Statistical mechanical model of coupled transcription from multiple promoters due to transcription factor titration

Mattias Rydenfelt
Department of Physics, California Institute of Technology, Pasadena, California 91125, USA

Robert Sidney Cox III
Technology Research Association of Highly Efficient Gene Design, Kobe University, Hyogo 657-8501, Japan

Hernan Garcia
Department of Physics, Princeton University, Princeton, New Jersey 08544, USA

Rob Phillips
Department of Applied Physics, California Institute of Technology, Pasadena, California 91125, USA
and Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California 91125, USA
(Received 15 October 2012; revised manuscript received 4 October 2013; published 6 January 2014)

Transcription factors (TFs) with regulatory action at multiple promoter targets is the rule rather than the exception, with examples ranging from the cAMP receptor protein (CRP) in *E. coli* that regulates hundreds of different genes simultaneously to situations involving multiple copies of the same gene, such as plasmids, retrotransposons, or highly replicated viral DNA. When the number of TFs heavily exceeds the number of binding sites, TF binding to each promoter can be regarded as independent. However, when the number of TF molecules is comparable to the number of binding sites, TF titration will result in correlation ("promoter entanglement") between transcription of different genes. We develop a statistical mechanical model which takes the TF titration effect into account and use it to predict both the level of gene expression for a general set of promoters and the resulting correlation in transcription rates of different genes. Our results show that the TF titration effect could be important for understanding gene expression in many regulatory settings.

DOI: 10.1103/PhysRevE.89.012702 PACS number(s): 87.16.Yc, 05.40.−a

I. INTRODUCTION

Organisms respond to a variety of environmental stimuli by regulating gene expression through the action of transcription factors (TFs). An increasingly quantitative description of transcriptional regulation has made it possible to construct predictive physical models based on equilibrium statistical mechanics. A number of biologically relevant parameters have been identified in these models, including the copy number of RNA polymerase (RNAP), TFs, the strengths of their corresponding binding sites, their interaction energies and the mechanical properties of the DNA [1–3]. Another such model parameter which so far has received less attention is the number of promoters N (or operators) that a TF regulates. One reason might be that implicitly it has been assumed that the number of TFs is much greater than N, hence making TF binding to different promoters independent.

In this work we use a statistical mechanical model to show that when the number of TF molecules is comparable to the number of targets, depletion of the TF can result in nontrivial dependence of the regulatory effect on the relative abundance of targets and TF molecules. The existence of this effect has been previously explored in the context of ultrasensitive regulatory networks [4], as well as the impact of decoy binding sites on TF lifetimes and the response of particular genetic circuits [5,6]. Here we present a generalized model of gene expression in the presence of TF competition. An advantage with this model is that any system of entangled promoters can be explicitly described in terms of its individual components. Moreover, quantities of interest can be expressed analytically, which, for example, allows us to easily study the role of model parameters, explore certain limits of, e.g., strong/weak TF binding, and efficiently compute TF titration curves without the need of running thousands of time-consuming Gillespie simulations.

A recent study asserts that half of the proteins in *E. coli* come in fewer than 10 copies [7] (30 for TFs), a number comparable to the gene copy number in many important biological situations, including plasmids [8], viral infections [9], gene duplications [10], (retro)transposons [11–13], rapid cell growth [14], and transfection of DNA into animal cells [15]. Even for some TFs the number of regular chromosomal binding sites could be large enough to titrate TFs (see Appendix B). If this picture is correct, a quantitative understanding of TF titration due to multiple targets will be essential for making predictive models of transcription regulation. Such models could potentially also shed new light onto diseases where gene copy number abnormalities play a role, including cancers [16], neuropsychiatric diseases [17], and autoimmune disorders [15].

As case studies we use three specific promoter architectures, representing three different mechanisms of repressing a gene. All three of these examples have been studied extensively both experimentally and theoretically [18–23]. The *simple repression* promoter architecture is arguably the most common nonconstitutive architecture in *E. coli* [24] and refers to a single TF binding site blocking RNAP from binding the promoter. For promoters with more than one binding site for a particular TF, 34% of these promoters have two binding sites

 ©2014 American Physical Society

separated by more than 100 bp [24], indicating a frequent scenario of facilitated *repression with DNA looping* [25], Table 1]. A famous example of this promoter architecture is the well-studied *lac* operon. In a variant of this promoter architecture, reminiscent of GalR repression at the P2 promoter [26], repression can *only* be achieved in the looped conformation. This *repression exclusively due to looping* promoter architecture has the interesting feature that the level of repression is not a monotonic function in number of TFs. Though we believe these three promoter architectures are both interesting and relevant, the particular choices are not central and the formalism presented here makes it possible to calculate the titration effect for any arbitrary regulatory architecture.

The organization of this paper is as follows. In Sec. II we introduce the thermodynamic models used throughout this work and discuss their validity. In Sec. III we compute individual ($N = 1$) partition functions for the three important promoter architecture case studies. This will be an instructive exercise before turning to the more abstract treatment of Sec. IV, where we compute the partition function for a general set of promoters ($N \geqslant 1$). In Sec. V we benefit from the hard work of the previous two sections to make predictions of a quantity of great biological importance, namely the *fold change* in gene expression, a quantity directly accessible experimentally. In Sec. VI we study correlation in transcription rates of different genes due to TF titration. In Sec. VII we extend the work of previous sections to include the case when TF and promoter copy numbers are not fixed but rather fluctuating according to a statistical distribution. Finally, in Sec. VIII we use Gillespie simulations to verify the thermodynamic model and derive a relationship between the stochastic model rate constants and thermodynamic free energy parameters for the three specific promoter architectures considered.

II. UNDERLYING ASSUMPTIONS OF THERMODYNAMIC MODEL

One of the most ubiquitous quantitative descriptions of transcription is founded upon the so-called thermodynamic models of regulation. In these models, the quantitative behavior of a given promoter is characterized in terms of the occupancy of that promoter by the transcription apparatus and a constellation of molecular partners such as TFs and nucleosomes [1–3,27]. One of the reasons for the success of these thermodynamic approaches is that in some cases the time scale associated with the production of mRNA is often much slower than the rate at which most proteins, such as TFs, move around within the cell [28] and bind or unbind DNA. For example, the effective (1D + 3D [29]) diffusion constant of LacI has been measured as $D_{\text{eff}} = 0.4 \pm 0.02\ \mu\text{m}^{-2}\,\text{s}^{-1}$ [28], which means that a LacI molecule can explore the full length of an *E. coli* cell in a few seconds. This should be compared to the significantly slower production rate of LacI which, averaged over the cell cycle, corresponds to around ∼0.3 per min [30]. Thus, there is reason to believe that LacI, and probably other TFs, can significantly explore the DNA over the time scales at which LacI is produced, providing circumstantial support for a quasiequilibrium approximation. This separation of time scales permits the use of statistical mechanics at promoters that satisfy this condition in order to compute the probabilities of different configurations of TFs and RNAP on the promoter targets. The thermodynamic approach has been used far and wide for characterizing a host of different regulatory processes [1,2,27,31–37]. Interestingly, this approach not only serves as a very powerful conceptual framework for predicting the behavior of different architectures, but even in those cases where it fails it is useful for suggesting new hypotheses [22,38–42].

Of course, this thermodynamic approach is really only the simplest first idea that one can exploit, but at a deeper level it is just a caricature of the real complications of the transcription process and the next layer of sophistication involves using rate equations. However, even in those cases in which models of transcription are built using rate equations, they too essentially appeal to thermodynamic models through the functions describing the occupancy of TFs. Generically, in these cases one writes a rate of production for some protein as

$$\frac{dA}{dt} = -\gamma A + f_{\text{occupancy}}([\text{TF}]), \tag{1}$$

where $f_{\text{occupancy}}([\text{TF}])$ is an occupancy function that reflects the probability of occupancy of TF binding sites as a function of the concentration of these factors. To make the point concrete, consider the example of an activator that activates its own production. In this case, one typically writes a rate equation of the form

$$\frac{dA}{dt} = -\gamma A + r_0 + r_1 \frac{\left(\frac{A}{K_d}\right)^n}{1 + \left(\frac{A}{K_d}\right)^n}, \tag{2}$$

where the first term describes protein degradation and dilution from cell growth and the second term describes basal production at a rate r_0. The third term is a Hill function [43] relating production to the occupancy of the promoter by its activator. This is obtained using precisely the same statistical mechanics arguments that are common in thermodynamic models. The dissociation constant K_d is only meaningful in the context of equilibrium, and a rapid change in TF copy number cannot correspond to an instantaneous response in promoter occupancy. Therefore, one again needs to rest on the assumption of quasiequilibrium. The literature is replete with examples of both prokaryotic and eukaryotic transcription regulation based upon these kinds of occupancy-based rate equations [44–58], only further raising the stakes for exploring the limits and validity of this approach.

Using the thermodynamic formalism described above, we consider a (quasi) equilibrium system, where the number of RNAP (P), TFs (F), and target promoters (N) are fixed. The term *promoter* will be used either to refer to the RNAP binding site or the full promoter region, including TF binding sites, depending on context. The number of nonspecific binding sites N_{NS} is assumed to be much larger than the number of RNAPs, the number of TFs, and the number of promoters ($N_{NS} \gg P, F, N$). Representative values for these parameters in *E. coli* are given by $P \approx 10^3$ [59–61], $N_{NS} \approx 5 \times 10^6$ (the size of the *E. coli* genome), $F \approx 1$–10^3 [7,62], and $N \approx 1$–10^2 [8,9]. Unless stated otherwise we will use these given values of P and N_{NS} where concrete numbers are needed. Further, we assume that TFs and RNAP are always bound to DNA and do not roam freely in the cell. This is justified in the cases of RNAP and the

Lac repressor, for example, by studies using minicells [63,64], though this is not necessarily generically true. The results are easily adjusted to the case in which the TFs are free in the cytoplasm rather than nonspecifically bound. We furthermore assume that the promoters have no shared binding sites and that they do not interact except via the competition for TFs.

For each configuration of TFs and RNAP we associate a free energy and corresponding Boltzmann weight, which will determine the probability for the system to be in that particular state [1–3,27]. The partition function (Z) is the sum of all these weights. Using the partition function the probability of finding RNAP bound to the promoter of interest can be calculated. This probability can, in turn, be related to the level of gene expression, a quantity accessible through the use of genetic reporters, or *fold change*, defined as the ratio of the level of gene expression in the presence vs the absence of a TF of interest, by assuming that the RNAP binding probability and gene expression are linearly related [2,3]. Such a linear relationship has been observed *in vitro* between RNAP binding probability and open complex formation when RNAP binding is the rate limiting step in transcription initiation [65]. A fully generalized model of transcription initiation taking the rates of open complex formation, promoter escape as well as intermediate conformational changes into account [66–68] is beyond the scope of this paper. Likewise, we assume that TFs act by modifying the RNAP binding affinity to the promoter. For repressors we can argue that this is indeed a common mechanism of repression by noticing that almost half [69] of these operators overlap with the RNAP binding region spanning about 40 bp upstream from the transcription start site, hence blocking RNAP from binding the promoter. In some cases also other mechanisms of transcriptional regulation, such as modulation of the promoter escape rate, can be rephrased in the thermodynamic language above, e.g., in the case of fast open complex formation. In general, however, the regulatory effect of a TF on transcription initiation depends in a complex way on the TF (un)binding rates and the rates of the various transcription initiation steps of the particular promoter, which again is beyond the scope of this paper.

III. SINGLE PROMOTER PARTITION FUNCTION

A. Simple repression

Of 795 transcription units reported in RegulonDB 7.1 [24] to have at least one TF interaction, 125 correspond to simple repressors [22], making it the most common promoter architecture in *E. coli*. The simple repressor has a single binding site overlapping the promoter such that RNAP cannot bind (or form an open complex which is mathematically equivalent in the context of our model) in the presence of repressor hence inhibiting transcription [see Fig. 1(a)]. A classic example of this regulatory motif are the well-studied *lac* operon mutants [18,21].

The partition function for a simple repressor was derived in [2], but is for the independence of this paper recaptured here. We assume that when not bound to the promoter, RNAP can be found at any of N_{NS} nonspecific binding sites with a binding energy of ε_{pd}^{NS}. Treating the RNAP molecules as indistinguishable, there are $\binom{N_{NS}}{P}$ ways of arranging P RNAP molecules on this nonspecific reservoir. The partition function corresponding to this situation is

$$Z_{\mathrm{P}}^{NS} = \binom{N_{NS}}{P} e^{-\beta P \varepsilon_{pd}^{NS}}. \tag{3}$$

As stated above, we assume that $N_{NS} \gg P$, which allows us to make the approximation $\binom{N_{NS}}{P} = \frac{N_{NS}!}{P!(N_{NS}-P)!} \simeq \frac{N_{NS}^{P}}{P!}$.

Assuming that the repressor has only one binding head, leaving the more complicated case of two binding heads to Sec. III B, the logic for finding the contribution of R repressor molecules to the total partition function imitates that for RNAP, namely,

$$Z_{\mathrm{R}}^{NS} = \binom{N_{NS}}{R} e^{-\beta R \varepsilon_{rd}^{NS}}, \tag{4}$$

where ε_{rd}^{NS} is the nonspecific repressor binding energy. Again, assuming $N_{NS} \gg R$ allows us to approximate $\binom{N_{NS}}{R} \simeq \frac{N_{NS}^{R}}{R!}$.

Since the total number of nonspecific sites is in great excess with respect to both number of repressors and RNAP, we can treat nonspecific binding of repressors and RNAP as independent, and hence the total nonspecific partition function is given by the product

$$Z^{NS} = Z_{\mathrm{P}}^{NS} Z_{\mathrm{R}}^{NS}. \tag{5}$$

We use this nonspecific partition function to find the overall partition function Z that accounts for binding to the promoter. The promoter can be found in three different states: empty, occupied by RNAP, or occupied by a repressor. As a consequence the overall partition function is given by

$$Z(P,R) = \underbrace{Z^{NS}(P,R)}_{\text{empty}} + \underbrace{Z^{NS}(P-1,R)e^{-\beta\varepsilon_{pd}^{S}}}_{\text{RNAP bound}} + \underbrace{Z^{NS}(P,R-1)e^{-\beta\varepsilon_{rd}^{S}}}_{\text{repressor bound}}. \tag{6}$$

The first term corresponds to an empty promoter, the second term corresponds to taking an RNAP molecule from the nonspecific reservoir and binding it to the promoter with a specific binding energy of ε_{pd}^{S}, and the third term similarly corresponds to taking a repressor from the nonspecific reservoir and binding it to the promoter with a specific binding energy ε_{rd}^{S}. If we normalize by $Z^{NS}(P,R)$ to assign the empty promoter weight 1, the partition function is given by

$$Z = 1 + \frac{P}{N_{NS}} e^{-\beta\Delta\varepsilon_{pd}} + \frac{R}{N_{NS}} e^{-\beta\Delta\varepsilon_{rd}}, \tag{7}$$

where we have defined the energy differences $\Delta\varepsilon_{rd} = \varepsilon_{rd}^{S} - \varepsilon_{rd}^{NS}$ and $\Delta\varepsilon_{pd} = \varepsilon_{pd}^{S} - \varepsilon_{pd}^{NS}$. The factors $\frac{R}{N_{NS}}, \frac{P}{N_{NS}}$ in the last two terms are of entropic origin and associated with the cost of forcing one molecule to stay on a particular site on the DNA, rather than letting it explore the full range of possible nonspecific sites.

B. Repression with looping

In repression with looping, RNAP is still excluded from the promoter by repressor binding to a main operator in the vicinity of the promoter. In this case, however, the repressors have two binding heads that can simultaneously bind the main operator

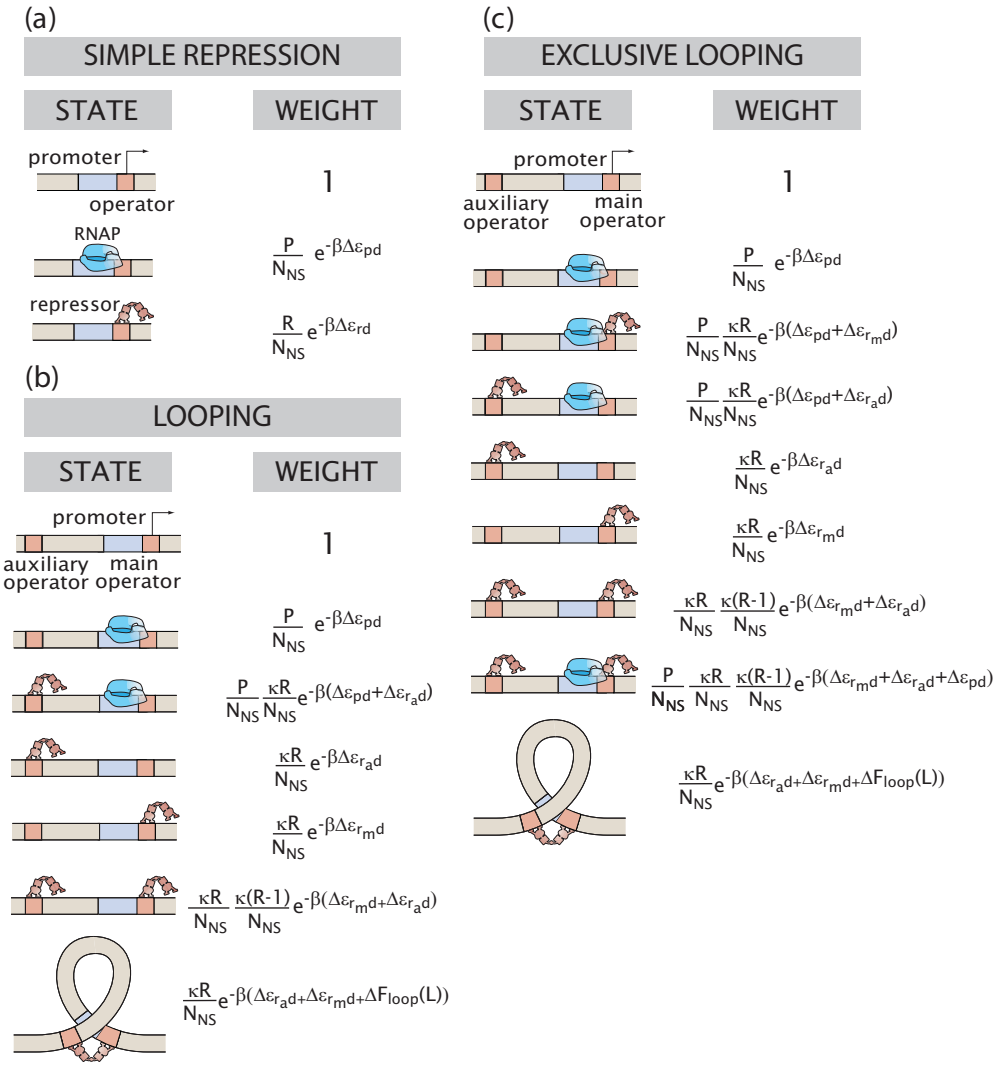


FIG. 1. (Color online) States and weights for three studied promoter architectures (a) simple repression, (b) repression with looping, and (c) repression exclusively due to looping. The last two promoter architectures differ only by the addition of two states to the exclusive looping architecture (third and eighth from the top), corresponding to RNAP and the main operator being simultaneously bound.

and an *auxiliary* operator through the formation of a DNA loop, though the auxiliary operator does not block the promoter on its own (see Fig. 2). As a result, there is an increase of effective concentration of repressor in the vicinity of the main operator leading to an increase in repression [18–20,70]. One of the most studied realizations of this promoter architecture is again based on modifications of the *lac* operon [18,19].

To compute the nonspecific partition function for repressors with two binding heads, we begin with a single repressor molecule ($R = 1$). Then, invoking the assumption that the nonspecifically bound repressors are noninteracting, it is easy to generalize the result to any number of repressors ($R > 1$). A single repressor molecule can be found in either a looped state, with both heads bound, or in a state with one head unbound. Each bound repressor head acquires a binding energy of ε_{rd}^{NS}, and for looped states there is an additional free energy cost (elastic plus entropic) $F_{\text{loop}}(i,j)$ of bringing two sites i and j together.

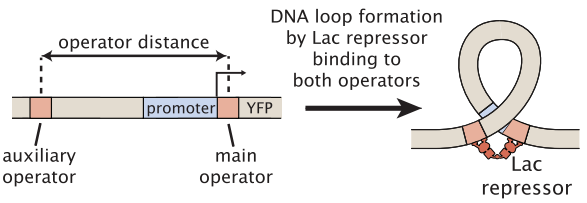


FIG. 2. (Color online) Repression through DNA looping. The repressor binds to the main and auxiliary operators simultaneously looping the intervening DNA.

Taking every possible such configuration into account we find the nonspecific single repressor partition function

$$Z_R^{NS}(R=1) = \underbrace{\sum_{i=1}^{N_{NS}} e^{-\beta\varepsilon_{rd}^{NS}}}_{\text{One head bound}} + \underbrace{\frac{1}{2} e^{-2\beta\varepsilon_{rd}^{NS}} \sum_{i=1}^{N_{NS}} \sum_{j=1, j\neq i}^{N_{NS}} e^{-\beta F_{\text{loop}}(i,j)}}_{\text{Two heads bound}}. \quad (8)$$

The factor of $\frac{1}{2}$ in the second sum is necessary to avoid double counting of the looped states. To simplify this expression we assume translational invariance, such that the last sum over j is independent of i. This assumes that on average DNA "looks the same" everywhere, at least locally. Using this assumption we get

$$Z_R^{NS}(R=1) = N_{NS} e^{-\beta\varepsilon_{rd}^{NS}} \left(1 + \frac{1}{2} e^{-\beta\varepsilon_{rd}^{NS}} \sum_{j=2}^{N_{NS}} e^{-\beta F_{\text{loop}}(1,j)}\right) \equiv N_{NS} e^{-\beta\varepsilon_{rd}^{NS}} e^{-\beta F_{\text{eff}}^{NS}}. \quad (9)$$

In the last step we defined the effective nonspecific free energy F_{eff}^{NS}. To extend $Z_R^{NS}(R=1)$ to an arbitrary number of repressors $R \geqslant 1$ we use a familiar result from statistical mechanics, namely,

$$Z_R^{NS}(R) = \frac{1}{R!}\left[Z_R^{NS}(R=1)\right]^R \quad (10)$$

$$= \frac{N_{NS}^R}{R!} e^{-\beta R(\varepsilon_{rd}^{NS} + F_{\text{eff}}^{NS})}, \quad (11)$$

which is applicable for indistinguishable and noninteracting repressors.

Finally, to find the total nonspecific partition function we combine $Z_R^{NS}(R)$ with the non-specific partition function for RNAP found in previous section [Eq. (3)] resulting in

$$Z^{NS}(R,P) = Z_R^{NS}(R) Z_P^{NS}(P) = \frac{N_{NS}^R}{R!} \frac{N_{NS}^P}{P!} e^{-\beta R\left(\varepsilon_{rd}^{NS} + F_{\text{eff}}^{NS}\right)} e^{-\beta P \varepsilon_{pd}^{NS}}. \quad (12)$$

Our next task is to determine the weights for all states of the promoter that are shown in Fig. 1(b). As an example we show how to determine the weight for the state with only the main operator bound by a repressor with binding energy $\varepsilon_{r_m d}^S$. For this state we need to consider all configurations for the second repressor head not bound to the main operator, as well as all configurations $Z_R^{NS}(R-1)$ for the remaining $R-1$ nonspecifically bound repressors. The weight associated with the specifically bound repressor is given by

$$Z_R^{NS}(R=1, \text{one repressor head bound to main operator}) = e^{-\beta\varepsilon_{r_m d}^S} \left(1 + e^{-\beta\varepsilon_{rd}^{NS}} \sum_{j=2}^{N_{NS}} e^{-\beta F_{\text{loop}}(1,j)}\right) \equiv e^{-\beta\varepsilon_{r_m d}^S} e^{-\beta\tilde{F}_{\text{eff}}^{NS}}, \quad (13)$$

where we have introduced another useful effective free energy $\tilde{F}_{\text{eff}}^{NS}$ (note the absent factor of $\frac{1}{2}$), which allows us to express the weight associated with the nonspecifically bound or free hanging repressor head simply as $e^{-\beta\tilde{F}_{\text{eff}}^{NS}}$.

Using the same normalization condition as above we find the Boltzmann weight for the state with only the main operator bound,

$$\text{Weight } (\text{  }) = \frac{e^{-\beta\varepsilon_{r_m d}^S} e^{-\beta\tilde{F}_{\text{eff}}^{NS}} Z_R^{NS}(R-1) Z_P^{NS}(P)}{Z_R^{NS}(R) Z_P^{NS}(P)} = \frac{R}{N_{NS}} e^{-\beta(\varepsilon_{r_m d}^S - \varepsilon_{rd}^{NS})} e^{-\beta(\tilde{F}_{\text{eff}}^{NS} - F_{\text{eff}}^{NS})} = \frac{\kappa R}{N_{NS}} e^{-\beta\Delta\varepsilon_{r_m d}}. \quad (14)$$

For convenience we introduce the following notation:

$$\Delta\varepsilon_{r_m d} = \varepsilon_{r_m d}^S - \varepsilon_{rd}^{NS}, \quad \Delta\varepsilon_{r_a d} = \varepsilon_{r_a d}^S - \varepsilon_{rd}^{NS}, \quad \Delta\varepsilon_{pd} = \varepsilon_{pd}^S - \varepsilon_{pd}^{NS}, \quad \Delta F_{\text{loop}} = F_{\text{loop}}^S - \left(\tilde{F}_{\text{eff}}^{NS} - \varepsilon_{rd}^{NS}\right), \quad (15)$$

$$\kappa = e^{-\beta(\tilde{F}_{\text{eff}}^{NS} - F_{\text{eff}}^{NS})}, \quad p = \frac{P}{N_{NS}} e^{-\beta\Delta\varepsilon_{pd}}.$$

Here $\Delta\varepsilon_{r_m d}$ and $\Delta\varepsilon_{r_a d}$ correspond to the main and auxiliary operators, respectively. From the definitions of $\tilde{F}_{\text{eff}}^{NS}$ and F_{eff}^{NS} it is easy to see that κ is always a number between 1 and 2. If $\tilde{F}_{\text{eff}}^{NS} \gg 1$ there is a large probability of nonspecific loop formation and $\kappa \simeq 2$. On the other hand, if $\tilde{F}_{\text{eff}}^{NS} \simeq 1$ then there is just a small probability of nonspecific loop formation and $\kappa \simeq 1$. Thus, κ can be viewed as a parameter related to how many repressor heads are effectively bound nonspecifically to DNA.

Using the same method we can compute the weights for all other states, and by adding these weights together we get the single promoter partition function

$$Z = 1 + p + p\left(\frac{\kappa R}{N_{NS}}\right) e^{-\beta\Delta\varepsilon_{r_a d}} + \left(\frac{\kappa R}{N_{NS}}\right) e^{-\beta\Delta\varepsilon_{r_a d}} + \left(\frac{\kappa R}{N_{NS}}\right) e^{-\beta\Delta\varepsilon_{r_m d}} + \frac{\kappa R}{N_{NS}} \frac{\kappa(R-1)}{N_{NS}} e^{-\beta(\Delta\varepsilon_{r_a d} + \Delta\varepsilon_{r_m d})} + \left(\frac{\kappa R}{N_{NS}}\right) e^{-\beta(\Delta\varepsilon_{r_a d} + \Delta\varepsilon_{r_m d} + \Delta F_{\text{loop}})}. \quad (16)$$

Here the states are listed in the same order as in Fig. 1(b).

C. Exclusive looping repression

For repression due exclusively to looping the situation is similar to the previous section but with the difference that RNAP is considered to be blocked from binding the promoter *only* in the looped state. Hence, it is not enough for just the main operator to be occupied to achieve repression. Such a model of repression is reminiscent of the mechanism of galactose metabolism repression by GalR at the $P2$ promoter [26] and the arabinose metabolism AraC repression at the P_C promoter in the absence of arabinose [71].

For this promoter architecture terms need to be added to the partition function of Eq. (16) corresponding to states with the main or auxiliary operator bound by repressor and the promoter bound by RNAP. In Fig. 1(c) these states are given by the third and eighth states from the top. After taking these new states into account we find the single promoter partition function

$$Z = 1 + p + p\left(\frac{\kappa R}{N_{NS}}\right) e^{-\beta\Delta\varepsilon_{r_m d}} + p\left(\frac{\kappa R}{N_{NS}}\right) e^{-\beta\Delta\varepsilon_{r_a d}} + \left(\frac{\kappa R}{N_{NS}}\right) e^{-\beta\Delta\varepsilon_{r_a d}} + \left(\frac{\kappa R}{N_{NS}}\right) e^{-\beta\Delta\varepsilon_{r_m d}} + (1+p)\frac{\kappa R}{N_{NS}} \frac{\kappa(R-1)}{N_{NS}} e^{-\beta(\Delta\varepsilon_{r_a d} + \Delta\varepsilon_{r_m d})} + \left(\frac{\kappa R}{N_{NS}}\right) e^{-\beta(\Delta\varepsilon_{r_a d} + \Delta\varepsilon_{r_m d} + \Delta F_{\text{loop}})}, \quad (17)$$

where again the states have been listed in the same order as in Fig. 1(c).

IV. MULTIPLE PROMOTER PARTITION FUNCTION

The simplest example of computing the total partition function for a set of promoters with individual partition functions $Z^{(1)}, Z^{(2)}, \ldots Z^{(N)}$ is when these are *independent*. In our model this happens when the promoters are unregulated or regulated by a TF whose copy number F greatly exceeds the number of promoters ($F \gg N$). Then if one TF binds to a promoter, the number of remaining available TFs is left essentially unchanged, and hence the other promoters are unaffected. By a familiar result from statistical mechanics the total partition function Z^{tot} for a system of independent promoters is given by

$$Z^{\text{tot}} = Z^{(1)} Z^{(2)} \cdots Z^{(N)} \quad \text{for} \quad F \gg N. \tag{18}$$

The complications associated with computing the partition function for a set of promoters regulated by the same TFs originate from the fact that at low TF copy numbers the promoters get "entangled." For entangled promoters, binding of one TF molecule to a promoter directly influences the TF binding probability to another promoter, due to an effective decrease in the number of available TFs. In the following sections we extend Eq. (18) and derive the total partition function for a general set of promoters without making any assumptions about the number of TFs or promoters. While this generality leads to somewhat more abstract derivations, it has the benefit of allowing us to apply the results to a wide range of interesting problems.

A. General set of promoters

We start by deriving the total partition function for a general set of, potentially different, promoters under control of a single type of TF (F). In Appendix A we generalize to regulation with an arbitrary number of TF types.

First we introduce the notation needed to make these calculations. Let f_n and p_n denote the number of TFs and RNAP bound to promoter $n \in \{1, \ldots, N\}$, respectively. Here f_n is constrained by the number of binding sites and the total number of TFs in the cell, namely, $\sum_n f_n \leqslant F$ and p_n is always either 0 or 1. Let s_n denote the state of promoter n (e.g., empty promoter, operator 1 occupied, operator 2 unoccupied, etc.), and let $F(s_n)$ and $P(s_n)$ denote number of TFs and RNAP bound at promoter n for state s_n. To compute the total partition function we take every allowed state into account by summing over the variables f_n and p_n, as well as the variables s_n for all states compatible with the choice (f_n, p_n). For each choice $\{f_n\}$ and $\{p_n\}$ there will be $F - \sum_i f_i$ TFs and $P - \sum_i p_i$ RNAPs left for nonspecific binding on the DNA "reservoir," and the statistical weight associated with these are given by the nonspecific partition functions $Z_F^{NS}(F - \sum_i f_i)$ and $Z_P^{NS}(F - \sum_i p_i)$, which we assume to have the forms $Z_F^{NS}(F) = \frac{N_{NS}^F}{F!} e^{-\beta F \varepsilon_{fd}^{NS}}$ and $Z_P^{NS}(P) = \frac{N_{NS}^P}{P!} e^{-\beta P \varepsilon_{pd}^{NS}}$, in accordance with our results for the simple repressor (Sec. III A) and repression by looping architecture (Sec. III B). The parameter ε_{fd}^{NS} is assumed to be independent of F. The specifically bound TFs and RNAP to promoter n will acquire a free energy $E(s_n)$ for state s_n. Since there might be many possible states s_n for a given choice (f_n, p_n) we need to sum over all states s_n compatible with this choice, to find the specific part $\sum_{s_n} e^{-\beta E(s_n)}|_{F(s_n)=f_n,\, P(s_n)=p_n}$ of the statistical weight for promoter n. If there are no states s_n for a given $(f_n,\ p_n)$, the sum over s_n is set equal to 0. This is, for example, the case for the simple repressor which cannot have both TF and RNAP specifically bound at the same time ($f_n = p_n = 1$) due to steric exclusion. The specific part of the weight for different promoters "commute," meaning that we can simply multiply these parts together. The promoter entanglement is fully contained inside the F dependent factorial terms, which motivates the order we have chosen to carry out the summations (f_n, p_n, s_n). Using a normalization where the state with N empty promoters is assigned weight 1, the total partition function is given by

$$\begin{aligned} Z^{\text{tot}} &= \sum_{\substack{f_1,\ldots,f_N \\ \sum_i f_i \leqslant F}} \sum_{p_1,\ldots,p_N} \frac{Z_F^{NS}\left(F - \sum_i f_i\right) Z_P^{NS}\left(P - \sum_i p_i\right)}{Z_F^{NS}(F) Z_P^{NS}(P)} \\ &\quad \times \prod_{n=1}^{N} \sum_{\substack{s_n \\ F(s_n)=f_n \\ P(s_n)=p_n}} e^{-\beta E(s_n)} \\ &= \sum_{\substack{f_1,\ldots,f_N \\ \sum_i f_i \leqslant F}} \sum_{p_1,\ldots,p_N} \frac{F!}{N_{NS}^{\sum_i f_i}\left(F - \sum_i f_i\right)!} \\ &\quad \times \frac{P!}{N_{NS}^{\sum_i p_i}\left(P - \sum_i p_i\right)!} \prod_{n=1}^{N} \sum_{\substack{s_n \\ F(s_n)=f_n \\ P(s_n)=p_n}} e^{-\beta \Delta E(s_n)}, \end{aligned} \tag{19}$$

where on the second line we have defined $\Delta E(s_n) = E(s_n) - f_n \varepsilon_{fd}^{NS} - p_n \varepsilon_{pd}^{NS}$.

We now use the "high RNAP copy number" assumption $P \gg N$ to make further progress on Eq. (19) by approximating $\binom{P}{P-i} \simeq \frac{P^i}{i!}$ for i specifically bound RNAP, resulting in

$$\begin{aligned} Z^{\text{tot}} &\simeq \sum_{\substack{f_1,\ldots,f_N \\ \sum_i f_i \leqslant F}} \sum_{p_1,\ldots,p_N} \frac{F!}{N_{NS}^{\sum_i f_i}\left(F - \sum_i f_i\right)!} \\ &\quad \times \prod_{n=1}^{N} \sum_{\substack{s_n \\ F(s_n)=f_n \\ P(s_n)=p_n}} \left(\frac{P}{N_{NS}}\right)^{p_n} e^{-\beta \Delta E(s_n)} \\ &= \sum_{\substack{f_1,\ldots,f_N \\ \sum_i f_i \leqslant F}} \frac{F!}{N_{NS}^{\sum_i f_i}\left(F - \sum_i f_i\right)!} Z_{f_1}^{(1)} Z_{f_2}^{(2)} \cdots Z_{f_N}^{(N)} \\ &= \sum_{f_1=0}^{\min(B_1, F)} \sum_{f_2=0}^{\min(B_2, F-f_1)} \cdots \sum_{f_N=0}^{\min(B_N, F-\sum_{i=1}^{N-1} f_i)} \\ &\quad \times \frac{F!}{N_{NS}^{\sum_i f_i}\left(F - \sum_i f_i\right)!} Z_{f_1}^{(1)} Z_{f_2}^{(2)} \cdots Z_{f_N}^{(N)}. \end{aligned} \tag{20}$$

Here B_n is the number of TF binding sites on promoter n, and $Z_{f_n}^{(n)}$ has been defined as

$$Z_{f_n}^{(n)} \equiv \sum_{p_n} \sum_{\substack{s_n \\ F(s_n)=f_n \\ P(s_n)=p_n}} \left(\frac{P}{N_{NS}}\right)^{p_n} e^{-\beta \Delta E(s_n)}. \tag{21}$$

A key observation is that the single promoter partition functions $Z^{(n)}$ are precisely given in terms of the $Z_i^{(n)}$ factors,

$$Z^{(n)} = \sum_{i=0}^{B_n} \frac{F!}{N_{NS}^i (F-i)!} Z_i^{(n)}, \tag{22}$$

which implies that once the single promoter partition functions are known, the total partition function for the set of promoters can be directly obtained from Eq. (20), independently of promoter architectures.

B. Identical promoters

Evaluating the total partition function for a general set of promoters can be computationally expensive. In Eq. (20) there are N summation indices $\{f_i\}$ and if these are not constrained by the number of TFs ($F \geqslant \sum_{i=1}^{N} B_i$) there are $\prod_{i=1}^{N}(1+B_i)$ different terms in the summation. As the number of promoters N increases this number grows exponentially, and computing the partition function presents a great challenge. In the important special case of N identical promoter copies each with partition function,

$$Z = \sum_{i=0}^{B} \frac{F!}{N_{NS}^i (F-i)!} Z_i; \tag{23}$$

however, the computational cost can be significantly reduced.

One way to keep track of the total number of bound TFs is to introduce numbers $\{k_i\}$, where k_i denotes the number of promoter copies occupied by i TFs, with the additional constraints $\sum_{i=0}^{B} k_i = N$ and $\sum_{i=0}^{B} i k_i \leqslant F$. To compute the partition function we first need to find the number of possible arrangements given numbers $\{k_i\}$, or the "degeneracy." As an example for $k_0 = N$ there is only one choice (all promoters empty), but for $k_0 = N-1, k_1 = 1$ there are N different choices, corresponding to N different ways of choosing a single promoter to be occupied by one TF (assuming $B, F \geqslant 1$). Here we treat the promoters as distinguishable physical objects, which is a valid assumption since the promoters have additional intrinsic degrees of freedom (e.g., position) that separate them. Starting with empty promoters, there are $\binom{N}{k_0}$ ways of choosing k_0 promoters without bound TF. From the remaining $N-k_0$ promoters we choose k_1 promoters with exactly one TF bound; this can be done in $\binom{N-k_0}{k_1}$ ways. Repeating this procedure B times gives us the degeneracy, namely,

$$\begin{aligned} \text{degeneracy } \{k_i\} &= \binom{N}{k_0}\binom{N-k_0}{k_1}\cdots\binom{N-\sum_{i=0}^{B-1} k_i}{k_B} \\ &= \binom{N}{k_0,k_1,\ldots,k_B}, \end{aligned} \tag{24}$$

where $\binom{N}{k_0,k_1,\ldots,k_B} = \frac{N!}{k_0!k_1!\cdots k_B!}$ is the multinomial coefficient. To find the total partition function Z^{tot} we need to sum over all allowed values of $\{k_0,k_1,\ldots,k_B\}$ and take the degeneracy into account. Using otherwise the same weights as in Eq. (20) we find the total partition function for identical promoter copies,

$$\begin{aligned} Z^{\text{tot}} &= \sum_{\substack{k_0,k_1,\ldots k_B \\ \sum_i k_i = N \\ \sum_i i k_i \leqslant F}} \binom{N}{k_0,k_1,\ldots,k_B} \frac{F!}{N_{NS}^{\sum_i i k_i}\left(F-\sum_i i k_i\right)!} \prod_{i=0}^{B} Z_i^{k_i} \\ &= \sum_{k_B=0}^{\min(N,\lfloor F/B \rfloor)} \cdots \sum_{k_j=0}^{\min(N-\sum_{i=j+1}^{B} k_i,\, \lfloor (F-\sum_{i=j+1}^{B} i k_i)/j \rfloor)} \\ &\quad \cdots \sum_{k_1=0}^{\min(N-\sum_{i=2}^{B} k_i,\, F-\sum_{i=2}^{B} i k_i)} \\ &\quad \times \binom{N}{k_0,k_1,\ldots,k_B} \frac{F!}{N_{NS}^{\sum_i i k_i}\left(F-\sum_i i k_i\right)!} \prod_{i=0}^{B} Z_i^{k_i}. \end{aligned} \tag{25}$$

Here k_0 is assigned the implicit value $k_0 = N - \sum_{i=1}^{B} k_i$ and $\lfloor \cdot \rfloor$ denotes the floor function.[1]

When the indices $\{k_i\}$ are not constrained by the number of TFs ($F \geqslant NB$), corresponding to the most computationally expensive case, the number of terms in the summation of Eq. (25) equals the number of non-negative integer solutions to the equation

$$k_0 + k_1 + \cdots + k_B = N. \tag{26}$$

This is a classical problem from combinatorics with the number of solutions given by $\binom{N+B}{N} \approx N^B/B!$, which grows *polynomially* with number of promoters N. Intuitively, we can understand the polynomial dependence from the fact that there are B different indices (not counting $k_0 = N - \sum_{i=1}^{B} k_i$), each of which can take N different values. Hence, the partition function for identical promoter copies can be computed for much higher values of promoter copies N than permitted by the general formula [Eq. (20)].

1. Simple repression

We now use our general results [Eq. (25)] to compute the partition function for multiple copies of the specific promoter architectures considered in Sec. III, starting with simple repression. From the single promoter partition function Z [Eq. (7)] one can easily termwise identify $Z_0 = 1 + \frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_{pd}}$ and $Z_1 = e^{-\beta \Delta \varepsilon_{rd}}$, where $Z = Z_0 + \frac{R}{N_{NS}} Z_1$. These factors are needed to compute the total partition function for multiple promoter copies.

[1]The floor function $\lfloor x \rfloor$ is the largest integer not greater than x, e.g., $\lfloor 1.8 \rfloor = \lfloor 1.2 \rfloor = 1$.

Plugging Z_0, Z_1 into the general formula of Eq. (25) gives us the total partition function for N promoters,

$$Z^{\text{tot}} = \sum_{k_1=0}^{\min(N,R)} \binom{N}{k_1} \frac{R!}{N_{NS}^{k_1}(R-k_1)!} e^{-\beta k_1 \Delta\varepsilon_{rd}} (1+p)^{N-k_1}, \tag{27}$$

where $p = \frac{P}{N_{NS}} e^{-\beta \Delta\varepsilon_{pd}}$. The summation in Eq. (27) can be carried out explicitly to yield a closed form expression of the partition function in terms of the Tricomi confluent hypergeometric function [72,73].

2. Repression with looping

From the single promoter partition function for repression with looping [Eq. (16)] we identify the Z_0, Z_1, Z_2 factors

$$\begin{aligned} Z_0 &= 1 + p, \\ Z_1 &= \kappa\left[e^{-\beta\Delta\varepsilon_{r_m d}} + (1+p)e^{-\beta\Delta\varepsilon_{r_a d}} + e^{-\beta(\Delta\varepsilon_{r_m d}+\Delta\varepsilon_{r_a d}+\Delta F_{\text{loop}})}\right], \\ Z_2 &= \kappa^2 e^{-\beta(\Delta\varepsilon_{r_m d}+\Delta\varepsilon_{r_a d})}, \end{aligned} \tag{28}$$

where $Z = Z_0 + \frac{R}{N_{NS}} Z_1 + \frac{R(R-1)}{N_{NS}^2} Z_2$. With the help of these we get the total promoter partition function for N promoter copies from Eq. (25),

$$\begin{aligned} Z^{\text{tot}} = &\sum_{k_2=0}^{\min(N,\lfloor R/2 \rfloor)} \sum_{k_1=0}^{\min(N-k_2, R-2k_2)} \binom{N}{k_2, k_1, N-k_2-k_1} \\ &\times \frac{R!}{N_{NS}^{k_1+2k_2}(R-k_1-2k_2)!} \\ &\times \kappa^{k_1+2k_2}\left[e^{-\beta\Delta\varepsilon_{r_m d}} + (1+p)e^{-\beta\Delta\varepsilon_{r_a d}} \right. \\ &\left. + e^{-\beta(\Delta\varepsilon_{r_m d}+\Delta\varepsilon_{r_a d}+\Delta F_{\text{loop}})}\right]^{k_1} \\ &\times (1+p)^{N-k_1-k_2} e^{-\beta k_2(\Delta\varepsilon_{r_m d}+\Delta\varepsilon_{r_a d})}. \end{aligned} \tag{29}$$

3. Exclusive looping repression

Again, using the single promoter partition function [Eq. (17)] we identify the Z_0, Z_1, Z_2 factors for the exclusive looping repression architecture,

$$\begin{aligned} Z_0 &= 1 + p, \\ Z_1 &= \kappa\left[(1+p)\left(e^{-\beta\Delta\varepsilon_{r_m d}} + e^{-\beta\Delta\varepsilon_{r_a d}}\right)\right. \\ &\quad \left. + e^{-\beta(\Delta\varepsilon_{r_m d}+\Delta\varepsilon_{r_a d}+\Delta F_{\text{loop}})}\right], \\ Z_2 &= \kappa^2(1+p)e^{-\beta(\Delta\varepsilon_{r_m d}+\Delta\varepsilon_{r_a d})}. \end{aligned} \tag{30}$$

By plugging these factors into Eq. (25) we find the total partition function for N promoter copies

$$\begin{aligned} Z^{\text{tot}} = &\sum_{k_2=0}^{\min(N,\lfloor R/2 \rfloor)} \sum_{k_1=0}^{\min(N-k_2, R-2k_2)} \binom{N}{k_2, k_1, N-k_2-k_1} \\ &\times \frac{R!}{N_{NS}^{k_1+2k_2}(R-k_1-2k_2)!} \\ &\times \kappa^{k_1+2k_2}\left[(1+p)\left(e^{-\beta\Delta\varepsilon_{r_m d}} + e^{-\beta\Delta\varepsilon_{r_a d}}\right)\right. \\ &\left. + e^{-\beta(\Delta\varepsilon_{r_m d}+\Delta\varepsilon_{r_a d}+\Delta F_{\text{loop}})}\right]^{k_1} \\ &\times (1+p)^{N-k_1} e^{-\beta k_2(\Delta\varepsilon_{r_m d}+\Delta\varepsilon_{r_a d})}. \end{aligned} \tag{31}$$

V. FOLD CHANGE

In order to create a bridge between experimental measurements and the thermodynamic model a key assumption is made stating that the level of expression of a gene is proportional to the probability of RNAP being bound to the promoter of the gene, or in the case of multiple gene copies, the expression is proportional to the average number of promoters bound by RNAP. Using this assumption we can predict the *fold change*, defined as the ratio of level of gene expression in the presence vs absence of a certain TF, which is a quantity commonly measured by experiments. We start by computing the fold change for a set of identical promoter copies and then move to the case with a general set of promoters. In the Supplemental Material [74] we show how to perform these computations using MATHEMATICA.

By assuming the number of RNAP molecules to be much bigger than the number of promoter copies, any state with i promoters bound by RNAP will have a weight of the form $\propto p^i$, with $p = \frac{P}{N_{NS}} e^{-\beta\Delta\varepsilon_{pd}}$. Here $\Delta\varepsilon_{pd}$ is the energy difference between specific and nonspecific RNAP binding to the promoter. Using this observation one can show that the expectation value for the number of promoters bound by RNAP is given by

$$\text{Occupancy} = p\frac{\partial}{\partial p} \ln Z^{\text{tot}}. \tag{32}$$

Equation (32) together with the partition function derived in the previous section allows us to compute the fold change f, defined as the ratio between occupancy in the presence and absence of a TF,

$$f = \frac{\text{Occupancy } (F)}{\text{Occupancy } (F=0)}. \tag{33}$$

In the particular case of simple repression, plugging the partition function [Eq. (27)] into Eqs. (32) + (33) leads, after a bit of algebra, to

$$f = \frac{1+p}{N} \frac{\sum_{k_1=0}^{\min(N,R)} \binom{N}{k_1} \frac{R!}{N_{NS}^{k_1}(R-k_1)!} e^{-\beta k_1 \Delta\varepsilon_{rd}} (N-k_1)(1+p)^{N-k_1-1}}{\sum_{k_1=0}^{\min(N,R)} \binom{N}{k_1} \frac{R!}{N_{NS}^{k_1}(R-k_1)!} e^{-\beta k_1 \Delta\varepsilon_{rd}} (1+p)^{N-k_1}}. \tag{34}$$

For weak promoters ($p \ll 1$) we can simplify this expression somewhat by dropping the last factor in the numerator and denominator. The summation can again be expressed in closed form using the Tricomi confluent hypergeometric function and a corresponding differentiation rule [75].

In Fig. 3, we show fold change as a function of number of repressors (R) for the three different promoter architectures considered in Sec. III. This figure shows the importance of TF titration as there can exist order of magnitude differences in predicted fold change for $N = 1$ vs $N \geqslant 1$ promoter copies. For the simple repressor [Fig. 3(a)], with $R < N$ the fold change will never be less than $\frac{1}{N}$, corresponding to a situation where all promoters but one are "turned off." However, as soon as $R \geqslant N$ all promoters can be repressed, which yields a steep decline in fold change around $R \approx N$, at least when the operators are strong enough to have high repressor binding probability (as is the case in Fig. 3). For weak operators the move across the "boundary" $R \approx N$ is uneventful and no such steep response occurs (see Fig. 4).

In the exclusive looping repression architecture [Fig. 3(c)], the fold change exhibits a sharp trough near $R \approx N$. This is explained by the fact that at high repressor copy number the operators will be bound by repressors separately (an unrepressed state), hence avoiding having to pay the energy cost of bending the DNA, and for low repressor copy number ($R < N$) the fold change is again never less than $\frac{1}{N}$. The observed trough corresponds to the middle range between these two extremes.

Finally, the repression with looping architecture [Fig. 3(b)] is a combination of the simple repression and exclusive looping repression architectures. Since both of these architectures show steep response around $R \approx N$ the repression by looping architecture will share this feature, as is apparent from Fig. 3(b). The free energy cost ΔF_{loop} of forming DNA loops is critical for this behavior. If ΔF_{loop} is increased such that it exceeds the binding energy of both operators, $\Delta F_{\text{loop}} > \max(|\Delta\varepsilon_{r_m d}|, |\Delta\varepsilon_{r_a d}|)$, the auxiliary operator serves only to titrate repressors and the fold change will resemble the simple repression case. For all architectures, the fold change curves converge in the high TF copy number limit ($R \gg N$) independently of promoter copy number. In this limit the number of TFs available for binding is essentially constant and transcription from each promoter can be regarded as independent.

So far we assumed that all promoters are identical; however, for a general set of promoters there might be several different "output" proteins, each with its own associated fold change. By analogy to the identical promoter case [Eq. (33)] we define the fold change $f^{(n)}$ with respect to promoter n as

$$f^{(n)} \equiv \frac{\text{Occupancy for promoter } n(F)}{\text{Occupancy for promoter } n(F=0)}, \tag{35}$$

where the occupancy is given by

$$\text{Occupancy for promoter } n = p^{(n)} \frac{\partial}{\partial p^{(n)}} \ln Z^{\text{tot}}, \tag{36}$$

with $p^{(n)} \equiv \frac{P}{N_{NS}} e^{-\beta \Delta\varepsilon_{p_n d}}$ and $\Delta\varepsilon_{p_n d}$ the energy difference between specific and nonspecific RNAP binding to promoter n.

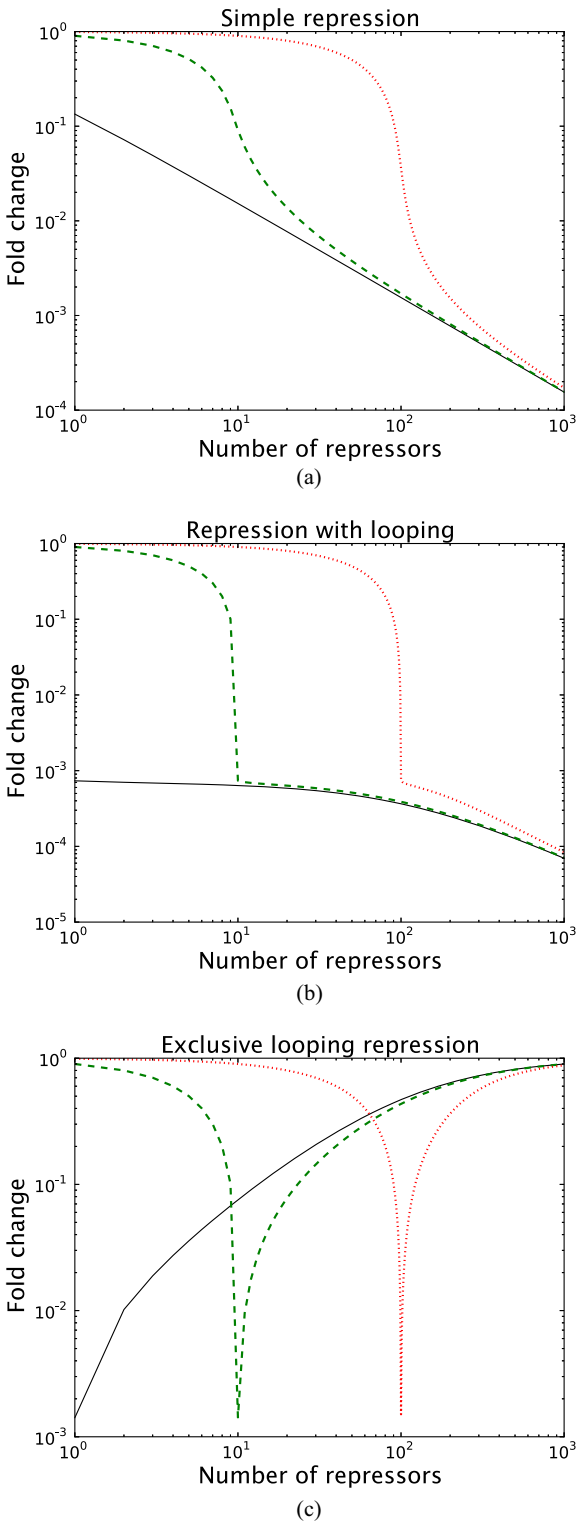


FIG. 3. (Color online) Fold change as a function of repressor copy number (R) for gene copy numbers $N = 1$ (solid line), $N = 10$ (dashed line), and $N = 100$ (dotted line) for three different promoter architectures: (a) simple repression, (b) repression with looping, and (c) exclusive looping repression. For these plots we used operator binding energy $-17.3\,k_B T$ (equivalent to the strongest known *lac* operator *Oid* [21]), the number of nonspecific sites as the genome length of *E. coli* ($N_{NS} = 5 \times 10^6$), number of RNAP $P = 1000$, and the looping energy $\Delta F_{\text{loop}} = 10\,k_B T$ [2]. The RNAP promoter binding energy is assumed to be weak ($p \ll 1$) [21,76].

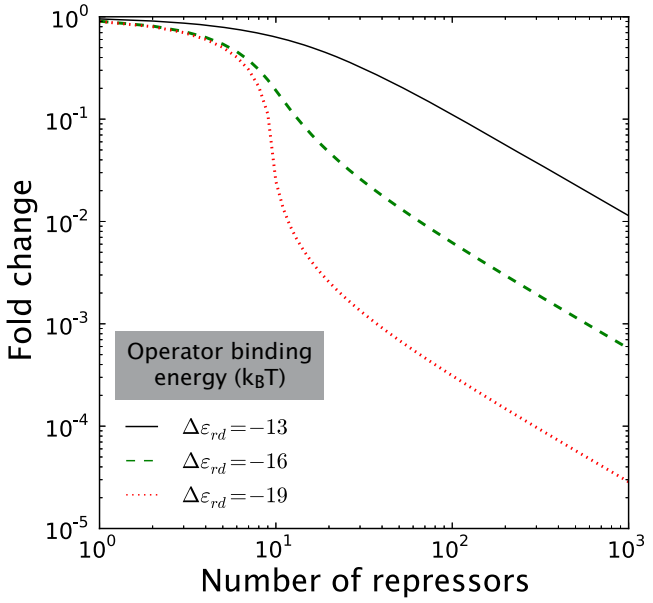


FIG. 4. (Color online) Fold change of a simple repressor with gene copy number $N = 10$ for three different TF binding site strengths, with strengths chosen to correspond to the range observed for real repressors. Stronger repressor binding leads to a steeper response in fold change around $R \approx N$. The RNAP promoter binding energy is assumed to be weak ($p \ll 1$), and the number of nonspecific sites $N_{NS} = 5 \times 10^6$.

If one promoter has stronger TF binding sites than the other promoters these binding sites will, in general, be filled first by TFs, but as soon as this happens the other promoter might experience a sudden regulatory response [4]. As an example [6], let us assume we have N_{pl} plasmids, each with one TF binding site of energy $\Delta\varepsilon_{pl}$ as shown in Fig. 5(a). The $Z_i^{(1)}$ factors associated with these N_{pl} binding sites are given by

$$Z_i^{(1)} = \binom{N_{pl}}{i} e^{-\beta i \Delta\varepsilon_{pl}}, \tag{37}$$

corresponding to $\binom{N_{pl}}{i}$ ways of distributing i repressors on N_{pl} plasmids, each with one binding site. Furthermore, let the same TF act as an inhibitor (see Sec. III A) for a single simply repressed gene located on the chromosome. We already know the $Z_i^{(2)}$ factors for this promoter architecture from Sec. IV B1, namely $Z_0^{(2)} = 1 + \frac{P}{N_{NS}} e^{-\beta\Delta\varepsilon}$ and $Z_1^{(2)} = e^{-\beta\Delta\varepsilon}$. Using Eq. (20) we find the total partition function of the system

$$\begin{aligned} Z^{\text{tot}} &= \sum_{i_1=0}^{\min(N_{pl},R)} \sum_{i_2=0}^{\min(1,R-i_1)} \frac{R!}{N_{NS}^{i_1+i_2}(R-i_1-i_2)!} Z_{i_1}^{(1)} Z_{i_2}^{(2)} \\ &= (1+p) \sum_{i_1=0}^{\min(N_{pl},R)} \frac{R!}{N_{NS}^{i_1}(R-i_1)!} \binom{N_{pl}}{i_1} e^{-\beta i_1 \Delta\varepsilon_{pl}} \\ &\quad + \sum_{i_1=0}^{\min(N_{pl},R-1)} \frac{R!}{N_{NS}^{i_1}(R-i_1-1)!} \binom{N_{pl}}{i_1} e^{-\beta i_1 \Delta\varepsilon_{pl}} e^{-\beta\Delta\varepsilon}. \end{aligned} \tag{38}$$

In Fig. 5(b) we show the fold change of the simple repressor on the chromosome for three choices of operator strength $\Delta\varepsilon < \Delta\varepsilon_{pl}$, $\Delta\varepsilon = \Delta\varepsilon_{pl}$, and $\Delta\varepsilon > \Delta\varepsilon_{pl}$. As expected when the plasmid binding sites are very strong, we do not get a response in fold change of the simple repressor until all these sites have been filled. However, if the simple repressor binding site is stronger than the plasmid binding sites, this is no longer the case and we see an immediate decline in fold change when repressors are added. Even on a logarithmic plot the fold change shows a rich structure, which makes it an ideal candidate for experimental verification since we expect that this functional form can be easily detected above the

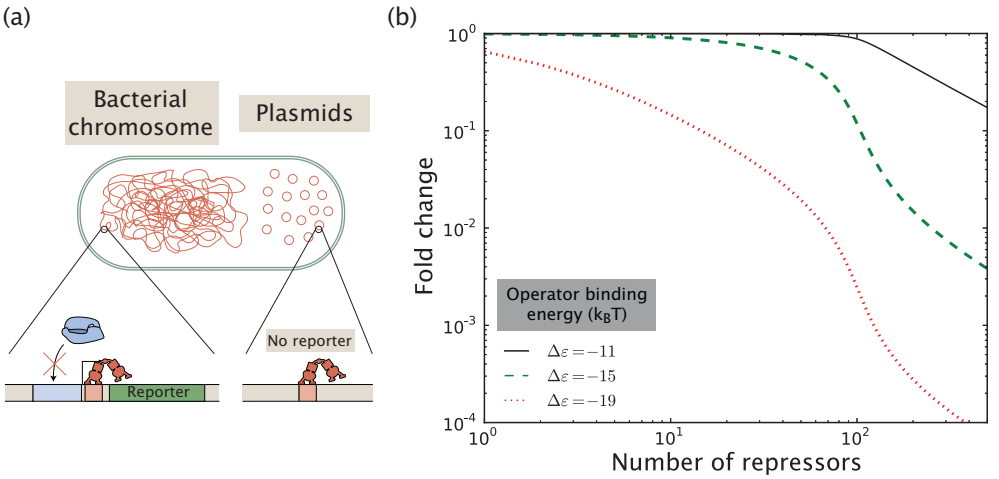


FIG. 5. (Color online) Effect of TF sequestration on fold change. (a) A repressor can bind to a reporter construct located in a single copy on the chromosome or to a binding site on a multicopy plasmid which leads to no gene expression. (b) Fold change of a simple repressor for different repressor binding site strengths, where the TF is subject to sequestration from 100 nonfunctional binding sites ($\Delta\varepsilon_{pl} = -15\, k_B T$). If the sequestration sites are much stronger than the simple repressor operator, the fold change remains constant until these sites have been filled. The RNAP promoter binding energy is assumed to be weak ($p \ll 1$), and the number of nonspecific sites $N_{NS} = 5 \times 10^6$.

intrinsic experimental noise in making such gene expression measurements.

Finally, for *independent* identical promoters, for example when the TFs are in great excess with respect to number of gene copies, the fold change for the set of promoters reduces to the fold change of an individual promoter. This intuitive result can be directly shown from Eqs. (32) + (33), using the fact that for N independent promoters, each with partition function Z, the total partition function is given by $Z^{\rm tot} = Z^N$. Let f_Z denote the fold change of a single promoter and $f_{Z^{\rm tot}}$ denote the fold change for N promoter copies, then

$$f_{Z^{\rm tot}} = \frac{\frac{p}{Z^{\rm tot}}\frac{\partial}{\partial p}Z^{\rm tot}}{\frac{p}{Z^{\rm tot}_{F=0}}\frac{\partial}{\partial p}Z^{\rm tot}_{F=0}} = \frac{(Z_{F=0})^N\,\frac{\partial}{\partial p}Z^N}{Z^N\,\frac{\partial}{\partial p}(Z_{F=0})^N}$$
$$= \frac{\frac{p}{Z}\frac{\partial}{\partial p}Z}{\frac{p}{Z_{F=0}}\frac{\partial}{\partial p}Z_{F=0}} = f_Z. \tag{39}$$

This equality, $f_{Z^{\rm tot}} = f_Z$, greatly simplifies calculating the fold change of the promoters.

VI. TRANSCRIPTIONAL CORRELATION

There are many reasons why expression of different genes might be correlated [77–79]. One obvious example is if a gene A regulates another gene B, then random intrinsic fluctuations in A will affect the expression of B (with a time delay), resulting in correlated expression of the two genes. For genes without direct regulatory connections, such random fluctuations due to intrinsic noise do not lead to correlated expression. Extrinsic noise, on the other hand, refers to fluctuations which affect the expression of *both* A and B simultaneously; this includes "global noise" such as fluctuating number of RNAP molecules or cell size, which leads to a positive correlation in transcription rates of the two genes. Another example of extrinsic noise, which we study in more depth in Sec. VII B, is fluctuations in TF copy number if A and B are regulated by the same TF.

In addition to these mechanisms we predict that promoter entanglement due to TF titration constitutes another source of correlation in transcription rates for genes regulated by the same TFs. Quantifying this effect is the topic of this section.

A. Toy model of transcriptional correlation

To develop intuition for the correlation in transcription from different promoters due to promoter entanglement, we first consider a hypothetical system of two unregulated promoters (P_A, P_B), transcribed by a single RNAP molecule $(P = 1)$. This system can be found in three different states: no promoter bound by RNAP, P_A bound by RNAP, or P_B bound by RNAP. Since the single RNAP molecule can only bind to one of the promoters at a time, transcription of the two promoters will become anticorrelated.

Let A, B denote the number (0 or 1) of RNAP bound to promoters P_A and P_B, respectively. These two random variables are correlated with the Pearson correlation coefficient,

$$\rho_{\rm corr} = \frac{\langle (A-\bar{A})(B-\bar{B})\rangle}{\sqrt{\langle (A-\bar{A})^2\rangle\langle (B-\bar{B})^2\rangle}} \tag{40}$$
$$= \frac{\langle (A-\bar{A})(B-\bar{B})\rangle}{\langle (A-\bar{A})^2\rangle}. \tag{41}$$

For the sake of simplicity we assume that the two promoters P_A, P_B have the same strength, and hence in Eq. (41) we set $\langle (A-\bar{A})^2\rangle = \langle (B-\bar{B})^2\rangle$. Let p_0 and $p_A = p_B$ denote the probabilities of the three states listed above. In terms of these probabilities the correlation coefficient translates to

$$\rho_{\rm corr} = -\frac{p_A}{1-p_A}, \tag{42}$$

which is plotted as a function of p_A in Fig. 6. When the promoters are very strong P_A *or* P_B will always be bound by RNAP ($p_A = p_B = \frac{1}{2}$); hence, knowledge of the state of one promoter is sufficient to tell the state of the other promoter ($\rho_{\rm corr} = -1$). However, when the promoters are weak, at most times both promoters are empty and the correlation between the promoters will be weak.

These results can be framed in terms of the familiar partition functions used throughout the paper. We now consider a statistical mechanical model of RNAP binding. The partition function for the two-promoter system is given by

$$Z = 1 + \frac{1}{N_{NS}}e^{-\beta\Delta\varepsilon_A} + \frac{1}{N_{NS}}e^{-\beta\Delta\varepsilon_B} \tag{43}$$
$$= 1 + \frac{2}{N_{NS}}e^{-\beta\Delta\varepsilon}, \tag{44}$$

where we again assume that both promoters have the same binding energy $\Delta\varepsilon = \Delta\varepsilon_A = \Delta\varepsilon_B$. The probability p_A for promoter A to be in the bound state is then given by

$$p_A = \frac{\frac{1}{N_{NS}}e^{-\beta\Delta\varepsilon}}{Z}. \tag{45}$$

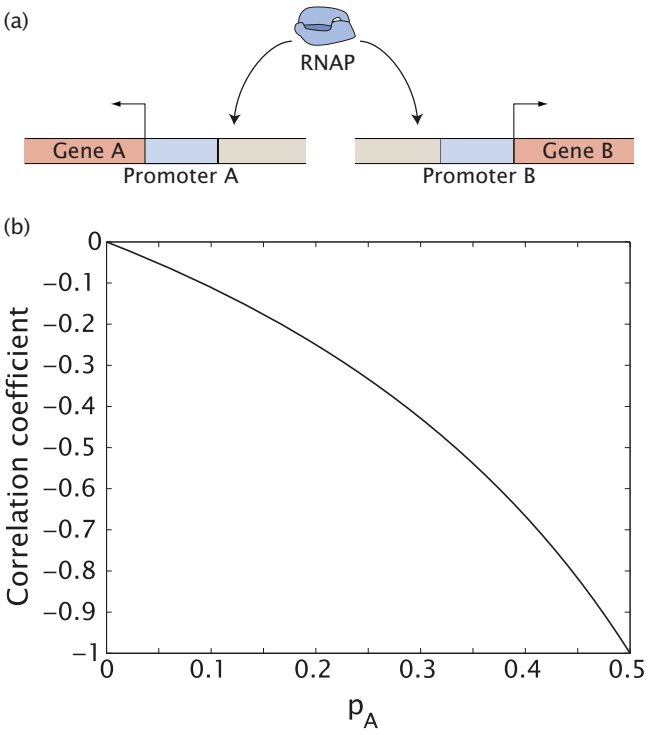


FIG. 6. (Color online) Correlation coefficient between transcription rates of two equally strong promoters P_A, P_B for a single RNAP molecule ($P = 1$), as a function of probability $p_A = p_B$ of one of the promoters being bound.

Plugging p_A back into the correlation coefficient [Eq. (42)] gives the transcriptional correlation as a function of promoter strength, namely,

$$\rho_{\text{corr}} = -\frac{1}{1 + N_{NS} e^{\beta \Delta \varepsilon}}. \tag{46}$$

These results are intended to illustrate how the correlations will be computed in the more general case considered next.

B. General theory

As reported in Sec. V, a state with i specifically bound RNAP molecules to a certain promoter type has a statistical weight of the form $\propto p^i$ with $p = \frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_{pd}}$. This weight generalizes for a set of N different promoter types to the form $\propto p_1^{i_1} \cdots p_N^{i_N}$. Using this observation it is easy to derive the statistical moments for promoter occupancies

$$\langle i_1, \ldots, i_m \rangle \equiv \frac{1}{Z^{\text{tot}}} p_{i_1} \frac{\partial}{\partial p_{i_1}} \cdots p_{i_m} \frac{\partial}{\partial p_{i_m}} Z^{\text{tot}}, \quad (1 \leqslant i_j \leqslant N, \forall\, j), \tag{47}$$

On the left hand side we use $\langle i_1, \ldots, i_m \rangle$ as a shorthand notation for the expectation value of the product of number of RNAP simultaneously bound to the promoters specified by the indices $i_1, \ldots, i_m$. For two promoter types ($N = 2$) the Pearson correlation coefficient can be expressed in terms of the partition function as

$$\rho_{i_1 i_2} = \frac{\langle (i_1 - \bar{i}_1)(i_2 - \bar{i}_2) \rangle}{\sqrt{\langle (i_1 - \bar{i}_1)^2 \rangle \langle (i_2 - \bar{i}_2)^2 \rangle}} \tag{48}$$

$$= \frac{p_1 p_2 \frac{\partial}{\partial p_1} \frac{\partial}{\partial p_2} \ln Z^{\text{tot}}}{\sqrt{\left[\left(p_1 \frac{\partial}{\partial p_1}\right)^2 \ln Z^{\text{tot}}\right]\left[\left(p_2 \frac{\partial}{\partial p_2}\right)^2 \ln Z^{\text{tot}}\right]}}. \tag{49}$$

Here $\bar{i}_{1,2}$ denotes the occupancy ($\langle i_{1,2} \rangle$) for promoters 1,2, respectively.

C. Two anticorrelated genes

Let us now study the specific example of transcriptional correlation for a system with two genes located together on N_{pl} identical plasmids, where both genes are regulated by the same A activating TFs (*activators*), as shown in Figs. 7 and 8. The transcription rates for the two genes will be anticorrelated, because when one gene is highly activated there are fewer activator molecules left to also activate the other gene. When there are no activators ($A = 0$) transcription of the two genes is clearly independent, but this is also true if $A \gg N_{pl}$ because the number of activators available for promoter binding will be essentially constant. Hence, we expect anticorrelation of transcription rates between the two genes to have a peak in magnitude when the number of activators is roughly comparable to the number of plasmids ($A \approx N_{pl}$).

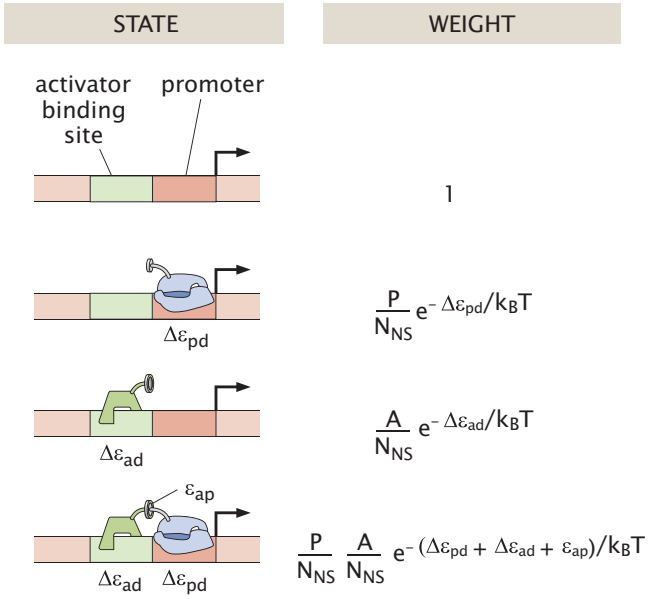


FIG. 7. (Color online) States and weights for the simple activation regulatory motif [2].

There are four different states for a simple activator promoter architecture (see Fig. 7): empty state, activator bound, promoter bound (by RNAP), and activator and promoter bound. The last state has a (negative) activator-RNAP interaction energy ε_{ap}, used by the activator to "recruit" RNAP to the promoter. For simplicity we assume that the two genes have the same operator strength and promoter strengths, and hence the same partition function,

$$Z = 1 + \frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_{pd}} + \frac{A}{N_{NS}} e^{-\beta \Delta \varepsilon_{ad}} + \frac{A}{N_{NS}} \frac{P}{N_{NS}} e^{-\beta(\Delta \varepsilon_{pd} + \Delta \varepsilon_{ad} + \varepsilon_{ap})}. \tag{50}$$

We use Eq. (20) to find the partition function for the two genes on one plasmid copy, then Eq. (25) to find the partition function for multiple plasmid copies. Once we have the total partition function we can calculate the transcriptional correlation using Eq. (49).

In Fig. 9(a) we show the transcriptional correlation of the system as a function of activator copy number for different numbers of plasmids. As expected, the correlation exhibits a peak when the number of activators is similar to the number of plasmids (peak value $\rho \approx -0.8$). As the number of activators outgrows the total number of binding sites ($2N_{pl}$)

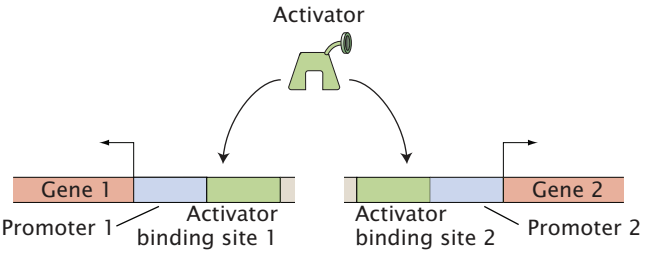


FIG. 8. (Color online) Two simple activators regulated by the same TF.

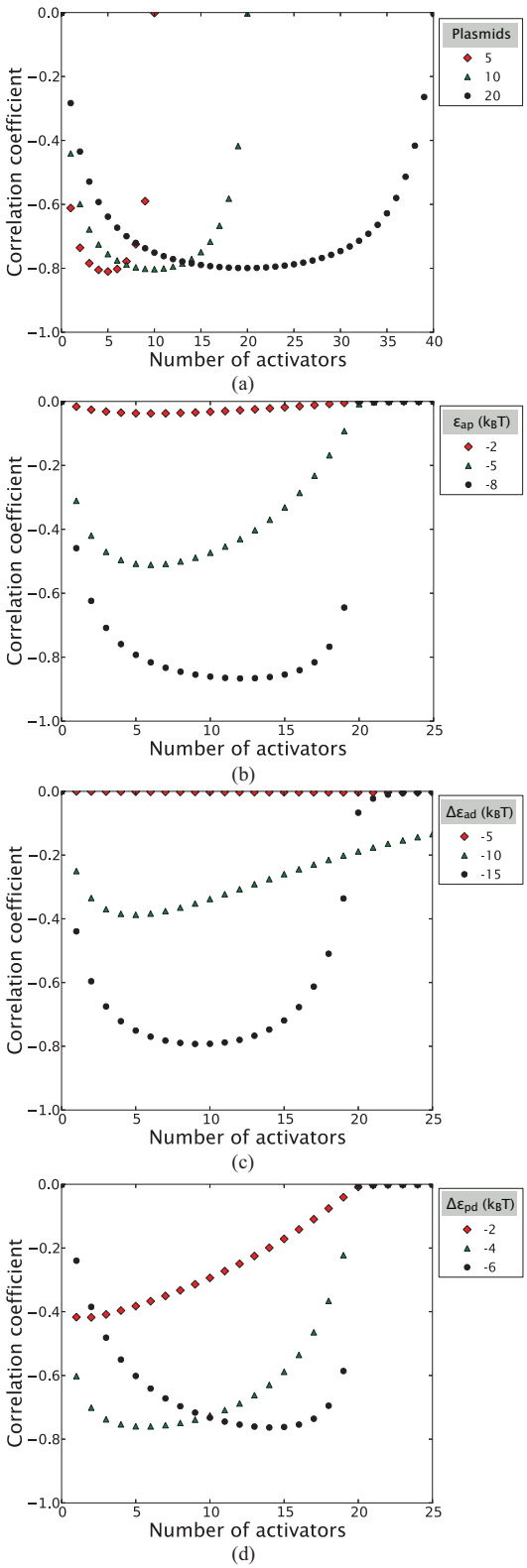


FIG. 9. (Color online) Correlation coefficient between transcription rates of two positively regulated genes on a plasmid, as a function of (a) number of plasmids, (b) RNAP-activator interaction energy ε_{ap}, (c) activator operator strength $\Delta\varepsilon_{ad}$, and (d) promoter strength $\Delta\varepsilon_{pd}$. For fixed parameter values we use number of nonspecific sites $N_{NS} = 5 \times 10^6$, 10 plasmids, operator strength $\Delta\varepsilon_{ad} = -17.3\, k_B T$, promoter strength $\Delta\varepsilon_{pd} = -5\, k_B T$, and interaction energy between TF and RNAP $\varepsilon_{ad} = -7\, k_B T$.

the correlation dies off rapidly, at least when the activator operators are strong. In Figs. 9(b) and 9(c) we show how the correlation depends on the RNAP-activator interaction energy ε_{ap} and the binding site strength of the activator $\Delta\varepsilon_{ad}$. As expected, the transcriptional correlation between the two genes increases in magnitude when these interactions are stronger (more negative). In Fig. 9(d) we show how the transcriptional correlation depends on the promoter binding strength. Weak promoters only recruit RNAP when bound by activators. With just one single activator molecule this system becomes similar to the toy model of Sec. VI A, and we see a fast response in correlation. Strong promoters can recruit RNAP well even without activators and hence it takes more of them before we see any substantial effect in fold change and correlation.

A necessary condition for the transcriptional correlation effect to be experimentally observable is that TFs stay bound to their binding sites a sufficient amount of time to avoid rapid switching between different promoter states. For example, if mRNA levels are measured at fixed time points (e.g., using FISH), TFs would need to stay bound longer to the operators than the mRNA lifetime. To see this, consider the opposite extreme when the mRNA lifetime is very long (or say infinite), then the observed mRNA expression merely corresponds to an averaged production over every possible promoter state and no effect of transcriptional correlation could be observed. On the other hand, if the mRNA lifetime is much shorter than the TF binding time, the observed mRNAs were likely produced from the same promoter state (or configuration of TFs). This condition is met, e.g., in the case of LacI regulating lacZ, where the TF on average stays bound approximately 10 min to the strongest operator in 37 °C [80], whereas the lacZ mRNA lifetime is only about 2 min [81]. Even when this condition is not met one might still be able to detect the transcriptional correlation effect by measuring mRNA or protein production during a relatively short time interval from fluorescence time traces, as long as the uncertainty in production (and maturation) time of mRNA or proteins is small compared to the TF binding time.

Another condition for the transcriptional correlation effect to be biologically relevant is that extrinsic noise sources, like fluctuations in plasmid or TF copy number, do not have a stronger impact on gene expression than the correlation effect due to TF titration. This matter is discussed at more length in Sec. VII B.

VII. STATISTICALLY DISTRIBUTED TF AND PROMOTER COPY NUMBERS

In a cell the number of TFs and promoter copies are, because of inherent stochasticity, not fixed but rather fluctuating according to a statistical distribution. These distributions vary greatly, with examples ranging from the tightly regulated low-copy F-plasmid [82], to the wide distribution of gene copies produced at viral infections [83]. In this section we see how the predicted fold change and transcriptional correlation are affected by fluctuations in promoter copy number and TF copy number. Given the wide range of possible copy number distributions, our goal is not necessarily to model any particular biological system but rather provide a general framework which allows us to compute the fold change and transcriptional

correlation for any given such distribution, as well as illustrate this effect on our previously derived results in a few specific cases.

A. Fold change

In Sec. V we show that the fold change of a promoter architecture can depend sensitively on the number of repressors R when this number is comparable to the number of promoter copies N (see Fig. 3). We now see how this sensitivity is affected when the number of repressors R or promoter copies N are not fixed but rather fluctuating according to a probability distribution $P(R,N)$. In this case the RNAP occupancy [Eq. (32)] to the promoters needs to be replaced by the expectation value with respect to $P(R,N)$,

$$\langle \text{Occupancy} \rangle_{P(R,N)} = \sum_{R,N} P(R,N)\,\text{Occupancy}\,(R,N),$$

which we can consequently insert into the definition of fold change,

$$f = \frac{\langle \text{Occupancy} \rangle_{P(R,N)}}{\langle \text{Occupancy}\,(R=0) \rangle_{P(N)}} \tag{51}$$

$$= \frac{\sum_{R,N} P(R,N)\,\text{Occupancy}\,(R,N)}{\sum_{N} P(N)\,\text{Occupancy}\,(R=0,N)}. \tag{52}$$

We can simplify the last line [Eq. (52)] by noticing that for $R=0$ the promoters are independent and hence the occupancy must be proportional to N

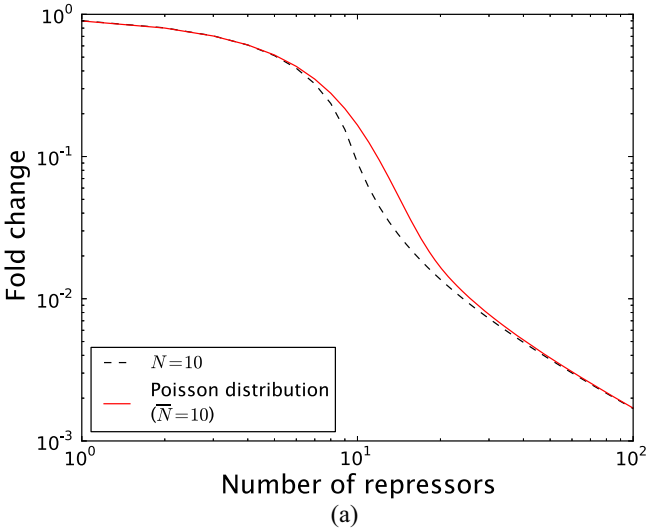
$$f = \frac{\sum_{R,N} P(R,N)\,\text{Occupancy}\,(R,N)}{\langle N \rangle_{P(N)}\,\text{Occupancy}\,(R=0,N=1)}. \tag{53}$$

As an example in Fig. 10 we investigate the effect of replacing the promoter copy number with a Poisson distribution in the simple repression (Sec. III A) and exclusive looping repression (Sec. III C) architectures. A set of simple repressors will only be effectively repressed when all the promoter copies are inhibited; therefore, the steep decline in fold change around $N \approx R$ will now be shifted up to higher repressor copy number. For the exclusive looping architecture we note that a trough is still clearly visible but less deep and slightly widened (at half peak depth) compared to the case with fixed promoter copy number. If we in Fig. 10 instead replace the repressor copy number by a Poisson distribution and keep the promoter copy number fixed the fold change will look close to identical (result not shown).

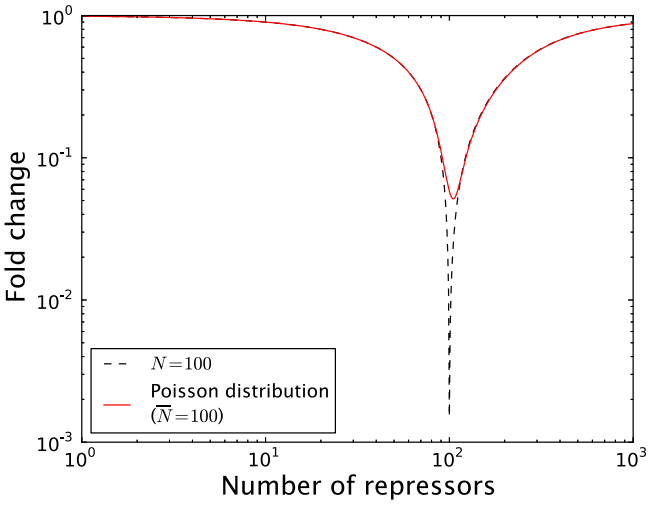
B. Transcriptional correlation

Fluctuations in TF copy number constitute an extrinsic form of noise that affects the transcription rate of all genes regulated by the TF. In this section we show that such fluctuations, when large enough, can hide the effect of transcriptional correlation due to TF titration. To include extrinsic noise into our calculation of transcriptional correlation [Eq. (49)] between two genes we compute the Pearson correlation coefficient using weighted moments,

$$\rho_{i_1 i_2} = \frac{\langle i_1 i_2 \rangle_{P(F)} - \langle i_1 \rangle_{P(F)} \langle i_2 \rangle_{P(F)}}{\sqrt{\langle (i_1 - \langle i_1 \rangle_{P(F)})^2 \rangle_{P(F)} \langle (i_2 - \langle i_2 \rangle_{P(F)})^2 \rangle_{P(F)}}}, \tag{54}$$



(a)



(b)

FIG. 10. (Color online) (a) Fold change in the simple repression architecture for fixed ($N = 10$) or Poisson distributed (mean $\bar{N} = 10$) promoter copy number. (b) Fold change in the exclusive looping repression architecture for fixed ($N = 100$) or Poisson distributed (mean $\bar{N} = 100$) promoter copy number. For these plots we use operator binding energy $-17.3\,k_BT$, looping energy $10\,k_BT$, and number of nonspecific sites $N_{NS} = 5 \times 10^6$. The RNAP promoter binding energy is assumed to be weak.

where the expectation value $\langle \cdot \rangle_{P(F)} \equiv \sum_{i=0}^{\infty} P(F=i) \langle \cdot \rangle_{F=i}$ is evaluated over the distribution of TFs. In Fig. 11 we use this formula to show how TF fluctuations affect the transcriptional correlation of the particular system of two genes activated by the same TF studied in Sec. VI C. In this case we use, for illustration purposes, a Gaussian distribution which allows us to vary the distribution width and see what effect is has on transcriptional correlation. Promoter entanglement and extrinsic noise will have an opposite effect on transcriptional correlation, and their relative strengths will determine the resulting sign of the correlation coefficient. For $A > 2N_{pl}$ there is no promoter entanglement but a positive correlation due to TF fluctuations remains until the average number of

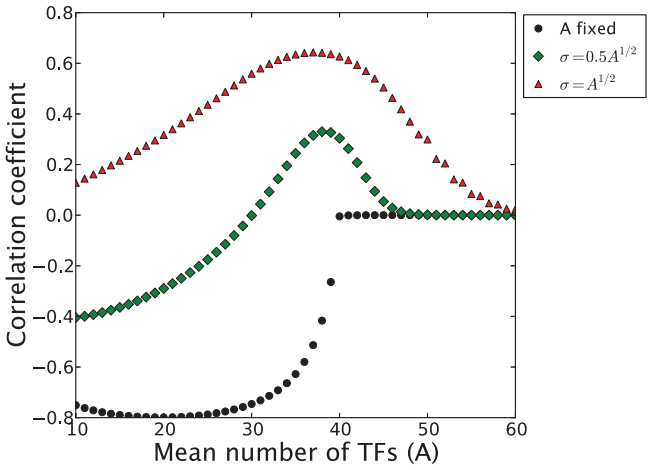


FIG. 11. (Color online) Correlation coefficient between transcription rates of two positively regulated genes located on 20 plasmids, as a function of number of TFs. Three different Gaussian TF copy number distributions are considered with standard deviations $\sigma = 0, \frac{1}{2}\sqrt{A}, \sqrt{A}$. TF fluctuations constitute extrinsic noise, affecting expression of both genes, that hides the anticorrelation in transcription rates due to promoter entanglement. As parameter values we choose number of RNAP $P = 1000$, nonspecific sites $N_{NS} = 5 \times 10^6$, 20 plasmids, operator strength $\Delta\varepsilon_{ad} = -17.3\, k_B T$, promoter strength $\Delta\varepsilon_{pd} = -5\, k_B T$, and interaction energy between TF and RNAP $\varepsilon_{ad} = -7\, k_B T$.

activators is so high that essentially all operators will be occupied.

As the TF copy number increases the TFs will distribute themselves more and more evenly among its targets, and the transcriptional correlation due to TF titration will have a smaller impact on gene expression. We therefore expect transcriptional correlation due to TF titration to be most relevant when the TF copy number is low and extrinsic noise limited. These conditions can be somewhat relaxed through recent advances in molecular biology, for example, cells with TFs labeled by a fluorescent reporter can be sorted by fluorescence to limit the effect of TF fluctuations on transcriptional correlation, hence allowing precision tests of the thermodynamic model.

VIII. VERIFYING THE THERMODYNAMIC MODEL OF TF TITRATION USING GILLESPIE SIMULATIONS

To examine the validity of the thermodynamic calculations, we use Gillespie simulations [84] to predict fold change and correlation in transcription rates. Although this is computationally more onerous than the thermodynamic models used throughout the paper, it has the benefit of simplicity, requiring only knowledge of the gene/TF copy numbers and allowed reactions. Consequently, the intricate details of TF binding combinatorics are given to us "for free."

To demonstrate the Gillespie algorithm we consider, as an example, free repressors (R) (un)binding to empty gene promoters (G) to form repressor-gene complexes (GR) through the reactions

$$G + R \underset{k_R^{\text{off}}}{\overset{k_R^{\text{on}}}{\rightleftarrows}} \text{GR}. \tag{55}$$

Here we assume, as in the law of mass action, that the total rate of repressor association is proportional to both the number of free repressors and empty promoters. The normalized rate parameter k_R^{on} gives number of associations per free repressor, per empty promoter, per time unit. Similarly the normalized disassociation rate parameter k_R^{off} gives number of disassociations per repressor-gene complex, per time unit. These rate parameters will depend on operator strength and number of competing nonspecific binding sites (N_{NS}), but not molecular numbers of the species involved. Notice that since the repressors are assumed to be always bound on DNA we do not consider cell volume, or cytosolic repressor/gene concentration, as parameters of our model. However, cell volume will have an indirect effect on above rate parameters through its influence on the nonspecific free energy of binding a repressor to DNA.

In the first step of the Gillespie algorithm we calculate the total accumulated reaction rate, $G \times R \times k_R^{\text{on}} + \text{GR} \times k_R^{\text{off}}$, for both reactions and then draw a random time step at which the next reaction will take place from an exponential distribution, with mean equal to the inverse of this rate. The decision which of the two reactions should be chosen is random but weighted by the accumulated rate for each reaction $G \times R \times k_R^{\text{on}}$ vs $\text{GR} \times k_R^{\text{off}}$. If the repressor binding reaction is chosen we update the corresponding state variables according to $G \to G - 1$, $R \to R - 1$, and $\text{GR} \to \text{GR} + 1$ (analogously for repressor unbinding). Notice that G, R, and GR are *discrete* quantities, not continuous concentrations. By repeating this procedure over and over we acquire time traces for G, R, and GR, which can be used to compute the (time averaged) occupancy of repressors to genes, fluctuations in G, R, and GR, etc. To compute fold change, a quantity of central importance throughout this work, we use Gillespie's method to find the average number of promoters bound by RNAP, with and without TFs present.

In order to connect the stochastic model with our thermodynamic calculations much effort in this section is dedicated to finding mathematical relations between the stochastic model rate constants and corresponding thermodynamic free energy parameters. This matter is alleviated by the fact that the rate constants are independent of gene copy number, TF copy number, and RNAP copy number, which allows us to determine the rates using stripped-down version of the full promoter architectures.

A. Simple repression

To determine the rate parameters corresponding to repressor (un)binding in the simple repression architecture (Sec. III A) we consider a minimal system with a single promoter ($N = 1$) and no RNAP. In this system there are only two states (see Fig. 12): repressor bound (state B) and empty promoter (state 0), with dynamics described by the following master equation

$$\frac{dP(B)}{dt} = Rk_R^{\text{on}} P(0) - k_R^{\text{off}} P(B). \tag{56}$$

Here $P(B), P(0)$ correspond to the respective state probabilities $[P(B) + P(0) = 1]$. In equilibrium there is no net

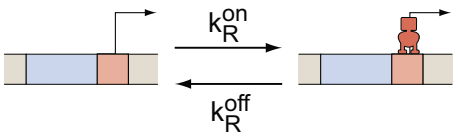


FIG. 12. (Color online) States and transition rates in a the simple repression architecture with no RNAP present. The rates correspond to "per molecule" rates, i.e., the total probability flux into the right, repressed state, is given by $Rk_R^{\text{on}}P(0)$.

probability flux between the two states, or mathematically,

$$Rk_R^{\text{on}}P(0) = k_R^{\text{off}}P(\text{B}) \Longrightarrow \frac{k_R^{\text{on}}}{k_R^{\text{off}}} = \frac{1}{R}\frac{P(B)}{[1-P(B)]}. \tag{57}$$

In the thermodynamic model we find the probability $P(B)$ from the partition function computed in Eq. (7),

$$P(B) = \frac{\frac{R}{N_{NS}}e^{-\beta\Delta\varepsilon_{rd}}}{1+\frac{R}{N_{NS}}e^{-\beta\Delta\varepsilon_{rd}}}, \tag{58}$$

which gives us a simple expression for the ratio between the rates

$$\frac{k_R^{\text{on}}}{k_R^{\text{off}}} = \frac{1}{N_{NS}}e^{-\beta\Delta\varepsilon_{rd}}. \tag{59}$$

This argument holds equally well for RNAP and we find

$$\frac{k_{\text{RNAP}}^{\text{on}}}{k_{\text{RNAP}}^{\text{off}}} = \frac{1}{N_{NS}}e^{-\beta\Delta\varepsilon_{pd}}. \tag{60}$$

In equilibrium each reaction will be balanced by its reverse reaction; hence, the final state probabilities can only depend on these ratios, also in the case of multiple gene copies.

We are now ready to apply Gillespie's method to simulate the full simple repression promoter architecture [see Fig. 1(a)], using the following set of reactions:

$$G + R \underset{k_R^{\text{off}}}{\overset{k_R^{\text{on}}}{\rightleftarrows}} \text{GR}, \quad G + P \underset{k_{\text{RNAP}}^{\text{off}}}{\overset{k_{\text{RNAP}}^{\text{on}}}{\rightleftarrows}} \text{GP}. \tag{61}$$

Here we use the notation: G (empty promoter), R (free repressor), P (free RNAP), GR (promoter bound by repressor), and GP (promoter bound by RNAP). From the resulting simulation time trace we can compute the average number of RNAP-promoter complexes (GP), which we use as a proxy for gene expression. By repeating the simulation with no repressors ($R = 0$) we can then determine the fold change.

Figure 13 shows a precise agreement in fold change between Gillespie simulations and thermodynamic theory, as one would expect.

B. Repression with looping

In the case of repression by looping (Sec. III B) we not only need to take the repressor (un)binding rates into account but also the rate of DNA (un)looping between the main and auxiliary binding site. To find the rate constants corresponding to the thermodynamic free energy parameters we consider a simplified system with a single promoter, no RNAP and only three states: empty promoter (state 0), main operator bound (state M), and looped state (state L). The transitions

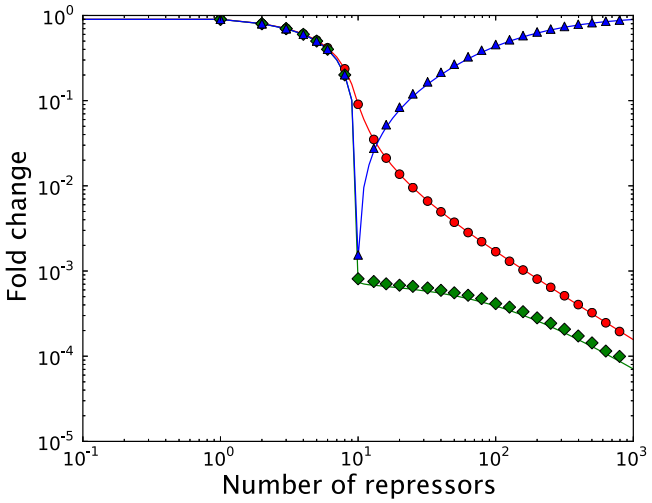


FIG. 13. (Color online) Fold change as a function of repressor copy number in the simple repression (●), repression with looping (◆), and repression exclusively due to looping (▲) promoter architecture, for $N = 10$ promoter copies. Solid lines correspond to thermodynamic model predictions and markers Gillespie simulated data. Here we use the parameters: $k_R^{\text{on}} = 1.0$, $k_R^{\text{off}} = 0.15$ (simple repression), $k_R^{\text{off}} = 0.075$ (looping), $k_{\text{RNAP}}^{\text{on}} = 3.0 \times 10^{-5}$, $k_{\text{RNAP}}^{\text{off}} = 1$, $k_{\text{loop}} = 1$, and $k_{\text{unloop}} = 6.8 \times 10^{-4}$ in arbitrary inverse time units, chosen according to Eqs. (59), (60), (64). The standard deviations, acquired from three separate runs, are smaller than the marker size. Since the rates only enter as *ratios* in the state probabilities we use this freedom to set larger of the two rates to 1. As initial condition we set all promoters to the empty state, $G = 10$, RNAP copy number $P = 1000$, and repressor copy number R indicated by the x axis.

between these states are illustrated in Fig. 14. We consider the state with only the auxiliary operator bound by a repressor to be forbidden. This does not affect the rate constants for repressor (un)binding or DNA loop formation as compared to the full repression with looping architecture, but makes the mathematical derivations more straightforward. The detailed balance equations for this system are

$$\begin{aligned} RP(0)k_R^{\text{on}} &= P(M)k_{\text{off}}, \\ P(M)k_{\text{loop}} &= P(L)k_{\text{unloop}}, \\ P(0) + P(M) + P(L) &= 1, \end{aligned} \tag{62}$$

which can be easily solved for $P(0)$, $P(M)$, and $P(L)$.

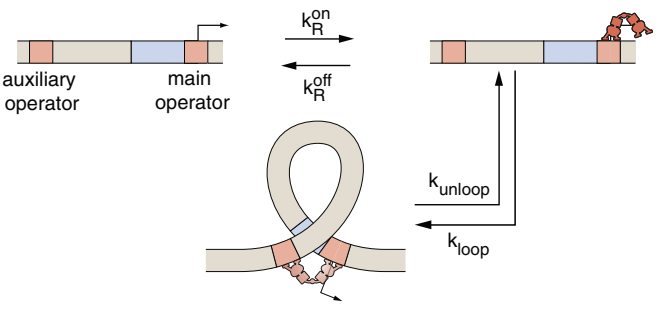


FIG. 14. (Color online) States and transition rates in a simplified version of the repression with looping promoter architecture, with no RNAP and where the auxiliary operator is not allowed to be bound individually.

On the other hand, the state probabilities for this system can be derived using the statistical mechanical framework, similar to the procedure used in Sec. III B

$$P(0) = \frac{1}{1 + \frac{2R}{N_{NS}} e^{-\beta \Delta\varepsilon_{rd}} + \frac{2R}{N_{NS}} e^{-\beta(2\Delta\varepsilon_{rd} + \Delta F_{\rm loop})}},$$
$$P(M) = \frac{\frac{2R}{N_{NS}} e^{-\beta \Delta\varepsilon_{rd}}}{1 + \frac{2R}{N_{NS}} e^{-\beta \Delta\varepsilon_{rd}} + \frac{2R}{N_{NS}} e^{-\beta(2\Delta\varepsilon_{rd} + \Delta F_{\rm loop})}}, \quad (63)$$
$$P(L) = \frac{\frac{2R}{N_{NS}} e^{-\beta(2\Delta\varepsilon_{rd} + \Delta F_{\rm loop})}}{1 + \frac{2R}{N_{NS}} e^{-\beta \Delta\varepsilon_{rd}} + \frac{2R}{N_{NS}} e^{-\beta(2\Delta\varepsilon_{rd} + \Delta F_{\rm loop})}}.$$

Here we assume that the main and auxiliary operators have the same binding energy $\Delta\varepsilon_{rd}$. Equating the state probabilities found in the thermodynamic model with those from Eq. (62) allows us to express the (un)binding and (un)looping rates in term of the free energies $\Delta\varepsilon_{rd}, \Delta F_{\rm loop}$

$$\frac{k_R^{\rm on}}{k_R^{\rm off}} = \frac{2}{N_{NS}} e^{-\beta \Delta\varepsilon_{rd}}, \quad \frac{k_{\rm loop}}{k_{\rm unloop}} = e^{-\beta(\Delta\varepsilon_{rd} + \Delta F_{\rm loop})}. \quad (64)$$

Notice that, by assuming that the two TF operators have the same binding energy we only need one set of (un)looping rates. We use these rates to apply Gillespie's method on the full repression with looping architecture, where all states in Fig. 1(b) are allowed, using the reaction scheme

$$\begin{gathered} G + R \underset{k_R^{\rm off}}{\overset{k_R^{\rm on}}{\rightleftarrows}} {\rm GR}_M, \quad G + R \underset{k_R^{\rm off}}{\overset{k_R^{\rm on}}{\rightleftarrows}} {\rm GR}_A, \\ {\rm GR}_M \underset{k_{\rm unloop}}{\overset{k_{\rm loop}}{\rightleftarrows}} {\rm GL}, \quad {\rm GR}_A \underset{k_{\rm unloop}}{\overset{k_{\rm loop}}{\rightleftarrows}} {\rm GL}, \\ {\rm GR}_M + R \underset{k_R^{\rm off}}{\overset{k_R^{\rm on}}{\rightleftarrows}} {\rm GR}_{MA}, \quad {\rm GR}_A + R \underset{k_R^{\rm off}}{\overset{k_R^{\rm on}}{\rightleftarrows}} {\rm GR}_{MA}, \\ {\rm GR}_A + P \underset{k_{\rm RNAP}^{\rm off}}{\overset{k_{\rm RNAP}^{\rm on}}{\rightleftarrows}} {\rm GPR}_A, \quad G + P \underset{k_{\rm RNAP}^{\rm off}}{\overset{k_{\rm RNAP}^{\rm on}}{\rightleftarrows}} {\rm GP}, \end{gathered} \quad (65)$$

where we use the following notation: G (empty promoter), R (free repressor), P (free RNAP), ${\rm GR}_M$ (main operator bound), ${\rm GR}_A$ (auxiliary operator bound), ${\rm GR}_{MA}$ (main and auxiliary operator bound), GL (looped conformation), ${\rm GPR}_A$ (auxiliary operator bound by TF and promoter by RNAP), GP (promoter bound by RNAP).

In Fig. 13 we find that our statistical mechanical predictions for fold change are precisely replicated by Gillespie simulations. To achieve the level of precision shown in the figure required around 1 h of Gillespie simulations for 30 data points, compared to the analytical framework which allowed us to compute the fold change for 1000 data points in less than 1 s.

C. Repression exclusively due to looping

For repression exclusively due to looping (Sec. III C) we use the same rate parameters as found in Eqs. (60) + (64), but allow RNAP to bind all states except the looped state (see Fig. 1). This means we need to add the following reactions to the scheme in (65)

$${\rm GR}_M + P \underset{k_{\rm RNAP}^{\rm off}}{\overset{k_{\rm RNAP}^{\rm on}}{\rightleftarrows}} {\rm GPR}_M, \quad {\rm GR}_{MA} + P \underset{k_{\rm RNAP}^{\rm off}}{\overset{k_{\rm RNAP}^{\rm on}}{\rightleftarrows}} {\rm GPR}_{MA}, \quad (66)$$

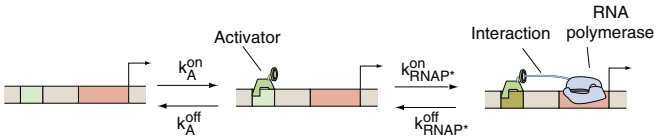


FIG. 15. (Color online) Simple activation promoter architecture in the weak promoter approximation, neglecting RNAP binding to the empty promoter.

where we use the notation ${\rm GPR}_M$ (main operator bound by TF and promoter by RNAP) and ${\rm GPR}_{MA}$ (main plus auxiliary bound by TF and promoter by RNAP).

In Fig. 13 we compare the fold change predicted by the thermodynamic model with Gillespie simulations and again find them to be in precise agreement.

D. Transcriptional correlation

In Sec. VI it was shown that under certain conditions the transcription rates of two genes can be correlated and we used the simple activation promoter architecture as a case study. To find the rate constants that correspond to the thermodynamic model free energy parameters for this promoter architecture we solve the detailed balance equations resulting from Fig. 15

$$\begin{aligned} {\rm AP}(0) k_A^{\rm on} &= P(A) k_{\rm off}, \\ P(A) P k_{\rm RNAP^*}^{\rm on} &= P({\rm AP}) k_{\rm RNAP^*}^{\rm off}, \\ P(0) + P(A) + P({\rm AP}) &= 1, \end{aligned} \quad (67)$$

where we use the following notation: empty promoter (state 0), activator bound to promoter (state A), activator and RNAP bound to promoter (state AP), and $k_{\rm RNAP^*}^{\rm off}, k_{\rm RNAP^*}^{\rm on}$ refer to RNAP (un)binding rate when the promoter is *already* bound by an activator. For mathematical convenience we invoke the weak promoter approximation and neglect the state with RNAP bound to an empty promoter.

In the thermodynamic model we can write down the corresponding state probabilities (see notation Fig. 7)

$$P(0) = \frac{1}{1 + \frac{A}{N_{NS}} e^{-\beta \Delta\varepsilon_{rd}} + \frac{AP}{N_{NS}^2} e^{-\beta(\Delta\varepsilon_{ad} + \varepsilon_{ap})}},$$
$$P(A) = \frac{\frac{A}{N_{NS}} e^{-\beta \Delta\varepsilon_{rd}}}{1 + \frac{A}{N_{NS}} e^{-\beta \Delta\varepsilon_{rd}} + \frac{AP}{N_{NS}^2} e^{-\beta(\Delta\varepsilon_{ad} + \varepsilon_{ap})}}, \quad (68)$$
$$P({\rm AP}) = \frac{\frac{AP}{N_{NS}^2} e^{-\beta(\Delta\varepsilon_{ad} + \varepsilon_{ap})}}{1 + \frac{A}{N_{NS}} e^{-\beta \Delta\varepsilon_{rd}} + \frac{AP}{N_{NS}^2} e^{-\beta(\Delta\varepsilon_{ad} + \varepsilon_{ap})}}.$$

Equating the state probabilities in Eqs. (67) and (68) allows us to express the TF and RNAP (un)binding rate in terms of the thermodynamic model parameters

$$\begin{aligned} \frac{k_A^{\rm on}}{k_A^{\rm off}} &= \frac{1}{N_{NS}} e^{-\beta \Delta\varepsilon_{ad}}, \\ \frac{k_{\rm RNAP^*}^{\rm on}}{k_{\rm RNAP^*}^{\rm off}} &= \frac{1}{N_{NS}} e^{-\beta(\Delta\varepsilon_{pd} + \varepsilon_{ap})} = \frac{k_{\rm RNAP}^{\rm on}}{k_{\rm RNAP}^{\rm off}} e^{-\beta \varepsilon_{ap}}. \end{aligned} \quad (69)$$

Using these rates we can apply Gillespie's method to the system of two genes considered in Sec. VI C, described by

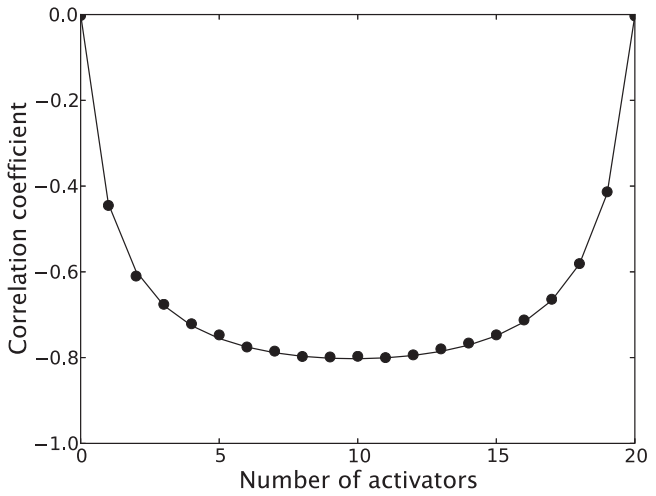


FIG. 16. Correlation coefficient between transcription rates of two positively regulated genes on a plasmid (copy number $N = 10$) as a function of activator copy number. The solid line corresponds to thermodynamic model prediction, and dots correspond to Gillespie simulated data. Here we use the parameters: $k_A^{\text{on}} = 1.0$, $k_A^{\text{off}} = 0.15$, $k_{\text{RNAP}}^{\text{on}} = 3.0 \times 10^{-5}$, $k_{\text{RNAP}}^{\text{off}} = 1$, $k_{\text{RNAP}^*}^{\text{on}} = 0.033$, and $k_{\text{RNAP}^*}^{\text{off}} = 1$ in arbitrary inverse time units, chosen according to Eqs. (59), (60), and (69). The standard deviations, acquired from three separate runs, are smaller than the marker size. Since the rates only enter as *ratios*, we use this freedom to set the larger of the two rates to 1. As initial condition we set all promoters to the empty state, $G_1 = G_2 = 10$, RNAP copy number $P = 1000$, and activator copy number A indicated by the x axis.

the reaction scheme

$$\begin{aligned} &G_1 + A \underset{k_A^{\text{off}}}{\overset{k_A^{\text{on}}}{\rightleftarrows}} \text{G}_1\text{A}, \quad G_2 + A \underset{k_A^{\text{off}}}{\overset{k_A^{\text{on}}}{\rightleftarrows}} \text{G}_2\text{A}, \\ &G_1 + P \underset{k_{\text{RNAP}}^{\text{off}}}{\overset{k_{\text{RNAP}}^{\text{on}}}{\rightleftarrows}} \text{G}_1\text{P}, \quad G_2 + P \underset{k_{\text{RNAP}}^{\text{off}}}{\overset{k_{\text{RNAP}}^{\text{on}}}{\rightleftarrows}} \text{G}_2\text{P}, \\ &G_1A + P \underset{k_{\text{RNAP}^*}^{\text{off}}}{\overset{k_{\text{RNAP}^*}^{\text{on}}}{\rightleftarrows}} \text{G}_1\text{AP}, \quad G_2A + P \underset{k_{\text{RNAP}^*}^{\text{off}}}{\overset{k_{\text{RNAP}^*}^{\text{on}}}{\rightleftarrows}} \text{G}_2\text{AP}. \end{aligned} \tag{70}$$

At each time step of the simulation the number of promoters of each type bound by RNAP is recorded, and using the time traces we can compute the correlation coefficient between the two quantities. Figure 16 again shows a precise agreement between our thermodynamic model and Gillespie simulations.

IX. CONCLUSION

In this work we have developed a general framework based on statistical mechanics to predict gene expression for systems with multiple genes or gene copies regulated by the same TFs. These kinds of systems arise in a multitude of biologically relevant circumstances. In particular, we have shown that when the number of TF binding sites is large enough to titrate the TFs, the predicted gene expression depends in a highly nontrivial way on the relative abundance of promoter and TF copy numbers. New data [7] on protein copy numbers in *E. coli* indicate that such titration might happen more often than previously thought. We have also quantitatively linked the effect of TF titration to correlation between transcription rates of different genes.

An advantage with the presented model is that quantities of interest, e.g., fold change or correlation in transcription rates, can be expressed analytically for a set of promoters explicitly in terms of the individual promoter architectures. This allows us to vary model parameters and TF copy number without the need of running thousands of time-consuming Gillespie simulations.

Recent advances in the field of molecular biology have made it possible to accurately measure and tune protein copy numbers in a cell [7,21,62,85], which provides an excellent opportunity to test the predictions presented here experimentally. This will indeed be the topic of an upcoming paper.[2]

ACKNOWLEDGMENTS

We wish to thank Robert Brewster and Franz Weinert for useful discussions. Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award No. R01 GM085286 and No. R01 GM085286-01S (M.R., H.H.G., R.P.), as well as National Institutes of Health Pioneer Award No. DP1 OD000217 (H.G.G., R.P.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

APPENDIX A: PARTITION FUNCTION FOR A SET OF PROMOTERS REGULATED BY MULTIPLE LOW-COPY TFS

One can easily show that the partition function derived in Eq. (20) for a set of promoters regulated by one TF type is valid also when the promoters are regulated by additional TFs, as long as these extra factors are not subject to titration effects and can be summed out together with RNAP in Eq. (21). However, in the case of regulation by multiple low-copy TFs the derivation needs to be generalized. To do this let us denote the different TFs by $F_1, \ldots, F_m$ and f_{n_j} the number of TFs of type $j \in \{1, \ldots, m\}$ bound to promoter $n \in \{1, \ldots, N\}$. By analogy to the treatment in Sec. IV A the total partition function is given by

$$Z^{\text{tot}} = \sum_{\substack{f_{n_j},\, \forall\, n,j \\ \sum_n f_{n_j} \leqslant F_j, \forall\, j}} \left[\prod_{j=1}^{m} \frac{F_j!}{N_{NS}^{\sum_n f_{n_j}} \left(F_j - \sum_n f_{n_j}\right)!} \right] \times \prod_{n=1}^{N} Z_{f_{n_1},\ldots,f_{n_m}}^{(n)}, \tag{A1}$$

where $Z_{g_1,\ldots,g_m}^{(n)}$ corresponds to states for promoter n occupied by g_1 number of TFs of type F_1, g_2 TFs of type F_2, etc. Analogously to Eq. (22) the single promoter partition functions

[2]Under review.

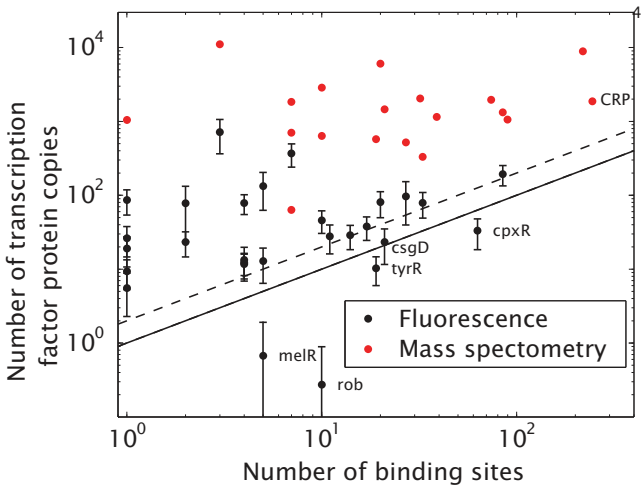


FIG. 17. (Color online) Transcription factor (TF) copy number vs number of binding sites, using two different protein censuses of *E. coli*. Protein copy numbers were determined using mass spectrometry [62] and fluorescence [7]. The number of binding sites was obtained from RegulonDB [24]. The solid line marks the boundary between depletable TFs (more binding sites than TF copies) and nondepletable (more TF copies than binding sites). For TFs forming dimers (e.g., CRP, Fis, GalR), this boundary is replaced by the dashed line. Due to incomplete knowledge about the *E. coli* regulatory system we expect the number of binding sites to be underestimated, and hence more TFs might belong to the depletable category than shown in the figure.

with multiple TF types are given by

$$Z^{(n)} = \sum_{g_1,\ldots,g_m} \left[\prod_{j=1}^{m} \frac{F_j!}{N_{NS}^{g_j}(F_j - g_j)!} \right] Z^{(n)}_{g_1,\ldots,g_m}. \quad \text{(A2)}$$

For the case when all promoter copies are identical we can also generalize the computationally more efficient Eq. (25) to multiple low-copy TF types $F_1, \ldots, F_m$

$$Z^{\text{tot}} = \sum_{\substack{k_{i_1,\ldots,i_m}, \forall\, i_1,\ldots,i_m \\ \sum_{i_1,\ldots,i_m} k_{i_1,\ldots,i_m} = N \\ \sum_{i_1,\ldots,i_m} i_j k_{i_1,\ldots,i_j,\ldots,i_m} \leq F_j, \forall\, j}} \binom{N}{\{k_{i_1,\ldots,i_m}\}} \times \left(\prod_{i_1,\ldots,i_m} Z_{i_1,\ldots,i_m}^{k_{i_1,\ldots,i_m}} \right) \times \prod_{j=1}^{m} \frac{F_j}{N_{NS}^{\sum_{i_1,\ldots,i_m} i_j k_{i_1,\ldots,i_m}} \left(F_j - \sum_{i_1,\ldots,i_m} i_j k_{i_1,\ldots,i_m}\right)!}. \quad \text{(A3)}$$

Here $k_{i_1,\ldots,i_m}$ is the number of promoters which have i_1 TF of type F_1 bound, i_2 TF of type F_2 bound, etc., $Z_{i_1,\ldots,i_m}$ corresponds to states with i_j TFs of type F_j bound, and the notation $\binom{N}{\{k_{i_1,\ldots,i_m}\}}$ refers to the multinomial coefficient $N! \prod_{i_1,\ldots,i_m} \frac{1}{k_{i_1,\ldots,i_m}!}$.

APPENDIX B: NUMBER OF BINDING SITES VS TF COPY NUMBER IN *E. coli*

For the specific case of *E. coli*, hundreds of TFs and their corresponding vast array of binding sites have been identified [24]. As a result, one can make an educated guess about regulatory architectures where the TF titration effect might play a role by looking for cases where the number of binding sites (N) approaches the number of TF molecules (F) per cell. An attempt to amass such data is shown in Fig. 17. The majority of genes belong to a regime where we do not expect strong titration of TFs, however, with a handful of exceptions, especially in the borderline regime $F \approx 2N$ where TFs binding as dimers could experience depletion. As new binding sites are discovered more TFs might fall into this category.

[1] N. E. Buchler, U. Gerland, and T. Hwa, Proc. Natl. Acad. Sci. USA **100**, 5136 (2003).
[2] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, Curr. Opin. Genet. Dev. **15**, 116 (2005).
[3] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman, and R. Phillips, Curr. Opin. Genet. Dev. **15**, 125 (2005).
[4] N. Buchler and M. Louis, J. Mol. Biol. **384**, 1106 (2008).
[5] A. Burger, A. M. Walczak, and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **107**, 4016 (2010).
[6] T. H. Lee and N. Maheshri, Mol. Syst. Biol. **8**, 576 (2012).
[7] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, Science **329**, 533 (2010).
[8] C. Zhong, D. Peng, W. Ye, L. Chai, J. Qi, Z. Yu, L. Ruan, and M. Sun, PLoS ONE **6**, e16025 (2011).
[9] S. E. Luria and R. Dulbecco, Genetics **34**, 93 (1949).
[10] K. Hanada, Y. Sawada, T. Kuromori, R. Klausnitzer, K. Saito, T. Toyoda, K. Shinozaki, W. H. Li, and M. Y. Hirai, Mol. Biol. Evol. **28**, 377 (2011).
[11] S. Wang, N. Liu, K. Peng, and Q. Zhang, Proc. Natl. Acad. Sci. USA **96**, 6824 (1999).
[12] A. Navarro-Quezada and D. J. Schoen, Proc. Natl. Acad. Sci. USA **99**, 268 (2002).
[13] E. K. Kentner, M. L. Arnold, and S. R. Wessler, Genetics **164**, 685 (2003).
[14] H. Bremer and P. P. Dennis, in *Escherichia coli and Salmonella Cellular and Molecular Biology*, edited by F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ASM Press, Washington, DC, 1996), pp. 1553–1569.
[15] T. J. Aitman, R. Dong, T. J. Vyse, P. J. Norsworthy, M. D. Johnson, J. Smith, J. Mangion, C. Roberton-Lowe, A. J. Marshall, E. Petretto *et al.*, Nature (London) **439**, 851 (2006).
[16] F. Cappuzzo, F. R. Hirsch, E. Rossi, S. Bartolini, G. L. Ceresoli, L. Bemis, J. Haney, S. Witta, K. Danenberg, I. Domenichini *et al.*, J. Natl. Cancer Inst. **97**, 643 (2005).
[17] E. H. Cook and S. W. Scherer, Nature (London) **455**, 919 (2008).

[18] S. Oehler, M. Amouyal, P. Kolkhof, B. von Wilcken-Bergmann, and B. Müller-Hill, EMBO J. **13**, 3348 (1994).
[19] J. Müller, S. Oehler, and B. Müller-Hill, J. Mol. Biol. **257**, 21 (1996).
[20] J. M. Vilar and S. Leibler, J. Mol. Biol. **331**, 981 (2003).
[21] H. G. Garcia and R. Phillips, Proc. Natl. Acad. Sci. USA **108**, 12173 (2011).
[22] H. G. Garcia, A. Sanchez, J. Q. Boedicker, M. Osborne, J. Gelles, J. Kondev, and R. Phillips, Cell Rep. **2**, 150 (2012).
[23] J. Q. Boedicker, H. G. Garcia, and R. Phillips, Phys. Rev. Lett. **110**, 018101 (2013).
[24] S. Gama-Castro, H. Salgado, M. Peralta-Gil, A. Santos-Zavaleta, L. Muniz-Rascado, H. Solano-Lira, V. Jimenez-Jacinto, V. Weiss, J. S. Garcia-Sotelo, A. Lopez-Fuentes *et al*., Nucleic Acids Res. **39**, 98 (2011).
[25] A. Cournac and J. Plumbridge, J. Bacteriol. **195**, 1109 (2013).
[26] M. J. Weickert and S. Adhya, Mol. Microbiol. **10**, 245 (1993).
[27] G. K. Ackers, A. D. Johnson, and M. A. Shea, Proc. Natl. Acad. Sci. USA **79**, 1129 (1982).
[28] J. Elf, G. W. Li, and X. S. Xie, Science **316**, 1191 (2007).
[29] R. B. Winter, O. G. Berg, and P. H. von Hippel, Biochemistry **20**, 6961 (1981).
[30] B. Müller-Hill, *The lac Operon: A Short History of a Genetic Paradigm* (Walter de Gruyter, Berlin, New York, 1996).
[31] E. Segal, T. Raveh-Sadka, M. Schroeder, U. Unnerstall, and U. Gaul, Nature (London) **451**, 535 (2008).
[32] T. Raveh-Sadka, M. Levo, and E. Segal, Genome Res. **19**, 1480 (2009).
[33] J. Gertz, E. D. Siggia, and B. A. Cohen, Nature (London) **457**, 215 (2009).
[34] H. H. He, C. A. Meyer, H. Shin, S. T. Bailey, G. Wei, Q. Wang, Y. Zhang, K. Xu, M. Ni, M. Lupien *et al*., Nat. Genet. **42**, 343 (2010).
[35] J. B. Kinney, A. Murugan, J. C. G. Callan, and E. C. Cox, Proc. Natl. Acad. Sci. USA **107**, 9158 (2010).
[36] W. D. Fakhouri, A. Ay, R. Sayal, J. Dresch, E. Dayringer, and D. N. Arnosti, Mol. Syst. Biol. **6**, 341 (2010).
[37] M. S. Sherman and B. A. Cohen, PLoS Comput. Biol. **8**, e1002407 (2012).
[38] T. Kuhlman, Z. Zhang, J. Saier, M. H., and T. Hwa, Proc. Natl. Acad. Sci. USA **104**, 6043 (2007).
[39] S. H. Meijsing, M. A. Pufall, A. Y. So, D. L. Bates, L. Chen, and K. R. Yamamoto, Science **324**, 407 (2009).
[40] M. E. Wall, D. A. Markowitz, J. L. Rosner, and R. G. Martin, PLoS Comput. Biol. **5**, e1000614 (2009).
[41] T. C. Voss, R. L. Schiltz, M. H. Sung, P. M. Yen, J. A. Stamatoyannopoulos, S. C. Biddie, T. A. Johnson, T. B. Miranda, S. John, and G. L. Hager, Cell **146**, 544 (2011).
[42] T. E. Kuhlman and E. C. Cox, Mol. Syst. Biol. **8**, 610 (2012).
[43] M. Santillán, Math. Modell. Nat. Phenom. **3**, 85 (2008).
[44] T. S. Gardner, C. R. Cantor, and J. J. Collins, Nature (London) **403**, 339 (2000).
[45] M. B. Elowitz and S. Leibler, Nature (London) **403**, 335 (2000).
[46] J. L. Cherry and F. R. Adler, J. Theor. Biol. **203**, 117 (2000).
[47] H. Bolouri and E. H. Davidson, Proc. Natl. Acad. Sci. USA **100**, 9371 (2003).
[48] G. M. Suel, J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz, Nature (London) **440**, 545 (2006).
[49] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman & Hall/CRC Mathematical and Computational Biology Series (Chapman & Hall/CRC, Boca Raton, FL, 2007).
[50] H. D. Kim and E. K. O'Shea, Nat. Struct. Mol. Biol. **15**, 1192 (2008).
[51] T. Y. Tsai, Y. S. Choi, W. Ma, J. R. Pomerening, C. Tang, and J. J. E. Ferrell, Science **321**, 126 (2008).
[52] T. Riley, E. Sontag, P. Chen, and A. Levine, Nat. Rev. Mol. Cell. Biol. **9**, 402 (2008).
[53] T. Cagatay, M. Turcotte, M. B. Elowitz, J. Garcia-Ojalvo, and G. M. Suel, Cell **139**, 512 (2009).
[54] I. S. Peter and E. H. Davidson, FEBS Lett. **583**, 3948 (2009).
[55] D. Sprinzak, A. Lakhanpal, L. Lebon, L. A. Santat, M. E. Fontes, G. A. Anderson, J. Garcia-Ojalvo, and M. B. Elowitz, Nature (London) **465**, 86 (2010).
[56] D. Sprinzak, A. Lakhanpal, L. Lebon, J. Garcia-Ojalvo, and M. B. Elowitz, PLoS Comput. Biol. **7**, e1002069 (2011).
[57] N. Balaskas, A. Ribeiro, J. Panovska, E. Dessaud, N. Sasai, K. M. Page, J. Briscoe, and V. Ribes, Cell **148**, 273 (2012).
[58] A. Warmflash, Q. Zhang, B. Sorre, A. Vonica, E. D. Siggia, and A. H. Brivanlou, Proc. Natl. Acad. Sci. USA **109**, E1947 (2012).
[59] M. Jishage and A. Ishihama, J. Bacteriol. **177**, 6832 (1995).
[60] I. L. Grigorova, N. J. Phleger, V. K. Mutalik, and C. A. Gross, Proc. Natl. Acad. Sci. USA **103**, 5332 (2006).
[61] S. Klumpp and T. Hwa, Proc. Natl. Acad. Sci. USA **105**, 18159 (2008).
[62] P. Lu, C. Vogel, R. Wang, X. Yao, and E. M. Marcotte, Nat. Biotechnol. **25**, 117 (2007).
[63] Y. Kao-Huang, A. Revzin, A. P. Butler, P. O'Conner, D. W. Noble, and P. H. von Hippel, Proc. Natl. Acad. Sci. USA **74**, 4228 (1977).
[64] W. Runzi and H. Matzura, J. Bacteriol. **125**, 1237 (1976).
[65] T. P. Malan, A. Kolb, H. Buc, and W. R. McClure, J. Mol. Biol. **180**, 881 (1984).
[66] W. R. McClure, Annu. Rev. Biochem. **54**, 171 (1985).
[67] N. B. Reppas, J. T. Wade, G. M. Church, and K. Struhl, Mol. Cell **24**, 747 (2006).
[68] L. M. Hsu, Biochim. Biophys. Acta **1577**, 191 (2002).
[69] H. Salgado, M. Peralta-Gil, S. Gama-Castro, A. Santos-Zavaleta, L. Muniz-Rascado, J. S. Garcia-Sotelo, V. Weiss, H. Solano-Lira, I. Martinez-Flores, A. Medina-Rivera *et al*., Nucleic Acids Res. **41**, D203 (2013).
[70] H. G. Garcia, P. Grayson, L. Han, M. Inamdar, J. Kondev, P. C. Nelson, R. Phillips, J. Widom, and P. A. Wiggins, Biopolymers **85**, 115 (2007).
[71] R. Schleif, Bioessays **25**, 274 (2003).
[72] F. Tricomi, Ann. Mat. Pura Appl. **26**, 141 (1947).
[73] http://functions.wolfram.com/07.33.03.0040.01
[74] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.89.012702 for information on how we perform these computations using MATHEMATICA.
[75] http://functions.wolfram.com/07.33.20.0005.01
[76] R. C. Brewster, D. L. Jones, and R. Phillips, PLoS Comput. Biol. **8**, e1002811 (2012).
[77] P. S. Swain, M. B. Elowitz, and E. D. Siggia, Proc. Natl. Acad. Sci. USA **99**, 12795 (2002).
[78] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).

[79] M. J. Dunlop, R. S. Cox, J. H. Levine, R. M. Murray, and M. B. Elowitz, Nat. Genet. **40**, 1493 (2008).
[80] P. Hammar, Ph.D. thesis, Uppsala University, Computational and Systems Biology, 2013.
[81] S. T. Liang, P. P. Dennis, and H. Bremer, J. Bacteriol. **180**, 6090 (1998).
[82] K. L. Jones and J. D. Keasling, Biotechnol. Bioeng. **59**, 659 (1998).
[83] M. Delbruck, J. Bacteriol. **50**, 131 (1945).
[84] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[85] R. G. Martin, E. S. Bartlett, J. L. Rosner, and M. E. Wall, J. Mol. Biol. **380**, 278 (2008).