stronger competition for nucleosomes and thus contribute more to fast activation time compared to low-affinity sites. In the group with exposed high-affinity sites, we expect the predicted and experimental order of the different variants to follow the same logic. However, although these differences exist in our predictions, owing to the large (>4) fold differences in strength between the strong and weak Pho4p binding sites (Lam et al. 2008), the activation curves for all of the variants in this group are highly similar. Taking into account the existence of some level of experimental inaccuracies, it is thus not surprising that the experimental data of Lam and colleagues do not detect a robust ordering in this case.

We note that Lam et al. (2008) did not discuss the role of covered sites in determining the time of activation. However, our ability to reproduce the correct ordering observed experimentally for the variants with exposed low-affinity sites suggests that the contribution of covered sites does exist, even though it is considerably weakened by the competition with nucleosomes compared to the contribution of exposed sites.

Taken together, the analysis above shows that our simple mechanistic framework is able to explain experimental data with much detail and to offer mechanistic understanding of how these observations are generated. In the case above, we clearly see that the positioning of nucleosomes and competition with transcription factors on access to sites is of significant importance to the expression of the *PHO5* yeast promoter.

The *PHO5* yeast promoter is composed of a specific combination of a number of sequence elements. These include several Pho4p sites and a highly constrained affinity landscape for nucleosomes, making it hard to decipher the actual contribution of each promoter element to the resulting promoter activation kinetics, and diminishing our ability to generalize from the behavior seen in this promoter to predicted behaviors in other promoters. In the remaining sections, we perform a more systematic exploration of the role of nucleosomes in transcription by focusing on more abstract promoters that code for much simpler affinity landscapes for nucleosome and transcription factor binding.

We start by considering a simple promoter with a single site for transcription factor binding. We will then gradually add promoter elements, generating more complex affinity landscapes for nucleosomes and transcription factors, and examine the resulting activation curves. As an initial step in this process, we study the effect of introducing nucleosomes in the simplest way possible, using a uniform landscape for their formation.

Uniform landscapes for nucleosome formation: Nucleosomes as general repressors

The first role attributed to nucleosomes in transcriptional regulation was their role as general repressors (Han and Grunstein 1988). Here, we show that our thermodynamic framework easily reproduces this observation.

Table 1. Parameters and their typical values

Parameter	Description	Typical value
L_P	Promoter length	5000
L_N	Length of sequence bound by each nucleosome	147
L_{TF}	Length of sequence bound by each TF	6
W_N	Statistical weight of a single bound nucleosome	60
τ_{TF}	Transcription factor concentration	0.01
$F_{(TF, P_{site}, P_{site}+L_{TF})}$	Binding affinity of a single transcription factor	600

Unless otherwise stated, we used these parameters and values in our computations. L_P was arbitrarily set to be large enough to avoid boundary effects stemming from the sequence edges.

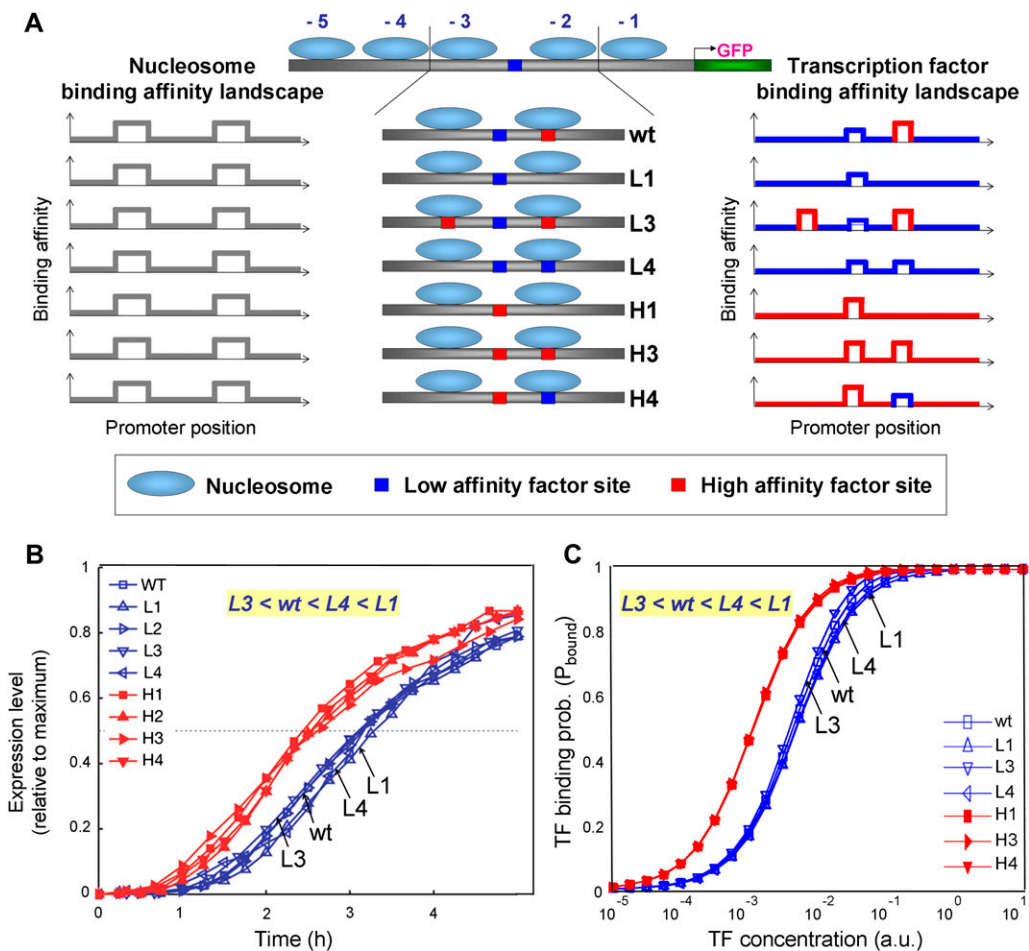


Figure 2. Predicted and observed activation curves for the *PHOS* promoter variants. (A) The *PHOS* promoter variants constructed by Lam et al. (2008), in which wild-type Pho4p sites were engineered to create various combinations of high- and low-affinity sites for the transcription factor, which were either exposed or covered by the -2 and -3 nucleosomes. For each variant, shown is a schematic representation of its binding affinity landscapes for transcription factors and nucleosomes. We excluded variants L2 and H2 since they are identical (because of symmetry) under our model to variants wild type and H3, respectively. (B) Experimental measurements by Lam et al. (2008) of the expression curves (scaled to maximum expression) of each variant following P_i starvation. [Reprinted with permission from Macmillan Publishers Ltd. © 2008, Lam et al. 2008.] (C) Prediction of the binding probability of Pho4p, P_{bound} , generated by applying our thermodynamic framework to each of the promoter variants from A at increasing Pho4p concentrations. In B and C, the yellow inset indicates the relative ordering of the onset times of the promoter variants with exposed low-affinity sites, where in both cases L3 is activated first, followed by wild type, L4, and L1. The energetic contribution ($F(TF, P_{site}, P_{site} + L_{TF})$) from transcription factor binding is set to 860 for a strong site, and 200 for a weak site, in accordance with the 4.3 ratio between strong and weak Pho4p sites reported in Lam et al. (2008).

In the past years, possible mechanisms were suggested for the observed repressive effects of nucleosomes, including their ability to change the topology of DNA and their ability to sterically block the binding of RNA polymerase and of the general transcriptional machinery. In addition, nucleosomes also compete with transcription factors, thereby reducing their binding probability. Assuming that binding of transcriptional activators is proportional to the expression output, this reduction in binding probability translates to a reduction in the output expression. Indeed, within our framework, even the incorporation of uniform landscapes for nucleosome formation results in the observed repressive effect on transcription factor binding probability (Fig. 1).

More formally, when nucleosomes are considered, we have several types of configurations, in addition to the two configurations possible (bound/unbound) in the nucleosome-free case (see the “Toy Example” section above). These types include configurations where the site is accessible (not occluded by nucleosomes) and the transcription factor is either not bound (with total

weight S_A) or bound (with total weight $W_{TF} \cdot S_A$), and configurations where the site is occluded by nucleosomes (with total weight S_O) (Fig. 1A, promoter 2). The probability of the site to be bound under this uniform binding landscape for nucleosomes, P_{bound}^{nuc} , is then given by:

$$\begin{aligned}
 P_{bound}^{nuc} &= \frac{W_{TF} \cdot S_A}{S_A + W_{TF} \cdot S_A + S_O} \\
 &< \frac{W_{TF} \cdot S_A}{S_A + W_{TF} \cdot S_A} \\
 &= \frac{W_{TF}}{1 + W_{TF}} = P_{bound}^{nuc-free}, \quad (3)
 \end{aligned}$$

which is smaller than the value $P_{bound}^{nuc-free}$ for the nucleosome-free situation. In fact, as can be seen in Figure 1B, when assuming a multiplicative increase in transcription factor concentrations, the P_{bound}^{nuc} graph is shifted toward activation at higher transcription

factor concentrations relative to the $P_{bound}^{nuc-free}$ graph, owing to the competition. This shift implies that the effect of adding nucleosomes is identical to that of lowering the affinity of the transcription factor to its site—in both cases to attain the same P_{bound} values as in the original nucleosome-free situation, the transcription factor concentration must be multiplied by a constant >1 . The value of this constant can be computed analytically (for details, see Supplemental material), and it depends on the length of the promoter, L_p , the location of the transcription factor site within the promoter, and the contribution from the formation (binding affinity and concentration) of each nucleosome, W_N . Note that if, alternatively, we assume that in living cells, changes in transcription factor concentrations are additive rather than multiplicative, the multiplication of the concentration by a constant results in an inflation (rather than a shift in log scale) of the activation curve. This means that each change in the P_{bound}^{nuc} graph requires larger additive changes in transcription factor concentration than in the corresponding nucleosome-free $P_{bound}^{nuc-free}$ graph.

Our assumption of a uniform landscape for nucleosome formation is a good approximation for promoters in which the differences between the probabilities for nucleosome formation at different locations are relatively small. Although nonuniform landscapes may produce other behaviors, transcription factor binding to a site placed under a stable nucleosome will in all cases exhibit this type of repression.

Addition of boundary elements for nucleosome formation shifts the activation curve

In this section, we take a step forward from the simplistic case of a promoter with a uniform affinity landscape for nucleosome formation and consider the addition of a new promoter element—the nucleosome boundary element.

The simplest nonuniform binding affinity landscape for nucleosomes is induced by sequences that act as boundaries for nucleosome formation, which are locations along the promoter where there are peaks or troughs in the otherwise uniform affinity landscape. Perfect boundaries are locations along the promoters where a nucleosome is either present with probability 1 (e.g., due to an extremely high-affinity sequence), or locations where a nucleosome is never bound (e.g., at chromosome ends). However, the more common boundaries are imperfect ones, where the probability for nucleosome formation is not extreme (integral) but rather higher/lower than in the uniform case. These types of boundaries can be formed by nucleosome attracting signals, or nucleosome repelling signals such as poly(dA:dT) tracts (Bernstein et al. 2004; Yuan et al. 2005; Lee et al. 2007; Field et al. 2008; Kaplan et al. 2009). Imperfect boundaries such as those that result over poly(dA:dT) tracts are common in eukaryotic genomes (Dechering et al. 1998), indicating a possible functionality for these sequence-encoded signals. Indeed, when located within promoter regions, these sequences are hypothesized to have important effects on transcription (Struhl 1985; Iyer and Struhl 1995). It is important to note that when dealing with strong boundaries, the effect of other weaker positioning signals if they exist is much less pronounced, and the otherwise uniform binding affinity landscape assumption becomes quite reasonable.

Below, we discuss the effects of adding boundaries for nucleosome formation to simple promoters with a single transcription factor binding site. The observations we make are valid not only for boundary elements as defined above, but also for the case of binding of some molecule (transcription factor or nucleosome)

whose concentration is constant relative to the concentration of the transcription factor under discussion. This is because such binding sterically hinders the formation of additional nucleosomes at the same sequence locations (Kornberg and Stryer 1988) and thus in practice serves as a boundary for nucleosome formation. The binding of molecules whose concentration is not constant shows different effects and is discussed separately in the next section.

The most basic characteristic of perfect boundaries is the independence they generate between binding events that occur at different sides of the boundary. Since binding events on two sides of the boundary cannot overlap, the distribution over binding configurations on one side does not depend on the distribution on the other side. This independence may be important for transcription regulation, for example, by providing a mechanism to nullify the effects of binding events at neighboring promoters. This way, a transcription factor or nucleosome that is positioned at one of these promoters will not affect binding events in another promoter located on the other side of the boundary. Although the more common imperfect boundaries do not generate complete independence, the stronger the boundary is, binding events on one of its sides will have less effect on binding events on its other side.

Another important effect generated by boundaries is an ordered positioning of nucleosomes (Fig. 3A). This periodic ordered positioning was predicted by Kornberg and Stryer (1988) and can also be observed in genome-wide measurements of nucleosomes (Yuan et al. 2005; Lee et al. 2007; Whitehouse et al. 2007). This pattern results from the fact that base pairs in the vicinity of a boundary can participate in a relatively small number of possible configurations in which these base pairs are covered by a nucleosome. As the distance from the boundary increases, the number of configurations in which the base pairs are covered by nucleosomes, and therefore their probability to be covered, increases. At distances from the boundary that are larger than L_N , an additional nucleosome can form between the boundary and the query base pair, thus acting as a secondary boundary and by that forming the observed periodic effect. For the more common case of imperfect boundaries, the observed pattern is generally the same, but is, as expected, less pronounced. In fact, this phenomenon is not specific to nucleosomes only, but rather to any molecule that can potentially bind to any sequence location. However, the effect is most pronounced at high molecule concentrations and is thus especially evident in the case of nucleosomes, because of their high abundance.

The described periodic pattern of nucleosome occupancy caused by a single boundary in a uniform binding affinity landscape can have significant implications with regard to transcription factor binding and the resulting promoter activation curves. From a transcription factor perspective, at any fixed concentration of the relevant transcription factor, the probability of binding a site at different distances from the boundary is inversely proportional to the probability of nucleosome formation at the same location (Fig. 3B). Compared to a promoter with the same length and transcription factor binding site but with no boundaries, one can easily observe that the reduced nucleosome occupancy near boundaries facilitates the binding of transcription factors to sites located in their immediate vicinity (or at other relatively exposed locations, i.e., troughs of Fig. 3A), thus allowing the boundary to act as a general indirect activator, contributing positively to transcription factor binding (in Fig. 3B, distances at which the red curve is above the light green curve). Indeed, the role of poly(dA:dT)

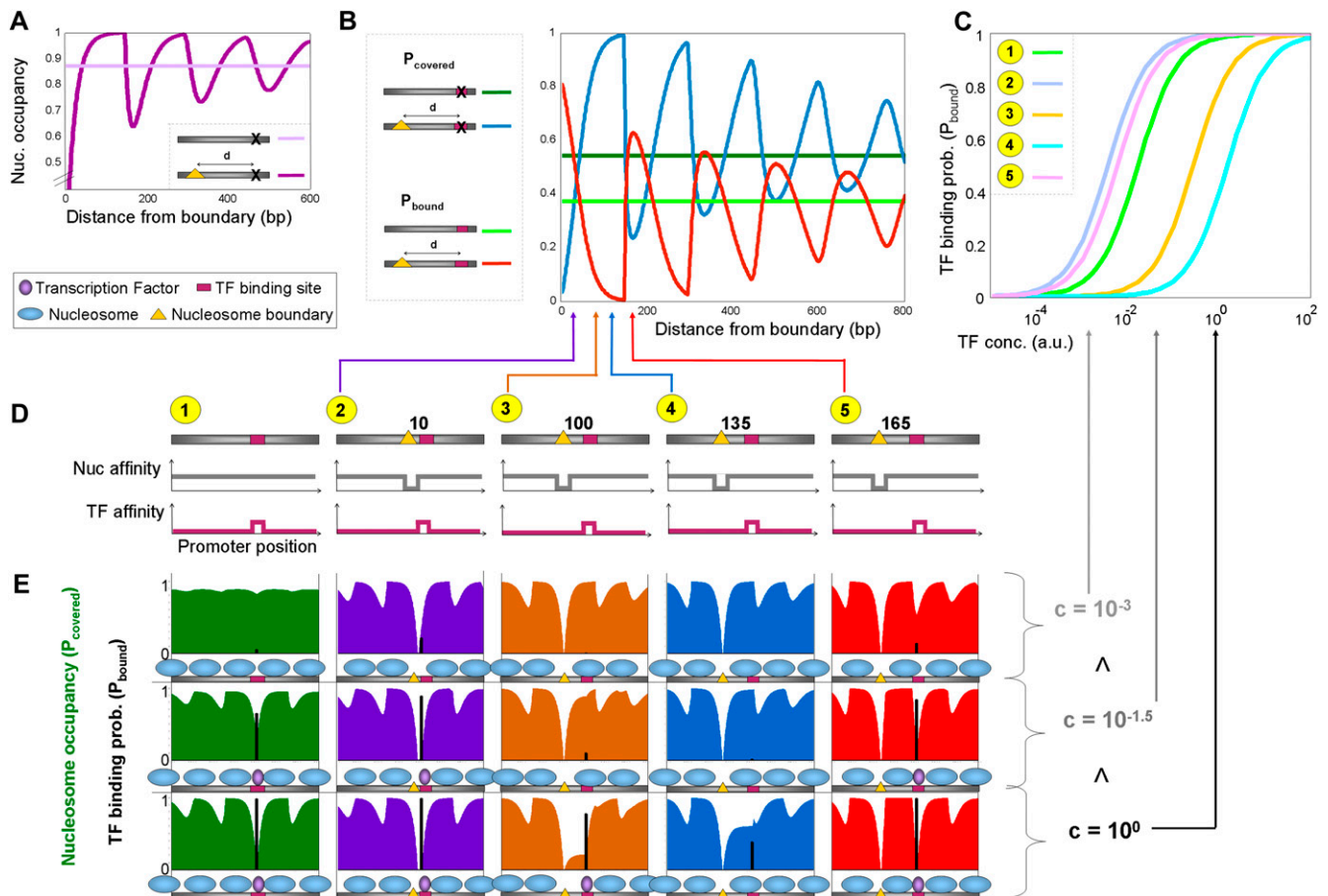


Figure 3. Addition of a boundary element for nucleosome formation to a simple promoter. (A) The periodic pattern of nucleosome occupancy induced by perfect boundaries for nucleosome formation (dark purple), as predicted by our thermodynamic model. Perfect boundaries are assumed to never be occluded by nucleosomes. Shown is the nucleosome formation probability at each base pair, $P_{covered}$, as a function of the distance from the boundary. The same probability is computed for a sequence with no boundaries (pink). (B) Probabilities for transcription factor binding and nucleosome formation as a function of the distance d of the site from the boundary, under a fixed transcription factor concentration (see Table 1). For comparison, P_{bound} and $P_{covered}$ values, under the same transcription factor concentration, for a corresponding binding site on a promoter with no boundaries (light green and dark green curves, respectively). (C) Shown are graphs of transcription factor binding probabilities, P_{bound} , at increasing transcription factor concentrations, for the promoters described in D, in which the transcription factor binding sites are located at different distances from a nucleosome boundary element. (D) Schematic illustrations for the promoters used in C and E and the associated binding affinities for nucleosomes and transcription factors. (E) Shown in color is the nucleosome formation probability at every base pair, $P_{covered}$, for each of the promoters in D. Shown in black is the probability of transcription factor binding, P_{bound} . For each promoter, we present $P_{covered}$ and P_{bound} values at three concentrations of the corresponding transcription factor, (10^{-3} , $10^{-1.5}$, and 10^0). We also illustrate the probable configuration of transcription factors and nucleosomes on these promoters, for each of the transcription factor concentrations, by employing a threshold (0.6) for both transcription factor binding and nucleosome formation. Note that in promoters (marked as 2 and 5) where the sites are relatively exposed at low transcription factor concentrations (owing to the effect of the boundary), the change in nucleosome occupancy in the base pairs surrounding the sites is less pronounced at higher transcription factor concentrations, compared to promoters (marked as 3 and 4) in which the sites were relatively occluded by nucleosomes at low transcription factor concentrations.

boundaries as activators was observed in the case of the yeast *HIS3* promoter (Iyer and Struhl 1995). When poly(dA:dT) elements were deleted from this promoter, there was a clear reduction in expression that could not be explained by binding of transcription factors to these elements. The activation observed was attributed to changes in nucleosome organization on the promoter. This positive effect was also observed *in vivo* upon the addition of a binding site for a transcription factor whose concentration is constant (Miller and Widom 2003). As mentioned in the beginning of this section, this case is virtually identical to the addition of an imperfect boundary for nucleosome formation. In this work, two foreign proteins (LexA, TetR) were shown to cooperatively

interact with Gcn4p to activate yeast genes when target sites for the foreign protein and for Gcn4p were placed at a short distance.

A more surprising theoretical effect, which can be easily observed in Figure 3C but which to the best of our knowledge has not been discussed in the literature, is the effect seen when the binding site is located at large distances from the boundaries, where there is increased probability for nucleosome formation (peaks of Fig. 3A). At these distances, the probability for transcription factor binding is reduced compared to a boundary-free landscape. This analysis suggests that boundary elements may also act as indirect repressors (in Fig. 3B, distances at which the red curve is below the light green curve), in addition to their

documented role as activators, where their exact role is determined by the distance between the transcription factor binding site and the boundary.

Both the repressive and the activating effects of boundaries can be observed when examining the effects of boundaries on the predicted promoter activation curves. As can be seen in Figure 3C, for a promoter with a single transcription factor site, the addition of a boundary element creates a shift effect (inflation effect in additive space) similar to the one created by a uniform landscape for nucleosome formation when compared to a nucleosome-free situation. However, unlike the case of the uniform landscape, in this case, the factor by which the concentration/affinity are multiplied can be either larger or smaller than 1 depending on the distance between the boundary and the transcription factor site (for details, see Supplemental material). As expected, when the site is located at a distance where there is decreased probability for nucleosome formation (troughs of Fig. 3A), the factor is <1 , causing the promoter to be activated at lower concentrations (a shift of the activation curve to the left), and by that acting as a general indirect activator, as mentioned above. Conversely, when the site is located at a distance where there is increased probability for nucleosome formation (peaks of Fig. 3A), the factor is >1 , causing the promoter to be activated at higher concentrations (a shift of the activation curve to the right, similar to the uniform nucleosome landscape case), and by that acts as a general indirect repressor. Additionally, there are cases in which the site is located at a distance where the probability for nucleosome formation is the same as in a promoter without a boundary (intersection points in Fig. 3A). In these cases, the factor is equal to 1, and the activation curve is identical to that of the boundaryless promoter (in Fig. 3B, distances at which the red curve intersects the light green curve).

As stated in the case of a uniform landscape for nucleosome formation, a shift effect is equivalent to changing the affinity of the transcription factor to its site. Thus, this observation suggests that the same activation curve can, in fact, be generated by at least two distinct promoters, one containing a boundary for nucleosome formation located at some distance from a transcription factor site, and one where there is no boundary element but the affinity of the transcription factor to its site is different (multiplied by the above mentioned factor) (for details, see Supplemental material). This can be useful when the affinity of a transcription factor to its site cannot be easily modulated, for example, for transcription factors that are not able to physically bind DNA with a wide range of different affinities. In these cases, a boundary element can serve as a means to fine-tune the promoter activation curve, allowing a significant increase in the possible activation behaviors that can be reached.

Finally, using our framework, we can predict the effect of increasing transcription factor concentration not only on the activation curves but also on the changes to nucleosome organization along the promoter. When examining the actual effect of increasing transcription factor concentrations on the probability ($P_{covered}$) of each base pair to be covered by nucleosomes (Fig. 3D,E) in promoters where a boundary is located at various distances from a transcription factor site, several observations can be made. First, we see that a boundary generates significant troughs in the $P_{covered}$ graph even when no transcription factor is present. In sequences where the sites are located in a region protected by the boundary element, the trough resulting from increasing concentrations of the transcription factor will slightly broaden or deepen the troughs that are already generated by the boundary (Fig. 3E, purple and red graphs). However, when the sites are located at

regions of high nucleosome coverage (Fig. 3E, blue and brown graphs), high transcription factor concentrations will generate a new trough in a previously peaked region, thereby changing the nucleosome occupancy in the area in a more dramatic manner. These results also make clear testable hypotheses regarding the changes in nucleosome positioning that may be expected at different transcription factor concentrations.

From the analysis above, it is clear that even simple non-uniform landscapes in the form of boundaries, which are present at many natural promoters, can have a significant effect on transcription factor binding. In addition to generating independence between binding events on different sides of the boundary, boundaries also affect the binding probability of a transcription factor to its cognate site. This effect can be either repressive or activating, depending on the distance of the site from the boundary. In general, the addition of a boundary element to a promoter with a single transcription factor binding site causes a shift (in log scale) in the promoter activation curve, thus allowing the generation of a diverse range of curves, simply by modifying the distance between the site and the boundary (Fig. 3C). This result suggests a mechanism by which the same transcription factor may use the same site at various promoter locations to regulate different subsets of its targets with different activation curves.

Addition of binding sites for transcription factors causes a change in the shape of the activation curve

In this section, we address the possible effects of adding another type of element to a simple promoter with a single transcription factor binding site—an additional binding site for a transcription factor whose concentration is not constant, but rather is changing as some function of the concentration of the transcription factor binding the original site. The addition of such an element creates a new peak in the affinity landscape for transcription factor binding in the location along the promoter where the site is added. For simplicity, throughout most of this section, we assume that the added site is an additional site for the same transcription factor, but the observations we discuss are also valid for a site for a different transcription factor whose concentration is not constant.

The addition of a binding site, which forms a two-site promoter architecture, is of particular interest in the presence of nucleosomes, since it has already been discussed in the literature. Owing to the abundance of nucleosomes in eukaryotic cells and the fact that a long stretch of generally inaccessible DNA is wrapped around the histone core of each nucleosome, it was previously hypothesized that nucleosomes can generate widespread obligate cooperative binding interactions between transcription factors. The idea was that a transcription factor that is able to displace a nucleosome and bind somewhere in the DNA previously occupied by that nucleosome frees up adjacent sites for binding of additional transcription factors, and by that positively affects their binding probabilities (Adams and Workman 1995; Polach and Widom 1996; Miller and Widom 2003). As mentioned in the previous section, this type of obligate cooperativity was already observed in vivo upon the addition of a site for a transcription factor whose concentration is constant (Miller and Widom 2003). Here we will examine the case of two binding sites for the same transcription factor or for related transcription factors, both with regard to the generation of obligate cooperativity as well as other possible effects of the added site on the binding probability to the original site.

In many aspects, the case of an additional binding site is similar to the case of addition of boundary elements that was described above. Under any fixed concentration of the cognate transcription factor, the addition of such a site is identical to the addition of a boundary element, and therefore we observe the same periodic effect that was observed in the boundary case. Thus, the effect of the new binding site on the probability of binding to the original site can again be either positive or negative depending on the distance between the sites, exactly as in Figure 3B. Hence, the hypothesized obligate cooperativity mentioned above is, in fact, what we described as the predicted positive effect when the two sites are placed at distances such that there is decreased probability for nucleosome formation (troughs in Fig. 3A). However, similar to the boundary case above, here too, we predict that in addition to the hypothesized cooperative effect, there should also exist a destructive effect between binding sites when placed at distances such that there is increased probability for nucleosome formation (peaks in Fig. 3A). To the best of our knowledge, this effect, like the predicted negative effect of the boundary, has not been discussed in the literature.

Both the cooperative and the destructive effects can be observed when plotting the probabilities of transcription factor binding to its site, on a promoter that contains an additional site for a different transcription factor, under a wide range of possible concentrations for the two transcription factors (Fig. 4). When the sites are located at a short distance of 10 bp apart, a cooperative effect can be observed. In most of the range of transcription factor concentrations, the most prevalent configuration is one where both sites are occupied, even when the concentration of one of the transcription factors is relatively low. This is due to the positive

effect on the binding probability of one transcription factor created when the other transcription factor binds to its site. The destructive effect can be observed when the two sites are located at a distance of 135 bp away. Here we see that the configuration where both sites are occupied is highly unlikely and occurs only at extremely high concentrations of both transcription factors. This is because the binding of one transcription factor to its site negatively affects the probability of the other transcription factor to bind its site. In this case, we can observe a different nonmonotonic effect: even when the concentration of one transcription factor is increased, its binding probability can decrease, if the concentration for the other factor is also increased at the same time, creating a stronger destructive effect between the sites.

The suggested mechanism for a nucleosome-induced obligate cooperative/destructive effect is appealing since it is general and does not require specialized evolution of proteins to allow for complex protein–protein interactions. This mechanism also suggests a function for the recent observations regarding abundant weak binding (Tanay 2006; Li et al. 2008), since weak binding sites are more sensitive to cooperative effects, as can be seen in Figure 5D–F. Other experimental observations that can potentially be explained by this obligate cooperativity are related to transcription factors with dual repressor–activator activities (Rubin-Bejerano et al. 1996; Ma 2005). Although we found no experimental evidence for this mechanism, a transcription factor that normally operates as a repressor (activator) can seem to act as an activator (repressor) if it promotes the binding of a nearby activator (repressor) by displacing a nucleosome that covers both sites. Still, most documented cases of transcription factor binding cooperativity like the Gal4p case (Giniger and Ptashne 1988) are usually

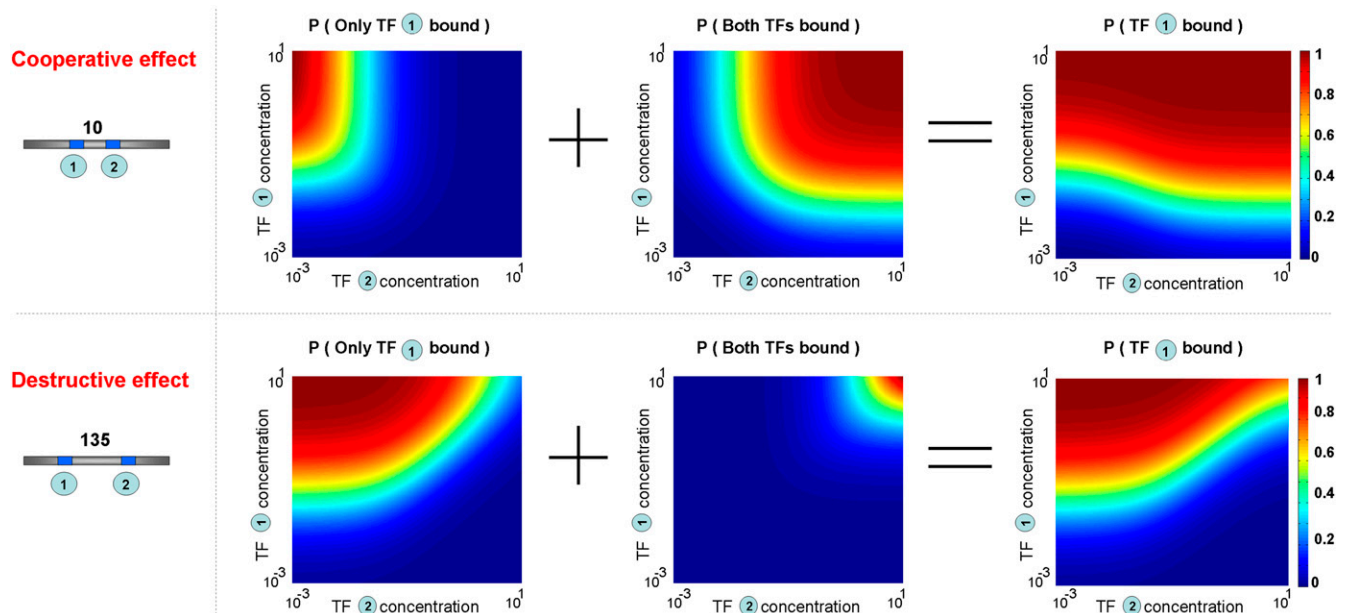


Figure 4. Obligate cooperative/destructive effects between transcription factor sites. For two promoters containing two sites for different transcription factors (marked as transcription factors 1 and 2) at different distances (10/135 bp), shown is the probability of binding for transcription factor 1 to its site (*left* site) for increasing concentrations of both transcription factors (*right* heat maps). This probability is, in fact, the sum of two probabilities: the probability that only transcription factor 1 is bound to its site while transcription factor 2 is not bound to its site (*left* heat maps), and the probability that both sites are occupied (*middle* heat maps). Note that when the sites are located at a distance of 10 bp from each other (*top* heat maps), a cooperative effect is observed—the binding of one transcription factor positively affects the probability of the other transcription factor to be bound, and thus the most prevalent configuration is one where both sites are occupied, even when the concentration of one of the transcription factors is relatively low. However, when the distance between the sites is 135 (*bottom* heat maps), a destructive effect is observed. The binding of one transcription factor negatively affects the probability of the other transcription factor to be bound, and thus the configuration where the two sites are occupied is rare and occurs only at extreme concentrations of both factors.

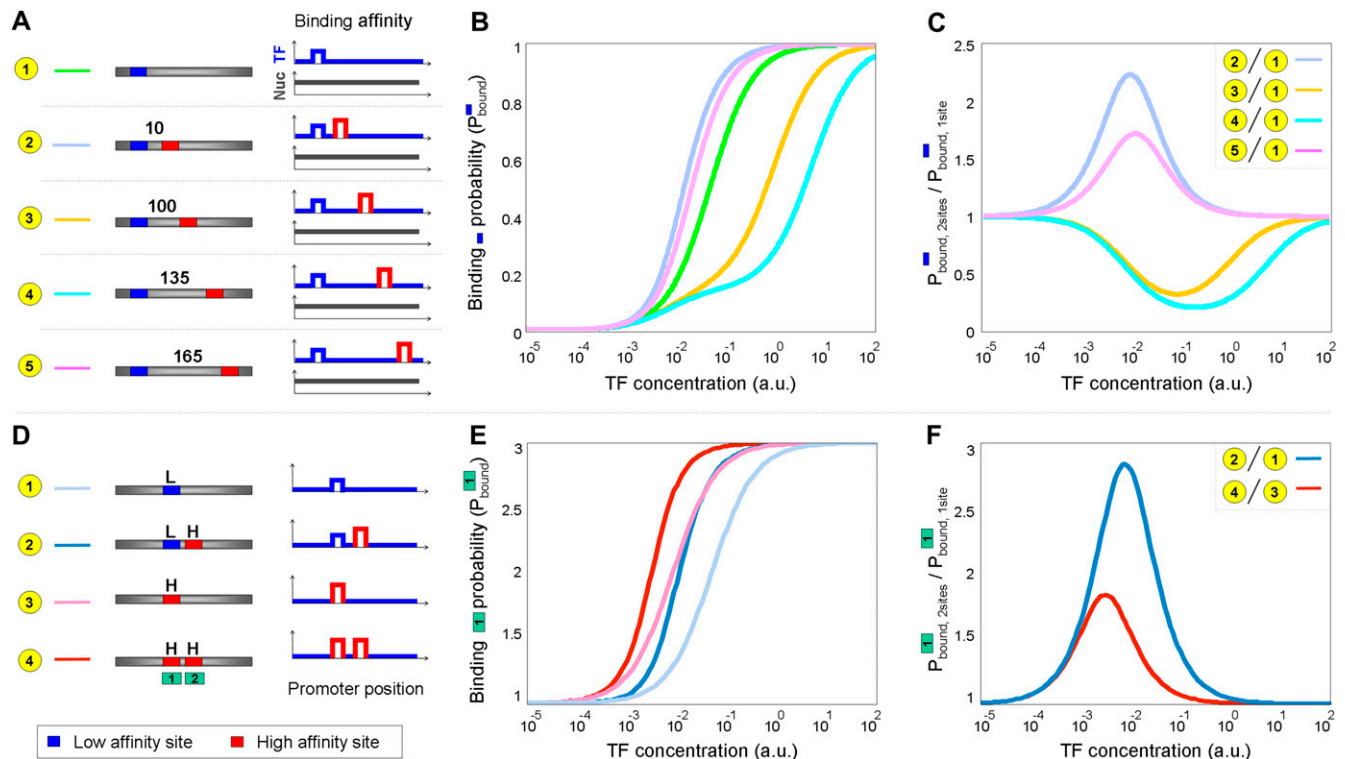


Figure 5. Addition of a transcription factor site to a simple promoter and its effect on the P_{bound} graph. (A) Illustrations of the promoters considered and the associated binding affinities for nucleosomes and transcription factors. (B) Shown are transcription factor binding probabilities to a low-affinity binding site for a simple promoter with a single low-affinity site (green), and for promoters with an additional high-affinity site located at various distances from the low-affinity site at increasing transcription factor concentrations. The addition of a transcription factor binding site can have both a cooperative effect (light purple and pink curves) and a destructive effect (yellow and cyan curves). Note that the addition of a transcription factor site results in a change in the shape of the activation curve. (C) Shown is the ratio between the probability of transcription factor binding to the low affinity site in a promoter with two transcription factor sites ($P_{bound,2sites}$) and in a promoter with one transcription factor site ($P_{bound,1site}$) at increasing transcription factor concentrations. This ratio can be viewed as the strength of the cooperative/destructive binding effect between the transcription factors (see main text). The ratio obtained changes as concentration increases and can be larger or smaller than one indicating a cooperative/destructive effect, respectively. (D–F) Examination of the cooperative effect generated by adding an adjacent high-affinity site to a promoter with a single, either low- or high-affinity, transcription factor site. (D) Illustrations of the promoters considered and the associated binding affinities for nucleosomes and transcription factors. (E) Transcription factor binding probability graphs for promoters with one high-affinity site (pink curve), two high-affinity sites (red curve), one low-affinity site (light blue), or one low-affinity and one high-affinity site (blue). The energetic contribution ($F(TF, P_{site}, P_{site} + L_{TF})$) from binding is set to 1200 for a high-affinity site, and 200 for a low-affinity site. (F) Shown is the ratio between the probability of transcription factor binding to the left site (marked as site 1) in a promoter with two transcription factor sites ($P_{bound,2sites}$) and in a promoter with one transcription factor site ($P_{bound,1site}$) at increasing transcription factor concentrations. Note that the cooperative effect between a low-affinity site and a high-affinity site is larger than the cooperative effect between two high-affinity sites.

attributed to protein–protein interactions. Thus, it remains to be seen whether the suggested mechanism is, indeed, a prevalent mechanism for cooperative/destructive binding *in vivo*.

Thus far, we have seen several aspects in which the addition of a binding site exhibits the same effect as the addition of a boundary element. Nevertheless, the two cases are not identical. The difference between the above boundary case and the case of adding a transcription factor binding site becomes apparent when we carefully examine the effect on the resulting activation curve. As in the addition of a boundary, the addition of a site changes the location of the activation curve. (Cooperative effects result in a curve to the left of the single-site curve, while a destructive effect results in a curve to the right of the single-site curve [see Fig. 5B].) However, this is no longer a simple shift effect but rather a more complex effect that changes the shape and the steepness of the activation curve (for details, see Supplemental material). Although the binding of a transcription factor to one of its sites effectively forms a boundary, since the concentration of the transcription factor is not constant, the strength of the boundary formed is not

constant. Thus, at each concentration, the strength of the effect on the binding probability is different—inducing a shape change in the P_{bound} graph (Fig. 5B). In Figure 5B, we see the predicted activation curves resulting from different promoter architectures, either with a single site (green curve) or with an additional site at various distances, in the presence of a uniform landscape for nucleosome formation. The ratio between the binding probability in a single-site architecture and the binding probability in a two-site architecture, for any given concentration, can be viewed as the cooperative/destructive effect (Fig. 5), that is, the contribution made by the presence of one site on the probability for binding to another site. As can be seen, the strength of the observed effect changes with concentration and, as stated above, can be both positive (ratio >1) or negative (ratio <1), depending on the distance between the sites.

In summary, it is clear that nucleosomes generate not only an obligate cooperative effect but also a more surprising destructive effect between transcription factors depending on the distance between the sites, as in the simple case of boundary addition.

However, unlike the addition of a boundary element, the addition of a transcription factor site also induces a shape change in the activation curve, creating a wide range of possible promoter activation curves.

Combinations of promoter elements can produce a diverse range of activation curves

In the above section, we discussed the effects of adding a boundary element or a transcription factor binding site to a simple promoter containing a single site for a transcription factor in the presence of nucleosomes. However, native promoters often contain a combination of several such elements. In this section, we consider the effects of such combinations on promoter activation curves.

Recall that in the case of the simple single-site promoter, the addition of a boundary element was associated with a simple shift effect (in log scale), while the addition of a transcription factor is associated with a change in the shape and steepness of the activation curve. However, in more complex promoters, which contain more than one transcription factor site, both the addition of a boundary and the addition of a transcription factor can induce a change in the shape and steepness of the activation curve (for details, see Supplemental material).

Examples for activation curves resulting from such combinations of promoter elements and the strength of cooperative/destructive effects formed between the transcription factor binding sites can be seen in Figure 6, where we consider promoters containing two adjacent sites forming a cooperative interaction. When we add to these promoters one or two boundaries at various distances from the sites, we can enhance or reduce the strength of their cooperative interaction. Note the architecture in which the sites are located between two boundaries (black curve), resulting in a significantly sharper, almost digital, activation curve. Thus, a combination of a few promoter elements is sufficient in order to create a wide range of activation curves.

It is important to note that using our modeling framework, we are able to efficiently produce the predicted activation curve for any given promoter architecture. Moreover, it enables us to examine analytically the nature of changes to the activation curve when comparing any two promoter architectures. It should be noted that such effects can also be computed simply by describing the transition between the two promoter architectures in terms of single promoter element addition/deletion steps, and serially composing the effect of such steps to form the overall effect (for details, see Supplemental material).

To conclude, from our theoretical framework it follows that, indeed, nucleosomes may have significant effects on transcription regulation, since the mere presence of nucleosomes enables the production of a wide range of possible promoter activation curves using even a small number of promoter elements, where each element is either a boundary element for nucleosome formation or a binding site for a transcription factor.

Nucleosomes as generators of distinct noise behaviors

In recent years, there has been surging interest in understanding the causes of cell-to-cell variability, or transcriptional noise, in gene expression levels among cells from the same population (Elowitz et al. 2002; Raser and O'Shea 2004). Such expression noise is speculated to be beneficial when phenotypic diversity is required to allow subpopulations to have increased fitness under unexpected environmental changes.

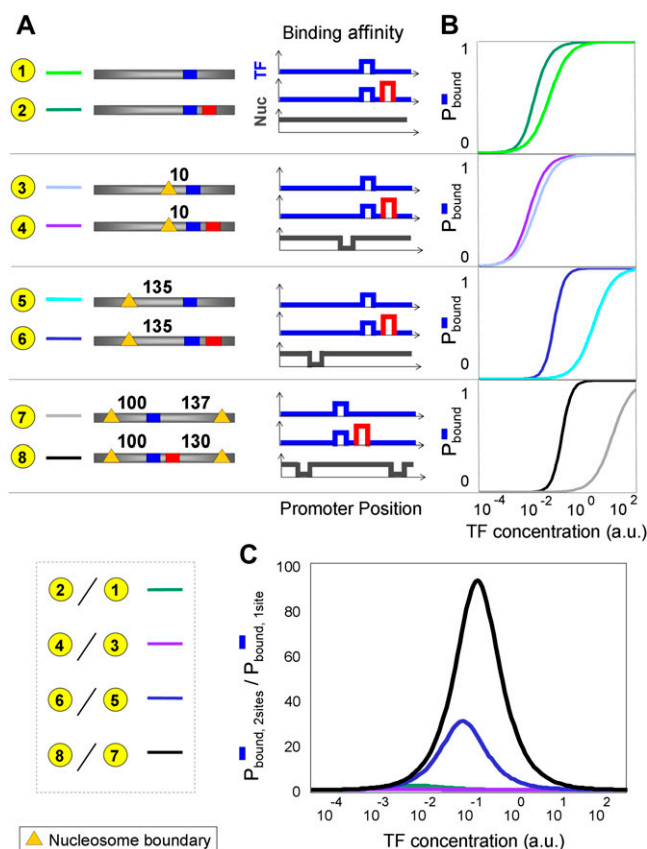


Figure 6. Combination of promoter elements produces a diverse range of activation curves. (A) Schematic illustrations for the promoters used in this figure and the associated binding affinities for nucleosomes and transcription factors. (B) Probability of transcription factor binding to a low-affinity site at increasing transcription factor concentrations for promoters with this low-affinity site as a single site (light colors, $P_{bound, 1site}$) or for promoters with an additional high-affinity site for the same transcription factor (dark colors, $P_{bound, 2sites}$). The distance between the two binding sites is set to 1 bp. The sites are located on the promoters at varying distances from one or two boundary elements for nucleosome formation. The energetic contribution ($F_{(TF, P_{site}, P_{site} + L_{TF})}$) from binding is set to 1200 for a high-affinity site, and 200 for a low-affinity site. (C) The ratio between $P_{bound, 2sites}$ to $P_{bound, 1site}$ from B at increasing transcription factor concentrations. This ratio represents the strength of the cooperative/destructive binding effect between the transcription factors. The ratio obtained is >1 , indicating a positive cooperative effect. However, the strength of the effect depends on the locations of the sites relative to the boundary: the ratio is higher for sites that are relatively covered by nucleosomes at low transcription factor concentrations.

Recently, several papers have implicated nucleosome-free regions (Tirosh and Barkai 2008) in general and sequences that form boundaries for nucleosome formation in particular (Field et al. 2008), with low expression noise. Here we examine the theoretical roles of boundaries and transcription factor sites on transcriptional noise.

Recall that the value of P_{bound} reflects the probability of a site in a single cell to be occupied. Therefore, a population of cells with an extreme P_{bound} value corresponds to low expression noise, since such a value implies that most cells in the population are likely to share the same state of transcription factor binding (either bound or unbound), and are therefore likely to share similar expression values. On the other hand, intermediate P_{bound} values correspond to high expression noise, since in this case, the population is

heterogeneous, where some cells exhibit one state of transcription factor binding, and others exhibit a different state (Fig. 7).

As stated in the above sections, assuming a multiplicative increase in transcription factor concentration, the effect of adding a boundary element to a promoter with a single site for a transcription factor is a shift in the activation curve, where the shape of the activation curve (P_{bound} graph) remains unchanged. Thus, in this case, the range of transcription factor concentrations responsible for “noisy” P_{bound} values is altered (shifted in some direction), but its extent remains constant (Fig. 7A). Therefore, we predict that boundaries are not necessarily linked with low expression noise. The observed effect on transcriptional noise is, according to our model, dependent on the physiological range of transcription factor concentrations. If the boundary resulted in a shift toward activation at transcription factor concentrations lower/higher than the physiological concentration, we will observe a fully activated or fully repressed, and therefore not “noisy” promoter at the native concentrations. If, however, the boundary caused a shift such that the native transcription factor concentration is responsible for an intermediate P_{bound} value, the promoter will appear to be “noisy.” Specifically, in the common case in which transcription factors regulate both promoters with a boundary (adjacent to the binding site) and promoters without a boundary, the general association between boundaries and low noise (Field et al. 2008) may be explained if we assume that both types of promoters are activated under the physiological concentration of the transcription factor to a reasonable extent. In these cases, owing to the shift in activation curves, promoters containing a boundary near a transcription factor binding site will exhibit higher P_{bound} values that are likely to be beyond the “noisy” range, and thus, such promoters will appear to be less noisy.

Conversely, the addition of a transcription factor site to a simple promoter with a single transcription factor, or the addition of either a boundary element or a transcription factor site to a promoter containing more than a single transcription factor site, results in a different effect on expression noise. As seen in the above sections, such additions result in a change to the shape and steepness of the activation curve. Therefore, both the location and the extent of the range of concentrations in which the promoter is “noisy” can be altered depending on the strength of cooperative/destructive effects between sites and positive/negative effects induced by a boundary (Fig. 7B,C).

Note, however, that the above analysis is valid only if we assume a multiplicative increase in the concentration of the activating transcription factor. If the increase is additive instead of multiplicative, then even for the simple case of adding a boundary to a promoter with a single transcription factor site, we get an inflation effect that changes the extent of the range of transcription factor concentrations responsible for “noisy” P_{bound} values, making it either larger or smaller, depending on whether the factor by which the concentration is multiplied is smaller or larger than 1.

The actual “noise” behavior of native promoters thus depends not only on their architecture, but also on the physiological concentrations of the transcription factors that regulate them. While the addition of boundary elements causes—assuming a multiplicative increase in transcription factor concentration—a mere shift in the range of transcription factor concentrations responsible for intermediate, “noisy” values of P_{bound} , the addition of other transcription factor sites changes both the location and the extent of this range, making the promoter “noisy” over a wider/narrower range of concentrations.

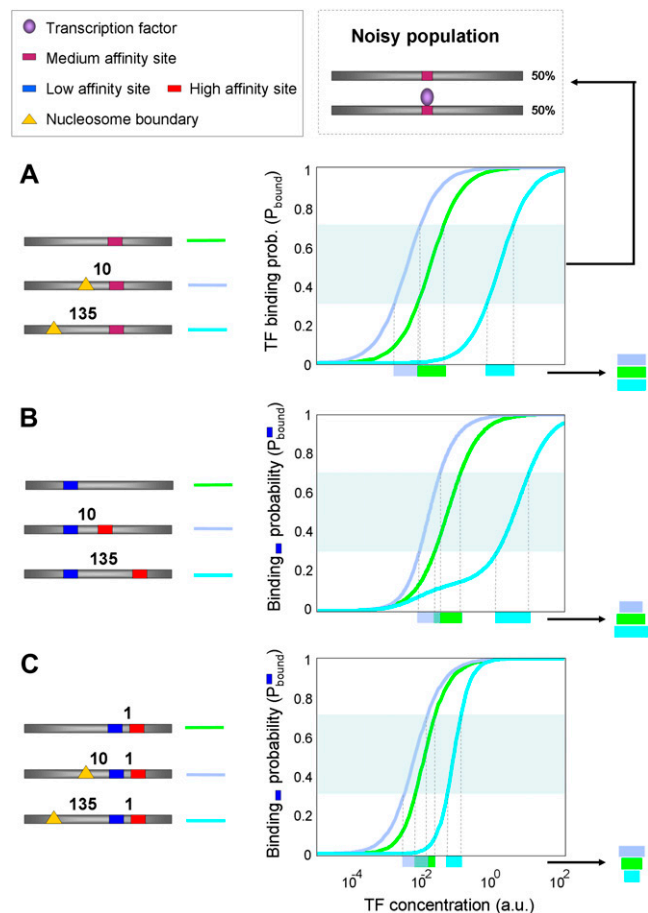


Figure 7. Nucleosomes as generators of distinct noise behaviors. Shown is the “noisy” regime of intermediate P_{bound} values (defined as 0.3–0.7) for the low-affinity (left) site in different promoters. These intermediate P_{bound} values correspond to a heterogeneous cell population where some cells exhibit one state of transcription factor binding, and others exhibit a different state. (A) Promoters (from Fig. 3C) with a single transcription factor site and a single boundary element, located at various distances from the site. Note that in all promoters the extent of the range of transcription factor concentrations responsible for “noisy” P_{bound} values is the same, but its location changes. (B) Promoters (from Fig. 5B) with high- and low-affinity sites for some transcription factor. Sites are located at various distances from each other. Note that both the location and the extent of the range of transcription factor concentrations responsible for “noisy” P_{bound} values are altered between promoters. When the sites are close together (10 bp), a cooperative effect is formed, and the extent of the range of concentrations responsible for the “noisy” regime is decreased. Conversely, when the sites are 135 bp apart, a destructive effect is observed, and the extent of the range of concentrations responsible for the noisy regime is increased. (C) Promoters (from Fig. 6B) with a high- and a low-affinity site for some transcription factor, located at various distances from a boundary element. The distance between the sites is set to 1 bp. Here, the boundary addition results in a change to both the location and the extent of the range of transcription factor concentrations responsible for “noisy” P_{bound} values. When the sites are close to the boundary (10 bp), the probability of a nucleosome covering the sites and generating a strong obligate cooperative effect is increased, and thus the extent of the range of concentrations responsible for the noisy regime is increased. Conversely, when the sites are far from the boundary (135 bp), the probability of a nucleosome covering the sites and generating a strong obligate cooperative effect is increased, and, thus, the extent of the range of concentrations responsible for the noisy regime is decreased.

Applicability to real promoters and binding affinities

As discussed above, we mostly consider simplified promoters and binding affinities. Using such controlled affinity landscapes allows us to analytically explore and characterize the effects of basic promoter elements and to make general observations regarding the activation curves of promoters that are composed of such elements. However, as we now show, our modeling framework is not restricted to such simplified affinities and can be readily applied to real sequences and binding affinity landscapes, thereby attempting to address the challenging task of predicting expression kinetics directly from promoter sequence.

Before we present our results on real sequences and affinity landscapes, we note that this task is particularly challenging for several reasons. First, our models rely on the ability to generate accurate affinity landscapes for each sequence using the binding sequence preferences of transcription factors and nucleosomes, but comprehensive models of such binding preferences have only recently become available, and there are still some differences between different such published preferences. Second, in addition to transcription factors and nucleosomes, there are other factors that we do not model such as RNA polymerase binding or the action of chromatin remodelers, which most likely influence the expression kinetics of real promoters. Finally, our model uses the binding occupancy of transcription factors on promoters as a proxy for expression levels, but the function that translates binding events into expression levels in living cells is likely more complex.

Despite these remaining challenges, we demonstrate below that our current modeling framework can successfully predict, from sequence alone, the expression behavior measured in two independent experimental studies (Iyer and Struhl 1995; Lam et al. 2008).

As a first application of our framework to real sequences, we return to the set of *PHO5* promoter variants (Lam et al. 2008) described above. Instead of the simplified affinity landscapes that we used above, we now use real binding preferences for both nucleosomes (Kaplan et al. 2009) and for Pho4p (Lam et al. 2008) to generate binding affinity landscapes for each sequence variant, and reapply our model to these real affinity landscapes. The use of real affinity landscapes also allows us to include two additional variants from the paper of Lam et al. (L2 and H2) that we previously omitted, since when using simplified affinity landscapes, they were symmetric to another variant (wild-type and H3 variants, respectively) and hence identical in this simplified affinity space.

Examining the activation curves predicted by our model using these real affinities, we find that as in the case of the simplified affinity landscapes, here too our model is able to distinguish between variants with high- and low-affinity exposed sites, thereby predicting the experimentally measured early activation of promoters with high-affinity exposed sites (Fig. 8B). That is, our framework correctly models the effect of the differences in the derived affinity landscapes on the competition between Pho4p and nucleosomes and translates the results of this competition into predictions of expression kinetics in a manner that matches the measured activation curves.

We also apply our framework to a set of variants constructed for the *HIS3* promoter (Iyer and Struhl 1995). Since the native *HIS3* promoter contains a poly(dA:dT) element, the use of this set allows us to test on real promoters, the effect on expression of nucleosome-disfavoring sequences that provide a boundary for nucleo-

some formation. In their paper, Iyer and Struhl show that variants of the *HIS3* promoter mutated to have longer poly(dA:dT) sequences exhibit stronger levels of expression. Their results further suggest that the effect of these poly(dA:dT) elements on the promoters is mediated through their effect on the nucleosome organization. Using real affinities for nucleosomes (Kaplan et al. 2009) and for the Gcn4p (MacIsaac et al. 2006) transcription factor, which targets the *HIS3* promoter, we find that our model indeed predicts higher levels of Gcn4p occupancy and thus higher levels of expression for variants containing longer tracts of poly(dA:dT) elements, thereby matching the experimental observations of Iyer and Struhl (Fig. 8E).

To summarize, we show that despite the challenge of predicting expression kinetics for real promoters using real molecule affinities, our model is able to match the experimentally measured expression behavior of two independent data sets, using real sequences and binding affinities for transcription factors and nucleosomes. These results further underscore the importance of incorporating nucleosomes into models of transcriptional regulation.

Discussion

The construction of quantitative models for transcription regulation represents one of the next great challenges in molecular biology. To build such models, we need to characterize the transcriptional components that are involved and learn how these components combine to produce expression patterns. Recently, it has become clear that chromatin is one such component, which plays a possibly crucial role in transcriptional processes. However, a quantitative understanding of the nature of its effects is still missing.

Here, by extending existing thermodynamic models to include both transcription factors and nucleosomes, we are able, for the first time, to theoretically explore possible roles and consequences that binding affinity landscapes of nucleosomes may have on transcriptional regulation. Through our framework, which generates a distribution over all possible configurations by combining the binding affinities and concentrations of bound transcription factors and nucleosomes in each configuration, we are able to distinguish between the effects of different promoter elements on the resulting promoter activation curves, and to provide several insights regarding the interplay between nucleosomes and transcription factors.

These insights advance our understanding of the effects of nucleosome positioning on transcription. Some of them suggest mechanistic explanations for experimental observations, such as the role of nucleosomes as general repressors and as generators of obligate cooperativity between transcription factors, while others offer several new testable hypotheses. These hypotheses include the possible role of different nucleosome landscapes in generating diverse promoter activation curves, and in determining cell-to-cell expression variability. Other hypotheses that result from our framework include the role of sequences that form nucleosome boundaries as activators or repressors of transcription, the obligate cooperative/destructive effects between transcription factor binding sites, which can induce a change to the shape of the activation curves, and the ability of nucleosomes to generate dual activator/repressor transcription factor behaviors.

Currently, most existing data regarding these issues are anecdotal or of low resolution and thus do not allow for a systematic testing of these hypotheses. As increasing amounts of high-quality data targeted at these questions (such as high-resolution data for

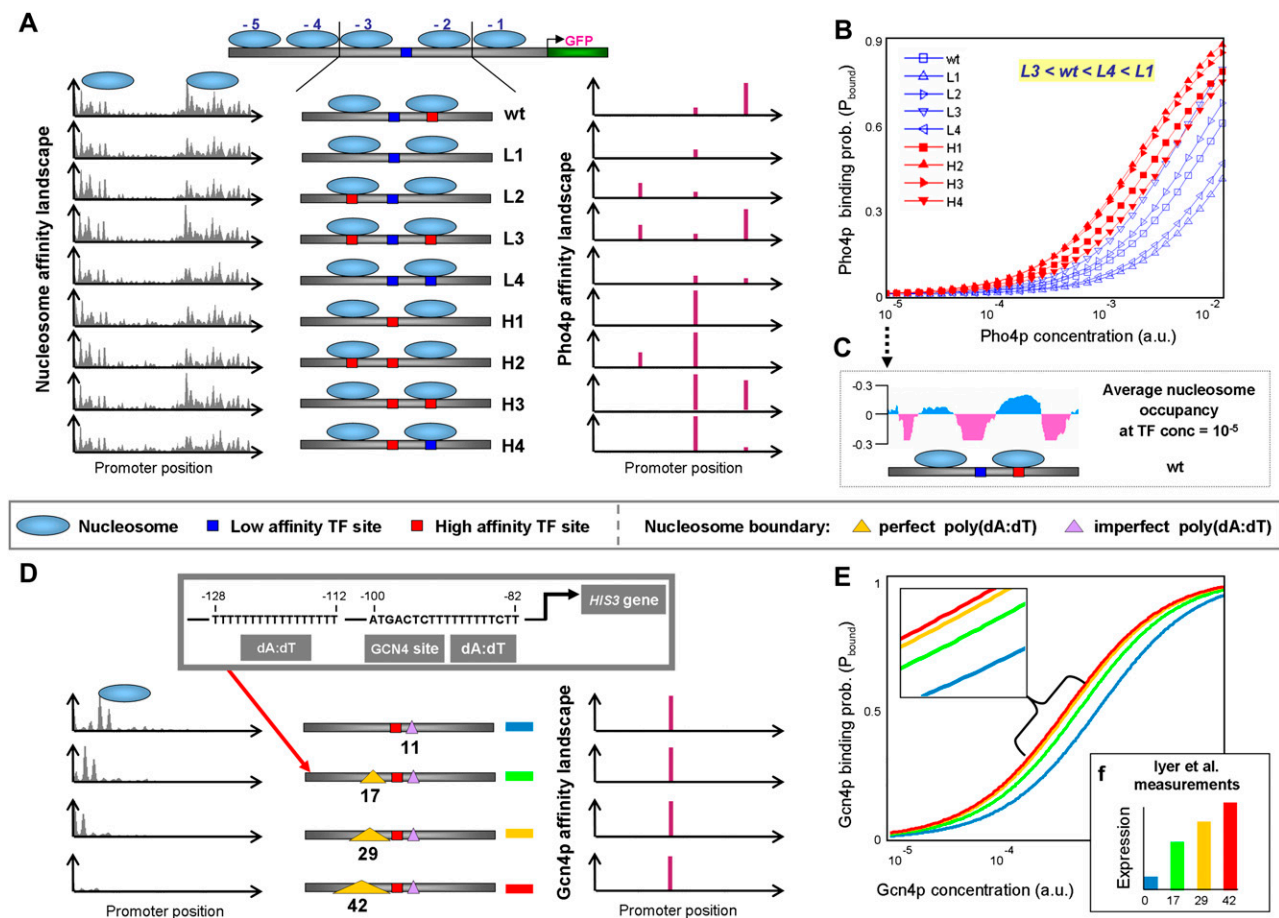


Figure 8. Our model predicts expression behavior of real promoter sequences. (A) Shown are real binding affinity landscapes for each of the *PHO5* promoter variants from Lam et al. (2008) (800 bp upstream of the *PHO5* gene) (see description in Fig. 2A), generated using the real binding preferences for Pho4p (Lam et al. 2008) and for nucleosomes (Kaplan et al. 2009). For each base pair, the binding affinity landscape displays the likelihood for a bound transcription factor or nucleosome to start at that sequence position. Note that since real sequences are used, unlike in Figure 2, variants L2 and H2 are no longer identical to variants wild type and H3 (respectively), and are therefore included in the analysis. (B) Model prediction of the binding probability of Pho4p, P_{bound} , generated by applying our thermodynamic framework to each of the promoter variants from A at increasing Pho4p concentrations. The yellow inset indicates the predicted relative ordering of the onset times of the promoter variants with exposed low-affinity sites. Note that here, too, similar to the case when using simplified promoters (Fig. 2), the order predicted is identical to the one measured (see Fig. 2B). (C) Shown is the predicted average nucleosome occupancy (log occupancy divided by the median) at low (10^{-5}) Pho4p concentration for each position along the sequence of the wild-type variant of *PHO5*. (D) Same as A, but for *HIS3* promoter variants constructed by Iyer and Struhl (1995), in which a native 17-bp nonperfect poly(dA:dT) element was either deleted or replaced by a perfect poly(dA:dT) element of length 17, 29, or 42 bp. For each variant, shown are real binding affinity landscapes (intergenic region between the *HIS3* gene and the upstream gene *MRM1* [also known as *PET56*]), generated using the real binding preferences for Gcn4p (MacIsaac et al. 2006) and for nucleosomes (Kaplan et al. 2009). (E) Model prediction of the binding probability of Gcn4p, P_{bound} , generated by applying our thermodynamic framework to each of the promoter variants from D at increasing Gcn4p concentrations. (F) Expression value measurements by Iyer and Struhl (1995) for each of the promoter variants from D.

noise and expression curves and nucleosome positioning data at various time points along promoter activation curves) are generated, these hypotheses could easily be tested. One promising direction for testing these hypotheses is studying the activation dynamics of synthetic promoter variants containing different combinations of the promoter elements discussed in this study (boundary elements and transcription factor sites). Such large-scale libraries aimed at exploring the effects of transcription factor binding on expression have already started to emerge (Ligr et al. 2006; Gertz et al. 2009). However, with the wealth of new hypotheses gained by this work, more elaborate variant libraries aimed also at studying the effects of nucleosomes on transcription can be designed.

In this study, we applied our model mostly to abstract promoters with relatively simple binding affinity landscapes for nu-

cleosomes and transcription factors. This allowed us to systematically explore the effects of nucleosomes on transcription by focusing on the effects of each promoter element separately. However, as we show above, our framework can also be used to address the challenging task of predicting expression from real promoter sequences. By applying our model to promoter sets from two independent studies, we demonstrate its ability to recapitulate experimentally measured expression using only binding preferences for transcription factors and nucleosomes, thereby underscoring the importance of the interplay between nucleosomes and transcription factors in determining expression kinetics.

Although the above successes in predicting expression from sequence are encouraging, our present model is still far from fully addressing this challenging task. Nonetheless, we are confident that our modeling approach and the insights gained from our

analysis will be useful for developing future models for transcriptional control. As a first step in this direction, we must alleviate some of the simplifying assumptions that we made in order to facilitate our theoretical examination. This can easily be done for most of our modeling assumptions. For example, the displacement model for nucleosomes can be replaced by a possibly more realistic model, according to which, in addition to displacement of nucleosomes, partial and transient unwrapping of some of the 147 bp wrapped around the histone core can occur. This would allow transcription factors to access their sites while the histone core remains bound to the remaining base pairs (Polach and Widom 1995). This change can be done by allowing the histone core to bind less than L_N ($= 147$) bp with some energetic cost paid for unwrapping several base pairs. The exposed base pairs can then be bound by transcription factors. Note that such an alternative model, which would model both nucleosome displacement and partial unwrapping, would still predict the cooperative and destructive effects between proximally bound transcription factors that we presented above, although the precise details at which such interactions occur might change.

Modeling the transition from binding to expression is also possible and requires the definition of such a transition function. An example for a possible transition function can be seen in Segal et al. (2008). Explicit cooperative effects that represent protein-protein interactions can also be added on top of the obligate cooperativity stemming from competition with nucleosomes. See a possible implementation in Segal et al. (2008). Finally, to attain a satisfactory quantitative, mechanistic model and to truly understand transcriptional processes, other components besides nucleosomes that are currently insufficiently characterized, such as RNA polymerase, the basal transcription machinery, and chromatin remodelers, must be properly incorporated into transcriptional models. Nevertheless, the incorporation of nucleosomes in transcriptional models is now within reach, representing a step forward toward a quantitative understanding of transcription and toward predicting expression patterns from DNA sequences.

URLs

Additional information including the source code for the implementation of our model can be found at http://genie.weizmann.ac.il/pubs/tf_nuc_model09/.

Acknowledgments

We thank Noam Vardi and Jon Widom for useful discussions. This work was supported by grants from the European Research Council (ERC) and Israel Science Foundation (ISF) to E.S. T.R.S. thanks the Azrieli Foundation for the award of an Azrieli Fellowship. E.S. is the incumbent of the Soretta and Henry Shapiro career development chair.

References

- Adams CC, Workman JL. 1995. Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol Cell Biol* **15**: 1405–1421.
- Anderson JD, Widom J. 2001. Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. *Mol Cell Biol* **21**: 3830–3839.
- Bernstein BE, Liu CL, Humphrey EL, Perlstein EO, Schreiber SL. 2004. Global nucleosome occupancy in yeast. *Genome Biol* **5**: R62. doi: 10.1186/gb-2004-5-9-r62.
- Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Kuhlman T, Phillips R. 2005. Transcriptional regulation by the numbers: Applications. *Curr Opin Genet Dev* **15**: 125–135.
- Buchler NE, Gerland U, Hwa T. 2003. On schemes of combinatorial transcription logic. *Proc Natl Acad Sci* **100**: 5136–5141.
- Bulyk ML, Huang X, Choo Y, Church GM. 2001. Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proc Natl Acad Sci* **98**: 7158–7163.
- Cairns BR. 2005. Chromatin remodeling complexes: Strength in diversity, precision through specialization. *Curr Opin Genet Dev* **15**: 185–190.
- Dechering KJ, Cuelenaere K, Konings RN, Leunissen JA. 1998. Distinct frequency-distributions of homopolymeric DNA tracts in different genomes. *Nucleic Acids Res* **26**: 4056–4062.
- Elowitz MB, Levine AJ, Siggia ED, Swain PS. 2002. Stochastic gene expression in a single cell. *Science* **297**: 1183–1186.
- Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, Lubling Y, Widom J, Segal E. 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput Biol* **4**: e1000216. doi: 10.1371/journal.pcbi.1000216.
- Gertz J, Siggia ED, Cohen BA. 2009. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature* **457**: 215–218.
- Giniger E, Ptashne M. 1988. Cooperative DNA binding of the yeast transcriptional activator GAL4. *Proc Natl Acad Sci* **85**: 382–386.
- Han M, Grunstein M. 1988. Nucleosome loss activates yeast downstream promoters in vivo. *Cell* **55**: 1137–1145.
- 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.
- Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO. 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**: 533–538.
- Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**: 318–356.
- Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**: 1497–1502.
- Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, Leproust EM, Hughes TR, Lieb JD, Widom J, et al. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**: 362–366.
- Kornberg RD. 1974. Chromatin structure: A repeating unit of histones and DNA. *Science* **184**: 868–871.
- Kornberg RD, Stryer L. 1988. Statistical distributions of nucleosomes: Nonrandom locations by a stochastic mechanism. *Nucleic Acids Res* **16**: 6677–6690.
- Lam FH, Steger DJ, O'Shea EK. 2008. Chromatin decouples promoter threshold from dynamic range. *Nature* **453**: 246–250.
- Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C. 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet* **39**: 1235–1244.
- Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, et al. 2008. Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol* **6**: e27. doi: 10.1371/journal.pbio.0060027.
- Ligr M, Siddharthan R, Cross FR, Siggia ED. 2006. Gene expression from random libraries of yeast promoters. *Genetics* **172**: 2113–2122.
- Ma J. 2005. Crossing the line between activation and repression. *Trends Genet* **21**: 54–59.
- MacIsaac KD, Wang T, Gordon DB, Gifford DK, Stormo GD, Fraenkel E. 2006. An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics* **7**: 113. doi: 10.1186/1471-2105-7-113.
- Maerkl SJ, Quake SR. 2007. A systems approach to measuring the binding energy landscapes of transcription factors. *Science* **315**: 233–237.
- Miller JA, Widom J. 2003. Collaborative competition mechanism for gene activation in vivo. *Mol Cell Biol* **23**: 1623–1632.
- Perlmann T, Wrangé O. 1988. Specific glucocorticoid receptor binding to DNA reconstituted in a nucleosome. *EMBO J* **7**: 3073–3079.
- 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.
- 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.
- Ptashne M. 1967. Isolation of the lambda phage repressor. *Proc Natl Acad Sci* **57**: 306–313.
- Rajewsky N, Vergassola M, Gaul U, Siggia ED. 2002. Computational detection of genomic cis-regulatory modules applied to body patterning in the early *Drosophila* embryo. *BMC Bioinformatics* **3**: 30. doi: 10.1186/1471-2105-3-30.
- Raser JM, O'Shea EK. 2004. Control of stochasticity in eukaryotic gene expression. *Science* **304**: 1811–1814.
- Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, et al. 2000. Genome-wide location and function of DNA binding proteins. *Science* **290**: 2306–2309.

- Richmond TJ, Davey CA. 2003. The structure of DNA in the nucleosome core. *Nature* **423**: 145–150.
- Rubin-Bejerano I, Mandel S, Robzyk K, Kassir Y. 1996. Induction of meiosis in *Saccharomyces cerevisiae* depends on conversion of the transcriptional repressor Ume6 to a positive regulator by its regulated association with the transcriptional activator Ime1. *Mol Cell Biol* **16**: 2518–2526.
- Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK, Wang JP, Widom J. 2006. A genomic code for nucleosome positioning. *Nature* **442**: 772–778.
- Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, Gaul U. 2008. Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* **451**: 535–540.
- 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* **18**: 735–748.
- Shea MA, Ackers GK. 1985. The OR control system of bacteriophage lambda. A physical–chemical model for gene regulation. *J Mol Biol* **181**: 211–230.
- Sinha S, van Nimwegen E, Siggia ED. 2003. A probabilistic method to detect regulatory modules. *Bioinformatics* **19**: i292–i301.
- Struhl K. 1985. Naturally occurring poly(dA–dT) sequences are upstream promoter elements for constitutive transcription in yeast. *Proc Natl Acad Sci* **82**: 8419–8423.
- Tanay A. 2006. Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Res* **16**: 962–972.
- Thastrom A, Lowary PT, Widlund HR, Cao H, Kubista M, Widom J. 1999. Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J Mol Biol* **288**: 213–229.
- Tirosh I, Barkai N. 2008. Two strategies for gene regulation by promoter nucleosomes. *Genome Res* **18**: 1084–1091.
- van Holde KE. 1989. *Chromatin* Springer, New York.
- Weiss S, Gladstone L. 1959. A mammalian system for the incorporation of cytidine triphosphate into ribonucleic acid. *J Am Chem Soc* **81**: 4118–4119.
- Whitehouse I, Rando OJ, Delrow J, Tsukiyama T. 2007. Chromatin remodelling at promoters suppresses antisense transcription. *Nature* **450**: 1031–1035.
- Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ. 2005. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**: 626–630.

Received November 25, 2008; accepted in revised form May 15, 2009.