

# Noise and Information Transmission in Promoters with Multiple Internal States

Georg Rieckh\* and Gašper Tkačik

Institute of Science and Technology Austria, Am Campus 1, Klosterneuburg, Austria

**ABSTRACT** Based on the measurements of noise in gene expression performed during the past decade, it has become customary to think of gene regulation in terms of a two-state model, where the promoter of a gene can stochastically switch between an ON and an OFF state. As experiments are becoming increasingly precise and the deviations from the two-state model start to be observable, we ask about the experimental signatures of complex multistate promoters, as well as the functional consequences of this additional complexity. In detail, we i), extend the calculations for noise in gene expression to promoters described by state transition diagrams with multiple states, ii), systematically compute the experimentally accessible noise characteristics for these complex promoters, and iii), use information theory to evaluate the channel capacities of complex promoter architectures and compare them with the baseline provided by the two-state model. We find that adding internal states to the promoter generically decreases channel capacity, except in certain cases, three of which (cooperativity, dual-role regulation, promoter cycling) we analyze in detail.

## INTRODUCTION

Gene regulation—the ability of cells to modulate the expression level of genes to match their current needs—is crucial for survival. One important determinant of this process is the wiring diagram of the regulatory network, specifying how environmental or internal signals are detected, propagated, and combined to orchestrate protein level changes (1). Beyond the wiring diagram, the capacity of the network to reliably transmit information about signal variations is determined also by the strength of the network interactions (the “numbers on the arrows” (2)), the dynamics of the response, and the noise inherent to chemical processes happening at low copy numbers (3–6).

How do these factors combine to set the regulatory power of the cell? Information theory can provide a general measure of the limits to which a cell can reliably control its gene expression levels. Especially in the context of developmental processes, where the precise establishment and readout of positional information has long been appreciated as crucial (7), information theory can provide a quantitative proxy for the biological function of gene regulation (8). This has led to theoretical predictions of optimal networks that maximize transmitted information given biophysical constraints (8–13), and hypotheses that certain biological networks might have evolved to maximize transmitted information (14). Some evidence for these ideas has been provided by recent high-precision measurements in the gap gene network of the fruit fly (15). In parallel to this line of research, information theory has been used as a general and quantitative way to compare signal processing motifs (16–29). Further theoretical work has demonstrated a rela-

tionship between the information capacity of an organism’s regulatory circuits and its evolutionary fitness (30–32).

Previously, information theoretic investigations primarily examined the role of the regulatory network. In this study, we focus on the molecular level, i.e., on the events taking place at the regulatory regions of the DNA. Little is known about how the architecture of such microscopic events shapes information transfer in gene regulation. Yet it is precisely at these regulatory regions that the mapping from the “inputs” in the network wiring diagram into the corresponding “output” expression level is implemented by individual molecular interactions. In this bottleneck various physical sources of stochasticity—such as the binding and diffusion of molecules (33–35), and the discrete nature of chemical reactions (36)—must play an important role. In the simplest picture, gene expression is modulated through transcriptional regulation. This involves molecular events such as the binding of transcription factors (TFs) to specific sites on the DNA, chemical events that facilitate or block TF binding (e.g., through chromatin modification), or events that are subsequently required to initiate transcription (e.g., the assembly and activation of the transcription machinery).

Although the exact sequence of molecular events at the regulatory regions often remains elusive (especially in eukaryotes), quantitative measurements have highlighted factors that contribute to the fidelity by which TFs can affect the expression of their target genes. These findings have been succinctly summarized by the so-called “telegraph model” of transcriptional regulation (37): the two-state promoter switches stochastically between the states “ON” and “OFF,” with switching rates dependent on the concentration(s) of the regulatory factor(s). This dependence can either be biophysically motivated (e.g., by a thermodynamic

Submitted August 6, 2013, and accepted for publication January 7, 2014.

\*Correspondence: [georg.rieckh@ist.ac.at](mailto:georg.rieckh@ist.ac.at)

Editor: Dennis Bray

© 2014 by the Biophysical Society  
0006-3495/14/03/1194/11 \$2.00

<http://dx.doi.org/10.1016/j.bpj.2014.01.014>



model of TF binding to DNA), or it can be considered as purely phenomenological. The switching itself is independent of mRNA production but determines the overall production rate. The production of mRNA molecules from one state is usually modeled as a Poisson process, with a first-order decay of messages; this is usually followed by a birth-death process in which proteins are translated from the messages. This two-state model is well-studied theoretically (37–43) and has been used extensively to account for measurements of noise in gene expression (44–47). An increasing amount of information about molecular details has motivated extensions to this model by introducing more than two states in specific systems (48–55), and recent measurements of noise in gene expression provided some support for such complex regulatory schemes (56–59).

In this article we address the general question of the functional effect of complex promoters with multiple internal states. How does the presence of multiple states affect information transmission? Under what conditions, if any, can multistate promoters perform better than the two-state model? We consider a wide spectrum of generic promoter models, such that many molecular “implementations” can share the same mathematical model. With this framework in hand, we derive the total noise in mRNA expression and discuss how measurements of this function can be diagnostic of the underlying mechanism of regulation. To answer the main question of this paper—namely if additional complexity at the promoter can lead to an improvement in controlling the output level of a gene—we compute the information transmission from transcription factor concentrations to regulated protein expression levels through two- and three-state promoters. Finally, we analyze in detail three complex promoter architectures that outperform the two-state regulation.

## METHODS

### Channel capacity as a measure of regulatory power

We start by considering a genetic regulatory element—e.g., a promoter or an enhancer—as a communication channel, shown in Fig. 1 A. As the concentrations of the relevant inputs (for example, transcription factors) change, the regulatory element responds by varying the rate of target gene expression. In steady state, the relationship between input  $k$  and expression level of the regulated protein  $g$  is often thought of as a “regulatory function” (60). Although attractive, the notion of a regulatory function in a mathematical sense is perhaps misleading: gene regulation is a noisy process, and so for a fixed value of the input we have not one, but a distribution of different possible output expression levels,  $P(g|k)$  (see Fig. 1 B). When the noise is small, it is useful to think of a regulatory function as describing the average expression level,  $\bar{g}(k) = \int dg gP(g|k)$ , and of the noise as inducing some random fluctuation around that average. The variance of these fluctuations,  $\sigma_g^2(k) = \int dg (g - \bar{g}(k))^2 P(g|k)$ , is thus a measure of noise in the regulatory element; note that its magnitude depends on the input,  $k$ .

The presence of noise puts a bound on how precisely changes in the input can be mapped into resulting expression levels on the output side—or

inversely, how much the cell can know about the input by observing the (noisy) outputs alone. In his seminal work on information theory (61), Shannon introduced a way to quantify this intuition by means of *mutual information*, which is an assumption-free, positive scalar measure in bits, defined as follows:

$$I(k; g) = \iint dk dg P(k)P(g|k) \log_2 \left[ \frac{P(g|k)}{P(g)} \right]. \quad (1)$$

In Eq. 1,  $P(g|k)$  is a property of the regulatory element, which we compute below, whereas  $P(k)$  is the distribution of inputs (e.g., TF concentrations) that regulate the expression; finally,  $P(g) = \int dk P(g|k)P(k)$  is the resulting distribution of gene expression levels. With  $P(g|k)$  set by the properties of the regulatory element and the biophysics of the gene expression machinery, there exists an optimal choice for the distribution of inputs,  $P^*(k)$ , that maximizes the transmitted information. This maximal value,  $I^*(k; g) = \max_{P(k)} I(k; g)$ , also known as the *channel capacity* (62), summarizes in a single number the “regulatory power” intrinsic to the regulatory element (10–14).

Our goal is to compute the channel capacity between the (single) regulatory input and the target gene expression level for information flowing through various complex promoters. Under the assumption that noise is small and approximately Gaussian for all levels of input, the complicated expression for the information transmission in Eq. 1 simplifies, and the channel capacity  $I^*(k; g)$  can be computed analytically from the regulatory function,  $\bar{g}(k)$ , and the noise,  $\sigma_g^2(k)$ . The result (10–14) is as follows:

$$I^*(k; g) = \log_2 \frac{Z}{\sqrt{2\pi e}}, \quad \text{with} \quad (2)$$

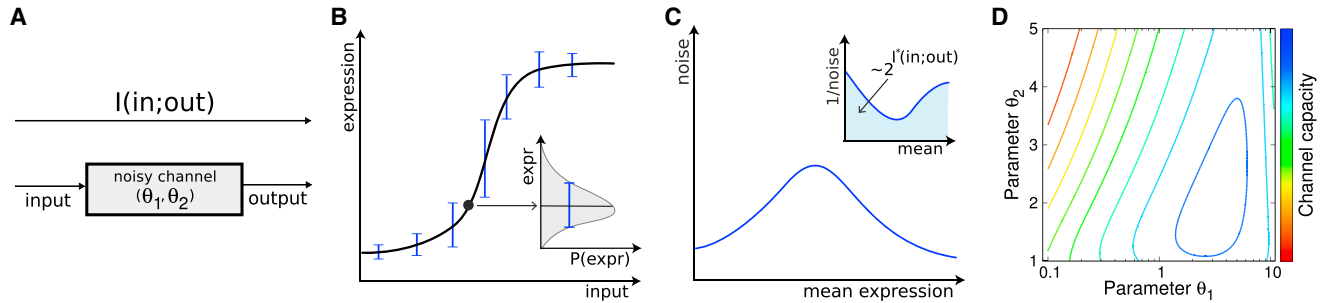
$$Z = \int_0^{k_{\max}} dk \frac{|d\bar{g}/dk|}{\sigma_g(k)} = \int_{\bar{g}_{\min}}^{\bar{g}_{\max}} \frac{d\bar{g}}{\sigma_g(\bar{g})}, \quad (3)$$

where in the last equality we changed the integration variables to express the result in terms of the average induction level,  $\bar{g}$ , using the regulatory function  $\bar{g} = \bar{g}(k)$ . This integral is graphically depicted in Fig. 1 C (inset). Finally, we will use this to explore the dependence of  $I^*(k; g)$  on parameters that define the promoter architecture (see Fig. 1 D), looking for those arrangements that lead to large channel capacities and thus high regulatory power.

Information as a measure of regulatory power has a number of attractive mathematical properties (for review, see (8)); interpretation-wise, the crucial property is that it roughly counts (the logarithm of) the number of distinguishable levels of expression that are accessible by varying the input—also taking into account the level of noise in the system. A capacity of 1 bit, therefore, suggests that the gene regulatory element could act as a binary switch with two distinguishable expression levels (note that this is different from saying that the promoter has two (or any other number) of possible internal states); capacities smaller than 1 bit correspond to (biased) stochastic switching, whereas capacities higher than 1 bit support graded regulation. An increase of information by 1 bit means that the number of tunable and distinguishable levels of gene expression has roughly doubled, implying that changes of less than a bit are meaningful. Careful analysis of gene expression data for single-input single-output transcriptional regulation suggest that real capacities can exceed 1 bit (14). Increasing this number substantially beyond a few bits, however, necessitates very low levels of noise in gene regulation, requiring prohibitive numbers of signaling molecules (10).

### Multistate promoters as state transition graphs

To study information transmission, we must first introduce the noise model in gene regulation, which consists of two components: i) the generalization



**FIGURE 1** A genetic regulatory element as an information channel. (A) Mutual information  $I$  is a quantitative measure of the signaling fidelity with which a genetic regulatory element maps inputs (e.g., TF concentrations) into the regulated expression levels. In this schematic example, the properties of the element are fully specified by two parameters  $(\theta_1, \theta_2)$ ; e.g., the switching rates between promoter states). (B) In steady state, the input/output mapping can be summarized by the regulatory function  $\bar{g}(k)$  (solid black line) for target protein expression (equivalently,  $\bar{m}(k)$  for target mRNA expression, not shown); noise,  $\sigma_g(k)$  (respectively  $\sigma_m(k)$  for mRNA), induces fluctuations around this curve (inset and error bars on the regulatory function). (C) The “noise characteristic” (noise vs. mean expression) is usually experimentally accessible for mRNA using in situ hybridization methods and can reveal details about the promoter architecture. The maximal transmitted information (channel capacity  $I^*$ , see Eqs. 2 and 3) is calculated from the area under the inverse noise curve for the target protein,  $\sigma_g^{-1}(\bar{g})$  (inset). (D) Channel capacity  $I^*$  is, in this example case, maximized for a specific choice of parameters  $\theta_1, \theta_2$  (blue peak). To see this figure in color, go online.

of the random telegraph model to multiple states, and ii), the model for input noise that captures fluctuations in the number of regulatory molecules. Starting with the first component, we compute the mean and variance for regulated mRNA levels, since these quantities are experimentally accessible when probing noise in gene expression. We assume that the system has reached steady state and that gene product degradation is the slowest timescale in the problem, i.e., that target mRNA or protein levels average over multiple state transitions of the promoter and that the resulting distributions of mRNA and protein are thus unimodal. Whereas for protein levels these assumptions hold over a wide range of parameters and include many biologically relevant cases, there exist examples where promoter switching is very slow and the system would need to be treated with greater care (e.g., (38,50,63)).

Let us represent the possible states of the promoter (and the transitions between them) by a state transition graph as in Fig. 2 A. Gene regulation occurs when an input signal modifies one (or more) of the rates at which the promoter switches between its states. To systematically analyze many promoter architectures, we choose not to endow from the start each graph with a mechanistic interpretation, which would map the abstract promoter states to various configurations of certain molecules on the regulatory regions of the DNA (as in Fig. 2 B). This is because there might be numerous molecular realizations of the same abstract scheme, which will yield identical noise characteristics and identical information transmission. In Fig. 3 and Fig. S1 in the Supporting Material, we discuss known examples related to different promoter architectures.

Given a specific promoter architecture, we would like to compute the first two moments of the mRNA distribution under the above assumptions. Here, we only sketch the method for the promoter in Fig. 2 A; for a general description and details see the Supporting Material. We will denote the rate of mRNA production from the active state(s) by  $r$  and its degradation rate by  $d$ . Let further  $p_i$  be the fractional occupancy of state  $i \in \{a, 1, 2\}$  and  $k_{ij}$  the rate of transitioning from state  $i$  to  $j$ ,  $i \neq j$ . Here,  $a$  is the active state, and 1, 2 are the nonexpressing states. Equations 4 and 5 then describe the behavior of the state occupancy and mRNA level  $m$ :

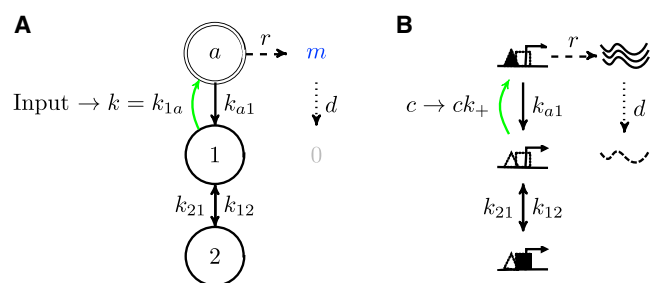
$$\partial_t \mathbf{p} = \mathbf{K} \mathbf{p} + \boldsymbol{\xi}, \quad \partial_t m = r p_a - d m + \xi_m, \quad \text{with} \quad (4)$$

$$\mathbf{K} = \begin{bmatrix} -(k_{a1} + k_{a2}) & k_{1a} & k_{2a} \\ k_{a1} & -(k_{1a} + k_{12}) & k_{21} \\ k_{a2} & k_{12} & -(k_{2a} + k_{21}) \end{bmatrix}, \quad (5)$$

and  $\mathbf{p} = (p_a, p_1, p_2)^T$ ;  $\boldsymbol{\xi} = (\xi_a, \xi_1, \xi_2)^T$  and  $\xi_m$  are Langevin white-noise random forces (36,64) (see the Supporting Material). In this setup it is easy to compute the mean and the variance in expression levels given a set of chosen rate constants. Using the assumption of slow gene product degradation,  $d \ll k_{ij}$ , we can write the noise in a generic way:

$$\sigma_m^2 = \bar{m} \left[ 1 + \frac{r}{d} p_{\text{act}} \Delta \right], \quad (6)$$

where  $p_{\text{act}}$  is the occupancy of the active states ( $p_a$  or  $p_a + p_b$ ), and the dimensionless expression for  $\Delta$  depends on the promoter architecture and can be read out from Fig. 3 A for different promoter models. The expression for noise in Eq. 6 has two contributions. The first, where the variance is equal to the mean ( $\sigma_m^2 = \bar{m} + \dots$ ) is the “output noise” attributable to the birth-death production of single mRNA molecules (also called “shot noise” or “Poisson noise”). The second contribution to the variance in Eq. 6 is attributable to stochastic switching of the promoter between internal states,



**FIGURE 2** Promoters as state transition graphs. (A) A state transition graph for an example three-state promoter. Active state  $a$  (double circle) expresses mRNA  $m$  at rate  $r$ , which are then degraded with rate  $d$ . Transition into  $a$  (green arrow) is affected by the input that modulates rate  $k = k_{1a}$ . Stochastic transitions between promoter states  $\{a, 1, 2\}$  are an important contribution to the noise,  $\sigma_m(k)$ . (B) A possible mechanistic interpretation of the diagram in (A): state 1 is an unoccupied promoter, state 2 is an inaccessible promoter (occupied by a nucleosome or repressor, solid square). Transition to the active state (green arrow) is modulated by changing the concentration  $c$  of activators (solid triangles) that bind their cognate site (open triangles) at the promoter with the rate  $ck_+$ . To see this figure in color, go online.

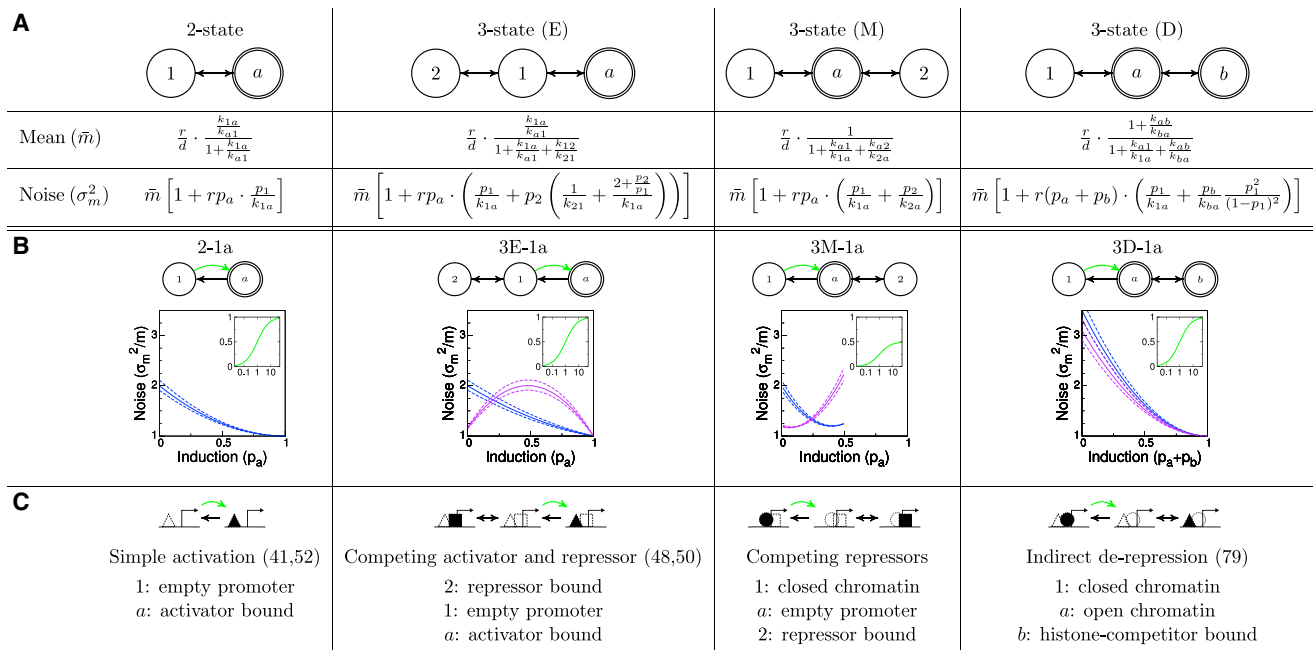


FIGURE 3 Mean expression level and noise for different promoter models. (A) Expressions for the mean (first row,  $\bar{m}$ ) and the variance (second row,  $\sigma_m^2$ ) of the mRNA distribution in steady state for different promoter architectures. In the limit of  $k_{12} \rightarrow 0$  (respectively  $k_{a2}, k_{ab} \rightarrow 0$ ), the expression for the two-state model is obtained from the models with three states. The names of the topologies indicate the position of the expressing state: E(nd), M(iddle), D(ouble). (B) Noise characteristics ( $p_a$  on the x axis vs. the Fano factor,  $\sigma_m^2/\bar{m}$ , on y axis) for different promoter models. Here, in all models  $k = k_{1a}$  is modulated (green arrow) to achieve different mean expression levels. For all rates (except  $k$ ) equal to 1 (blue lines), the functional form of the noise characteristics is very similar. This remains true for a variation of  $k_{a1}$  of  $\pm 10\%$  (blue dashed lines). Making the rates in/out of the third state (state 2 or b) slower by a factor of 5 (purple lines, dashed  $\pm 10\%$ ) yields qualitatively different results. Insets show the induction curve,  $p_{act}(k)$ , where  $k$  is the modulated rate. A full table and possible molecular interpretations of different promoter schemes are given in Fig. S1 (74–80). (C) Possible interpretations of the modulation schemes. Triangles represent activators, squares are repressors, and circles are histones. The dotted shapes denote (empty) binding sites. Cited references use similar models. To see this figure in color, go online.

referred to as the “switching noise.” This term does depend on the promoter architecture and has a more complicated functional form than being simply proportional to the mean. A first glance at the expressions for noise seems to imply that going from two to three promoter states can only increase the noise (and by Eq. 3 decrease information), since new, positive contributions appear in the expressions for  $\sigma_m^2$ ; we will see that, nevertheless, transmitted information can increase for certain architectures.

### Input noise

In addition to the noise sources internal to the regulatory mechanism, we also consider the propagation of fluctuations in the input, which will contribute to the observed variance in the gene expression level. Can we say anything general about the transmission of input fluctuations through the genetic regulatory element? Consider, for instance, the modulated rate  $k$  that depends on the concentration  $c$  of some transcription factor, as in  $k = k_+c$ , where  $k_+$  is the association rate to the TF’s binding site. Since the TF itself is expressed in a stochastic process, we could expect that there will be (at least) Poisson-like fluctuations in  $c$  itself, such that  $\sigma_c^2 \propto c$ ; this will lead to an effective variance in  $k$  that will be propagated to the output variance in proportion to the “susceptibility” of the regulatory element,  $(\partial \bar{g} / \partial k)^2$ . Extrinsic noise would affect the regulatory element in an analogous way, as suggested in a previous study (65). Independently of the noise origin, we can write

$$\sigma_g^2 = \dots + v \left( \frac{\partial \bar{g}}{\partial k} \right)^2 k, \quad (7)$$

where (...) indicate output and switching terms from Eq. 6 and  $v$  is the proportionality constant ( $\sigma_k^2 = vk$ ) that is related to the magnitude of the input fluctuations and, possibly, their subsequent time averaging (5).

Even if there were absolutely no fluctuations in the total concentration  $c$  of TF molecules in the cell (or the nucleus), the sole fact that they need to find their target by diffusion puts a lower bound on the variance of the local concentration at the regulatory site (33,34,66). This diffusion noise represents a substantial contribution to the total, as was shown by the analysis of high-precision measurements in gene expression noise (35,45). For this biophysical limit set by diffusion, we find yet again that the variance in the input is proportional to the input itself, and that the values of  $v$  are of order one when expressed in units of the relevant averaging time (see the Supporting Material and (66–68)). This, in sum, demonstrates that Eq. 7 can be used as a generic model for diverse kinds of input noise.

## RESULTS

### Experimentally accessible noise characteristics

Could complex promoter architectures be distinguished by their noise signatures, even in the easiest case where the input noise can be neglected (as is often assumed (45))? The expressions for the noise presented in Fig. 3 A hold independently of which transition rate the input is modulating. We can specialize these results by choosing the



*modulation scheme*, that is, making one (or more) of the transition rates the regulated one. This allows us to construct the regulatory function (insets in Fig. 3 B). Additionally, we can also plot the noise (shown in Fig. 3 B as the Fano factor,  $\sigma_m^2/\bar{m}$ ) as a function of the mean expression,  $\bar{m}$ , thus getting the *noise characteristic* of every modulation scheme. These curves, shown in Fig. 3 B, are often accessible from experiments (44,69), even when the identity of the expressing state or the mechanism of modulation is unknown. We systematically organize our results in Fig. 3 B (for the case when  $k = k_{1a}$  is modulated) and provide a full version in Fig. S1; we also list four molecular schemes implementing these architectures in Fig. 3 C, while providing additional molecular implementations in Fig. S1. We emphasize that very different molecular mechanisms of regulation can be represented by the same architecture, resulting in the same mathematical analysis and information capacity.

Measured noise-vs.-mean curves have been used to distinguish between various regulation models (44,69,70). For this, two conditions have to be met (45,71). First, it must be possible to access the full dynamic range of the gene expression in an experiment, and this sometimes seems hard to ensure. The second condition is that the input noise is not the dominant source of noise: input noise can mimic promoter switching noise and can, e.g., provide alternative explanations for noise measurements in (44) that quantitatively fit the data (not shown).

Even if these conditions are met, it would be impossible to distinguish between certain promoter architectures (e.g., 2-a1 vs. 3E-a1) with this method, whereas some would require data of a very high quality to distinguish (e.g., activating 3E-1a vs. repressing 3E-12, see Fig. S1), at least in certain parameter regimes. On the other hand, there exist noise characteristics that can only be obtained with multiple states (e.g., 3M-1a).

One feature that can easily be extracted from the measured noise characteristics is the asymptotic induction: it can be equal to 1 (e.g., in 2-1a), or bounded away from 1 (e.g., in 3M-1a). Although this distinction between architectures cannot be inferred from the shapes of the regulatory functions, the effect on the noise characteristics is unambiguous: in the case where the expressing state is never saturated, the Fano factor does not drop to the Poisson limit of 1 even at the highest-expression levels (which seems to have been the case in a previous study (44)).

Taken together, when the range of promoter architectures is extended beyond the two-state model, distinguishing between these architectures based on the noise characteristics seems possible only under restricted conditions, emphasizing the need for dynamical measurements that directly probe transition rates (e.g., (59,72)), or for the measurements of the full mRNA distribution (rather than only its second moment). We note that dynamic rates are often re-

ported *assuming* the two-state model, as they are inferred from the steady-state noise measurements (e.g., (44,47)), and only a few experiments probe the rates directly (e.g., (73)); for a brief review of the rates and their typical magnitudes, see the Supporting Material.

## Information transmission in simple gene regulatory elements

### Protein noise

In most cases the functional output of a genetic regulatory element is not the mRNA, but the translated protein. Incorporating stochastic protein production into the noise model does not affect the functional form of the noise, but only rescales the magnitude of the noise terms. To see this, we let proteins be produced from mRNA at a rate  $r'$  and degraded or diluted at  $d'$ , such that  $d' \ll d$  is the slowest timescale in the problem. Then the mean protein expression level is  $\bar{g}(k) = (r'/d')\bar{m}(k)$ . The output and promoter switching noise contributions are affected differently, so that the protein level noise can be written as (45):

$$\sigma_g^2 = \bar{g} \left[ (1 + w) + \frac{r'}{d'} p_{\text{act}} \Delta' \right] + (\text{input noise}), \quad (8)$$

where  $w = r'/d$  is the *burst size* (the average number of proteins translated from one mRNA molecule),  $\Delta' = (r'/d')(d'/d)\Delta$  and the other quantities are as defined in Eq. 6.

### Information transmission in the two-state model

To establish the baseline against which to compare complex promoters, we look first at the two-state promoter (2-1a). Here the transition into the active state is modulated by TF concentration  $c$  via  $k = k_{1a} = k_{+c}$ , as it would be in the simple case of a single TF molecule binding to an activator site to turn on transcription. Adding together the noise contributions of Eqs. 8 and 7, we obtain the following model for the total noise:

$$\sigma_g^2 = \bar{g} \left[ (1 + w) + \frac{rr'}{dk_{a1}}(1 - p_a)^2 + \frac{v}{k_{a1}} \frac{rr'}{dd'}(1 - p_a)^3 \right]. \quad (9)$$

To compute the corresponding channel capacity, we use Eq. 3 with the noise given by Eq. 9 as follows:

$$Z = \int_{\bar{g}_{\min}}^{\bar{g}_{\max}} \frac{d\bar{g}}{\sigma_g(\bar{g})} = \sqrt{N_{\max}} \int_0^{p_a^{\max}} dp_a \times p_a^{-1/2} \left[ 1 + \frac{1}{k_-}(1 - p_a)^2 + \frac{\tilde{v}}{k_-}(1 - p_a)^3 \right]^{-1/2} \quad (10)$$

$$= \sqrt{N_{\max}} Z_0. \quad (11)$$

Here,  $N_{\max} = (rr')/[(dd')(1+w)]$  and  $\tilde{k}_- = k_{a1}/(d'N_{\max})$  is the dimensionless combination of parameters related to the off-rate for the TF dissociation from the binding site.  $N_{\max}$  can be interpreted as the number of *independently* produced output molecules when the promoter is fully induced (11–13). In the case where mRNA transcription is the limiting step for protein synthesis,  $N_{\max}$  corresponds to the maximal average number of mRNA synthesized during a protein lifetime:  $N_{\max} = r/d' w/(1+w) \approx r/d'$ . With this choice of parameters,  $N_{\max}$  affects  $Z$  multiplicatively and thus simply adds a constant offset to the channel capacity  $I^*$  (see Eq. 2) without affecting the parameter values that maximize capacity. In what follows we therefore disregard this additive offset, and examine in detail only  $I^* = \log_2 Z_0$ . We also only use dimensionless quantities (as above, e.g., the rates are expressed in units of  $d'$ ), but leave out the tilde symbols for clarity.

### Optimizing information transmission

What parameters maximize the capacity of the two-state promoter 2-1a given by Eq. 10? Given that the dynamic range of input (e.g., TF concentration) is limited (10–13),  $k \in [0, k_{\max}]$ , and given a choice of  $v$  that determines the type and magnitude of input noise, the channel capacity  $I^*$  for the two-state promoter only depends nontrivially on the choice of a single parameter,  $k_-$ . Fig. 4 shows the tradeoff that leads to the emergence of a well-defined optimal value for  $k_-^*$ : at a fixed dynamic range for the input,  $k \in [0, k_{\max}]$ , the information-maximizing solution chooses  $k_-^*$  that balances the strength of binding (such that the dynamic range of expression is large), while simultaneously

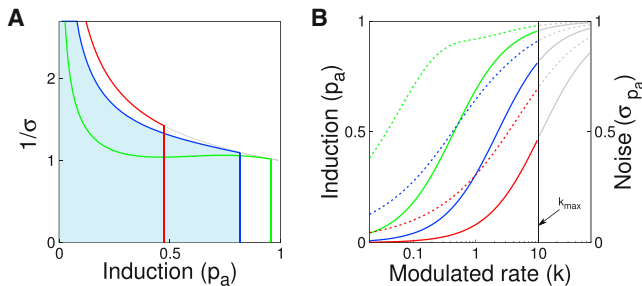


FIGURE 4 Finding  $k_-$  that optimizes information transmission in a two-state promoter. The strength of the input noise is fixed at  $v = 2$  and the input dynamic range for  $k$  is from 0 to  $k_{\max} = 10$ . (A) The integrand of Eq. 10 is shown for the optimal choice of  $k_-$  (blue), and for two alternative  $k_-$  values: a factor of 5 larger (red) or smaller (green) than the optimum. Although increasing  $k_-$  lowers the noise, it also decreases the integration limit, and vice versa for decreasing  $k_-$ . (B) The effect on the regulatory function (solid, left axis) and the noise (dashed, right axis), of choosing different  $k_-$  values. Optimal  $k_-$  (blue curves) from (A) leads to a balance between the dynamic range in the mean response (the maximal achievable induction), and the magnitude of the noise. Higher  $k_-$  (red curves), in contrast, lead to smaller noise, but fail to make use of the full dynamic range of the response. The gray part of the regulation curves cannot be accessed, since the input only ranges over  $k \in [0, k_{\max}]$ . To see this figure in color, go online.

keeping the noise as low as possible. If this abstract promoter model were interpreted in mechanistic terms where a TF binds to activate gene expression, then choosing the optimal  $k_-$  would amount to choosing the optimal value for the dissociation constant of our TF; importantly, the existence of such a nontrivial optimum indicates that, at least in an information-theoretic sense, the best binding is not the tightest one (10–13,27,81,82). This tradeoff between noise and dynamic range of outputs (also called “plasticity”) has also been noticed in other contexts (83,84).

### Improving information transmission with multistate promoters

We would like to know if complex promoter architectures can outperform the two-state model in terms of channel capacity. To this end, we have examined the full range of three-state promoters, summarized in Fig. S1, and found that generally—as long as only one transition is modulated and only one state is active—extra promoter states lead to a decrease in the channel capacity relative to two-state regulation. However, by relaxing these assumptions, architectures that outperform two-state promoters can be found.

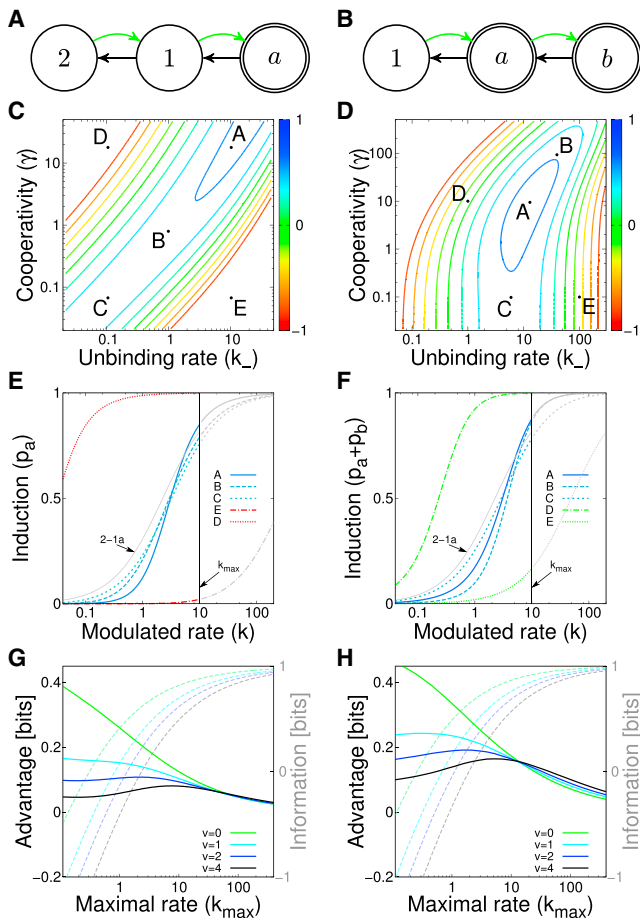
#### Cooperative regulation

The first such pair of architectures is illustrated in Fig. 5 A and B: three-state promoters with one (or two) expressing states, where *two* transitions into the expressing states are simultaneously modulated by the input. A possible molecular interpretation of these promoter state diagrams is an AND-architecture cooperative binding for the model with one expressing state, and an OR-architecture cooperative activation for the model with two expressing states. In case of an AND-architecture, a TF molecule hops onto the empty promoter (state 2) with rate  $2k$  (since there are two empty binding sites), whereas a second molecule can hop on with rate  $\rho k$  (called “recruitment” if  $\rho > 1$ ), bringing the promoter into the active state. The first of two bound TF molecules falls off with rate  $2\gamma^{-1}k_-$  (called “cooperativity” if  $\gamma > 1$ ), bringing the promoter back to state 1, and ultimately, the last TF molecule can fall off with rate  $k_-$ . The dynamics are now described (see Eq. 5) by the matrix

$$\mathbf{K} = \begin{bmatrix} -(2\gamma^{-1}k_-) & \rho k & 0 \\ 2\gamma^{-1}k_- & -(k_- + \rho k) & 2k \\ 0 & k_- & -(2k) \end{bmatrix}, \quad (12)$$

and  $\mathbf{p} = (p_a, p_1, p_2)^T$ , respectively  $\mathbf{p} = (p_b, p_a, p_1)^T$ . To compute the noise, we can use the solutions for the generic three-state model 3E from Fig. 3 A by making the following substitutions:  $k_{a1} = 2\gamma^{-1}k_-$ ,  $k_{1a} = \rho k$ ,  $k_{12} = k_-$ ,  $k_{21} = 2k$ .

To simplify our exploration of the parameter space, we choose  $\rho = 1$  (i.e., no recruitment), but keep  $k_-$  (unbinding rate) and  $\gamma$  (cooperativity) as free parameters; the modulated rate  $k$  is proportional to the concentration of TF



**FIGURE 5** Improving information transmission via cooperativity. (A,B) The state transition diagram for the AND-architecture three-state promoter (at left, one active state) and OR-architecture (at right, two active states). (C,D) The information planes showing the channel capacity (color code) for various combinations of  $k_-$  and  $\gamma$  at a fixed maximum allowed  $k_{\max} = 10$ , for the AND- and OR-architectures and  $\nu = 2$ . (E,F) The regulatory functions of various models selected from parameters denoted by dots in the information planes in (B,C). Solutions that maximize the information denoted with a solid blue line. Colors indicate the channel capacities. The gray part of the regulation curves cannot be accessed, since the input only ranges over  $k \in [0, k_{\max}]$ . For comparison, we also plot the regulation curve of the 2-1a scheme (with its optimal  $k^*$ ). (G,H) Channel capacity  $I_3^*$  (dashed lines, axes at right) and the advantage of the multistate scheme over the best two-state promoter,  $I_3^* - I_2^*$  (solid lines, axes at left), of the AND and OR models, as a function of the maximal input range,  $k_{\max}$ , and the strength of the input noise ( $\nu$ , color). Since there is no globally optimal choice for  $\gamma$  for the AND-architecture, we fix  $\gamma = 10$  in (G), and optimize only over  $k_-$  values. To see this figure in color, go online.

molecules and is allowed the range  $k \in [0, k_{\max}]$ . For every choice of  $(k_-, \gamma)$ , we computed the regulatory function and the noise, and used these to compute the capacity,  $I^*(k; g)$ , using Eqs. 2 and 3. This information is shown in Fig. 5 C and D for the AND- and OR-architecture, respectively.

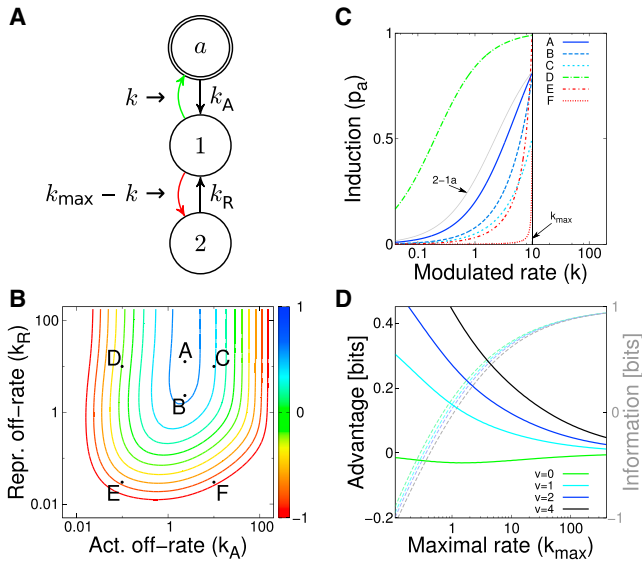
In the case of an AND-architecture, where both molecules of the TF have to bind for the promoter to express, there is a ridge of optimal solutions: as we move along the ridge in the

direction of increasing information, cooperativity is increased and thus the doubly occupied state is stabilized, while the unbinding rate increases as well. This means that the occupancy of state 1 becomes negligible, and the regulation function becomes ever steeper, as is clear from Fig. 5 E, while maintaining the same effective dissociation constant (the input  $k = k_{1/2}$  at which the promoter is half induced, i.e.,  $p_a(k_{1/2}) = 0.5$ ). In this limit, the shape of the regulation function must approach a Hill function with the Hill coefficient of 2,  $p_a(k) = k^2 / (k^2 + k_{1/2}^2)$ . Surprisingly, information maximization favors weak affinity of individual TF molecules to the DNA, accompanied by strong cooperativity between these molecules. The OR-architecture portrays a different picture: here, the maximum of information is well-defined for a particular combination of parameters  $(k_-, \gamma)$ , as shown in Fig. 5 D. As  $\gamma \rightarrow 0$  (increasing destabilization for  $\gamma < 1$ ), the second active state ( $b$ ) is never occupied, and the model reverts to a two-state model.

For both architectures we can assess the advantage of the three-state model relative to the optimal two-state promoter. Fig. 5 E and F show the information of the optimal solutions as a function of the input noise magnitude as well as the input range,  $k_{\max}$ . As expected, the information increases as a function of  $k_{\max}$  since the influence of input and switching noise can be made smaller with more input molecules. This increase saturates at high  $k_{\max}$  because output noise becomes limiting to the information transmission—this is why the capacity curves converge to the same maximum, the  $\nu = 0$  curve that lacks the input noise altogether. The advantage (increase in capacity) of the three-state models relative to the two-state promoter is positive for any combination of parameters  $k_{\max}$  and  $\nu$ . It is interesting to note that increasing  $k_{\max}$  and decreasing  $\nu$  have very similar effects on channel capacity, since both drive the system to a regime where the limiting factor is the output noise.

#### Regulation with dual-role TFs

In the second architecture that we consider a transcription factor can switch its role from repressor to activator, depending on the covalent TF modification state or formation of a complex with specific co-factors. A well-studied example is in Hedgehog (Hh) signaling, where the TF Gli acts as a repressor when Hh is low, or as an activator when Hh is high (85–87). Fig. 6 A shows a possible dual-role signaling scheme where the total concentration of dual-role TFs is fixed (at  $k_{\max}$ ), but the signal modulates the fraction of these TFs that play the activator role ( $k$ ) and the remaining fraction that act as repressors ( $k_{\max} - k$ ), which compete for the same binding site. The channel capacity of this motif is depicted in Fig. 6 B as a function of promoter parameters  $k_A$  and  $k_R$ , showing that a globally optimal setting (denoted “A”) exists for these parameters; with these parameters, the input/output function, shown in Fig. 6 C, is much steeper than what could be achieved with the best two-state promoter, and that is

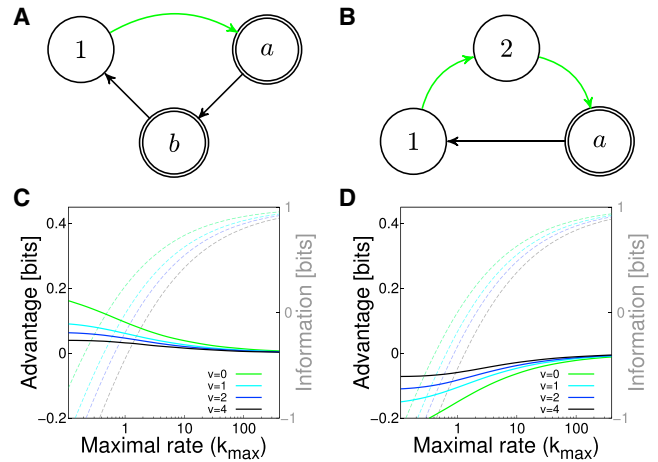


**FIGURE 6** Improving information transmission via dual-role regulation. (A) The signal  $k$  increases the concentration of TFs in the activator role that favor the transition (green) to the expressing state, while simultaneously decreasing the rate of switching (red) into the inactive state with the repressor bound. (B) Channel capacity (color) as a function of off-rates ( $k_A, k_R$ ) shows a peak at A.  $k_{max} = 10$  and  $v = 4$ . (C) The regulatory functions for the optimal solution A (solid blue line) and other example points (B to E) from the information plane, show that this architecture can access a rich range of response steepnesses and induction thresholds. For comparison, we also plot the regulation curve of the scheme 2-1a (with its optimal  $k^*$ ). (D) The channel capacity (dashed line) and information advantage over the optimal two-state architecture (solid line), as a function of  $k_{max}$ . To see this figure in color, go online.

true despite the fact that the molecular implementation of this architecture uses only a single binding site. The ability to access such steep regulatory curves allows this architecture to position the midpoint of induction at higher inputs  $k$ , thus escaping the detrimental effects of the input noise at low  $k$ , while still being able to induce almost completely (i.e., make use of the full dynamic range of outputs) as the input varies from 0 to  $k_{max}$ . This is how the dual-role regulation can escape the tradeoff faced by the two-state model 2-1a (shown in Fig. 4). Sharper transition at higher input would lead us to expect that the advantage of this architecture over the two-state model is most pronounced when input noise is dominant (small  $k_{max}$ , large  $v$ ), which is indeed the case, as shown in Fig. 6 D.

*Promoter cycling*

In the last architecture considered in this article, promoters “cycle” through a sequence of states in a way that does not obey detailed balance, e.g., when state transitions involve expenditure of energy during irreversible reaction steps. In the scheme shown in Fig. 7 A, the regulated transition puts the promoter into an active state  $a$ ; before decaying to an inactive state, the promoter must transition through another active state  $b$ . Effectively, this scheme is similar to



**FIGURE 7** Improving information transmission via cycling. (A) Promoter cycles through two active states ( $a, b$ ) expressing at identical production rates before returning to inactive state (1), from which the transition rate back into the active state (green) is modulated. For each value of  $k_{max}$ , we look for the optimal choice of  $k_{ab} = k_{b1}$ . The information and advantage relative to the optimal two-state model is shown in (C). (B,D) A similar architecture where turning the gene on is a multistep regulated process. This architecture always underperforms the optimal two-state model, indicated by a negative value of the advantage for all choices of  $v$  and  $k_{max}$ . To see this figure in color, go online.

the two-state model in which the decay from the active state is not first-order with exponentially distributed transition times, but rather with transition times that have a sharper peak. The benefits of this architecture are maximized when the transition rates from both active states are equal. Although it always outperforms the optimal two-state model, the largest advantage is achievable for small  $k_{max}$ . At large  $k_{max}$  the advantage tends to zero: this is because the optimal off-rates are high, causing the dwell times in the expressing states to be short. In this regime the gamma distribution of dwell times (in a three-state model) differs little from the exponential distribution (in a two-state model). Note that this model would not yield any information advantage if the state transitions were reversible.

Fig. 7 C and D show that irreversible transitions alone do not generate an information advantage: a promoter that needs to transition between two inactive states (1, 2) to reach a single expressing state  $a$  from which it exits in a first-order transition, is always at a loss compared with a two-state promoter. This is because here the effective transition rate to the active state in the equivalent two-state model is lower (since an intermediate state must be traversed to induce), necessitating the use of a lower off-rate  $k_{a1}$ , which in turn leads to higher switching noise.

It is interesting to note that recent experimental data on eukaryotic transcription seem to favor models in which the distribution of exit times from the expressing state is exponential, whereas the distribution of times from the inactive into the active state is not (59), pointing to the seemingly underperforming architecture of Fig. 7 B.



The three-state promoter suggested in this study is probably an oversimplified model of reality, yet it nevertheless makes sense to ask why irreversible transitions through multiple states are needed to switch on the transcription of eukaryotic genes (88,89) and how this observation can be reconciled with the lower regulatory power of such architectures. This is the topic of our ongoing research.

## DISCUSSION

When studying noise in gene regulation one is usually restricted to the use of phenomenological models, rather than a fully detailed biochemical reaction scheme. Simpler models, such as those studied in this paper, also allow us to decouple questions of mechanistic interpretation from the questions of functional consequences. Here, we extended a well-known functional two-state model of gene expression to multiple internal states. We introduced state transition graphs to model the “decision logic” by which changes in the concentrations of regulatory proteins drive the switching of our genes between various states of expression. This abstract language allowed us to systematically organize and explore nonequivalent three-state promoters. The advantage of this approach is that many microscopically distinct regulatory schemes can be collapsed into equivalent classes sharing identical state transition graphs and identical information transmission properties.

The functional description of multistate promoters confers two separate benefits. First, it is able to generate measurable predictions, such as the noise vs. mean induction curve. Existing experimental and theoretical work using the two-state model has demonstrated how the measurements of noise constrain the space of promoter models (44), how the theory establishes the “vocabulary” by which various measured promoters can be classified and compared with each other (90), and how useful a baseline mathematical model can be in establishing quantitative signatures of deviation that, when observed, must lead to minimal model revisions able to accommodate new data (59). Alternative complex promoters presented here could explain existing data better either because of the inclusion of additional states (see (52)), or because we also included and analyzed the effects of input (diffusive) noise, which can mimic the effects of promoter switching noise but is often neglected (45). As a caveat, it appears that in many cases discriminating between promoter architectures based on the noise characteristics alone would be very difficult, and thus dynamical measurements would be necessary.

The second benefit of our approach is to provide a convenient framework for assessing the functional impact of noise in gene regulation, as measured by the mutual information between the inputs and the gene expression level. We are interested in the question whether multistate promoters can, at least in principle, perform better than the simple ON/OFF two-state model. We find that generically, i.e.,

for all three-state models where one state is expressing and only one transition is modulated by the input, the multistate promoters underperform the two-state model. Higher information transmission can be achieved when these conditions are violated, and biological examples for such violations can be found. For example, we find that a multistate promoter with cooperativity has a higher channel capacity than the best comparable two-state promoter, even when promoter switching noise is taken into account (see (11)). Dual-regulation yields surprisingly high benefits, which are largest when input noise is high. In the context of metazoan development where the concentrations of the morphogen molecules can be in the nanomolar range and the input noise is therefore high (35), the need to establish sharp spatial domains of downstream gene expression (as observed (91),) might have favored such dual-role promoter architectures. Lastly, we considered the simplest ideas for a promoter with irreversible transitions and have shown that they can lead to an increase in information transmission by sharpening the distribution of exit times from the expressing state (92).

The main conclusion of this article—namely that channel capacity can be increased by particular complex promoters—is testable in dedicated experiments. One could start with a simple regulatory scheme in a synthetic system and then by careful manipulation gradually introduce the possibility of additional states (e.g., by introducing more binding sites), using promoter sequences that show weaker binding for individual molecules yet allow for stronger cooperative interaction. In both the simple and complex system, one could then measure the noise behavior for various input levels. Information theoretic analysis of the resulting data could be used to judge if the design of higher complexity, although perhaps noisier by some other measure, is capable of transmitting more information, as predicted.

The list of multistate promoters that can outperform the two-state regulation and for which examples in nature could be found is potentially much longer and could include combinations of features described in this article. Rather than trying to find more examples, we should perhaps ask about fundamentally different mechanisms and constraints that our analysis did not consider. By expending energy to keep the system out of equilibrium, one could design robust reaction schemes where, for example, the binding of a regulatory protein leads (almost) deterministically to some tightly controlled response cycle, perhaps evading the diffusion noise limit (93) and increasing information transmission. At the same time, cells might be confronted by sources of stochasticity we did not discuss in this article—for example, because of cross-talk from spurious binding of noncognate regulators. Finally, cells need to not only *transmit* information through their regulatory elements, but actually perform *computations*, that is, combine various inputs into a single output, thereby potentially discarding information. A challenging question for the future is thus

about extending the information-theoretic framework to these other cases of interest.

## SUPPORTING MATERIAL

Two figures and supplemental information are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(14\)00082-4](http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00082-4).

We thank W. Bialek, T. Gregor, I. Golding, M. Dravecká, K. Mitosch, G. Chevereau, J. Briscoe, and J. Kondev for helpful discussions.

## REFERENCES

- Levine, M., and E. H. Davidson. 2005. Gene regulatory networks for development. *Proc. Natl. Acad. Sci. USA*. 102:4936–4942.
- Ronen, M., R. Rosenberg, ..., U. Alon. 2002. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc. Natl. Acad. Sci. USA*. 99:10555–10560.
- Elowitz, M. B., A. J. Levine, ..., P. S. Swain. 2002. Stochastic gene expression in a single cell. *Science*. 297:1183–1186.
- Ozbudak, E. M., M. Thattai, ..., A. van Oudenaarden. 2002. Regulation of noise in the expression of a single gene. *Nat. Genet.* 31:69–73.
- Paulsson, J. 2004. Summing up the noise in gene networks. *Nature*. 427:415–418.
- Raj, A., and A. van Oudenaarden. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*. 135:216–226.
- Houchmandzadeh, B., E. Wieschaus, and S. Leibler. 2002. Establishment of developmental precision and proportions in the early *Drosophila* embryo. *Nature*. 415:798–802.
- Tkačik, G., and A. M. Walczak. 2011. Information transmission in genetic regulatory networks: a review. *J. Phys. Condens. Matter*. 23:153102.
- Ziv, E., I. Nemenman, and C. H. Wiggins. 2007. Optimal signal processing in small stochastic biochemical networks. *PLoS ONE*. 2:e1077.
- Tkačik, G., C. G. Callan, Jr., and W. Bialek. 2008. Information capacity of genetic regulatory elements. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 78:011910.
- Tkačik, G., A. M. Walczak, and W. Bialek. 2009. Optimizing information flow in small genetic networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 80:031920.
- Walczak, A. M., G. Tkačik, and W. Bialek. 2010. Optimizing information flow in small genetic networks. II. Feed-forward interactions. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 81:041905.
- Tkačik, G., A. M. Walczak, and W. Bialek. 2012. Optimizing information flow in small genetic networks. III. A self-interacting gene. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 85:041903.
- Tkačik, G., C. G. Callan, Jr., and W. Bialek. 2008. Information flow and optimization in transcriptional regulation. *Proc. Natl. Acad. Sci. USA*. 105:12265–12270.
- Dubuis, J. O., G. Tkačik, ..., W. Bialek. 2013. Positional information, in bits. *Proc. Natl. Acad. Sci. USA*. 110:16301–16308.
- Tostevin, F., and P. R. ten Wolde. 2009. Mutual information between input and output trajectories of biochemical networks. *Phys. Rev. Lett.* 102:218101.
- Tostevin, F., and P. R. ten Wolde. 2010. Mutual information in time-varying biochemical systems. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 81:061917.
- Cheong, R., A. Rhee, ..., A. Levchenko. 2011. Information transduction capacity of noisy biochemical signaling networks. *Science*. 334:354–358.
- de Ronde, W. H., F. Tostevin, and P. R. ten Wolde. 2010. Effect of feedback on the fidelity of information transmission of time-varying signals. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 82:031914.
- de Ronde, W. H., F. Tostevin, and P. R. ten Wolde. 2012. Feed-forward loops and diamond motifs lead to tunable transmission of information in the frequency domain. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 86:021913.
- de Ronde, W., F. Tostevin, and P. R. ten Wolde. 2011. Multiplexing biochemical signals. *Phys. Rev. Lett.* 107:048101.
- Tostevin, F., W. de Ronde, and P. R. ten Wolde. 2012. Reliability of frequency and amplitude decoding in gene regulation. *Phys. Rev. Lett.* 108:108104.
- Mugler, A., F. Tostevin, and P. R. ten Wolde. 2013. Spatial partitioning improves the reliability of biochemical signaling. *Proc. Natl. Acad. Sci. USA*. 110:5927–5932.
- Bowsher, C. G., and P. S. Swain. 2012. Identifying sources of variation and the flow of information in biochemical networks. *Proc. Natl. Acad. Sci. USA*. 109:E1320–E1328.
- Jost, D., A. Nowojewski, and E. Levine. 2013. Regulating the many to benefit the few: role of weak small RNA targets. *Biophys. J.* 104:1773–1782.
- Hormoz, S. 2013. Cross talk and interference enhance information capacity of a signaling pathway. *Biophys. J.* 104:1170–1180.
- Levine, J., H. Y. Kueh, and L. Mirny. 2007. Intrinsic fluctuations, robustness, and tunability in signaling cycles. *Biophys. J.* 92:4473–4481.
- Lalanne, J.-B., and P. François. 2013. Principles of adaptive sorting revealed by *in silico* evolution. *Phys. Rev. Lett.* 110:218102.
- Mancini, F., C. H. Wiggins, ..., A. M. Walczak. 2013. Time-dependent information transmission in a model regulatory circuit. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 88:022708.
- Taylor, S. F., N. Tishby, and W. Bialek. 2007. Information and fitness. *arXiv* 0712.4382.
- Rivoire, O., and S. Leibler. 2011. The value of information for populations in varying environments. *J. Stat. Phys.* 142:1124–1166.
- Donaldson-Matasci, M. C., C. T. Bergstrom, and M. Lachmann. 2010. The fitness value of information. *Oikos*. 119:219–230.
- Bialek, W., and S. Setayeshgar. 2005. Physical limits to biochemical signaling. *Proc. Natl. Acad. Sci. USA*. 102:10040–10045.
- Bialek, W., and S. Setayeshgar. 2008. Cooperativity, sensitivity, and noise in biochemical signaling. *Phys. Rev. Lett.* 100:258101.
- Gregor, T., D. W. Tank, ..., W. Bialek. 2007. Probing the limits to positional information. *Cell*. 130:153–164.
- van Kampen, N. 1981. Stochastic processes in physics and chemistry. North Holland, Amsterdam.
- Peccoud, J., and B. Ycart. 1995. Markovian modeling of gene-products synthesis. *Theor. Popul. Biol.* 48:222–234.
- Iyer-Biswas, S., F. Hayot, and C. Jayaprakash. 2009. Stochasticity of gene products from transcriptional pulsing. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 79:031911.
- Shahrezaei, V., and P. S. Swain. 2008. Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. USA*. 105:17256–17261.
- Dobrzynski, M., and F. J. Bruggeman. 2009. Elongation dynamics shape bursty transcription and translation. *Proc. Natl. Acad. Sci. USA*. 106:2583–2588.
- Kepler, T. B., and T. C. Elston. 2001. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* 81:3116–3136.
- Hu, B., D. A. Kessler, ..., H. Levine. 2011. Effects of input noise on a simple biochemical switch. *Phys. Rev. Lett.* 107:148101.
- Hu, B., D. A. Kessler, ..., H. Levine. 2012. How input fluctuations reshape the dynamics of a biological switching system. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 86:061910.
- So, L.-H., A. Ghosh, ..., I. Golding. 2011. General properties of transcriptional time series in *Escherichia coli*. *Nat. Genet.* 43:554–560.

45. Tkačik, G., T. Gregor, and W. Bialek. 2008. The role of input noise in transcriptional regulation. *PLoS ONE*. 3:e2774.
46. Raser, J. M., and E. K. O'Shea. 2004. Control of stochasticity in eukaryotic gene expression. *Science*. 304:1811–1814.
47. Raj, A., C. S. Peskin, ..., S. Tyagi. 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol*. 4:e309.
48. Sánchez, A., and J. Kondev. 2008. Transcriptional control of noise in gene expression. *Proc. Natl. Acad. Sci. USA*. 105:5081–5086.
49. Gutiérrez, P. S., D. Monteoliva, and L. Diambra. 2009. Role of cooperative binding on noise expression. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys*. 80:011914.
50. Karmakar, R. 2010. Conversion of graded to binary response in an activator-repressor system. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys*. 81:021905.
51. Coulon, A., O. Gandrillon, and G. Beslon. 2010. On the spontaneous stochastic dynamics of a single gene: complexity of the molecular interplay at the promoter. *BMC Syst. Biol*. 4:2.
52. Sanchez, A., H. G. Garcia, ..., J. Kondev. 2011. Effect of promoter architecture on the cell-to-cell variability in gene expression. *PLOS Comput. Biol*. 7:e1001100.
53. Zhou, T., and J. Zhang. 2012. Analytical results for a multistate gene model. *SIAM J. Appl. Math*. 72:789–818.
54. Zhang, J., L. Chen, and T. Zhou. 2012. Analytical distribution and tunability of noise in a model of promoter progress. *Biophys. J*. 102:1247–1257.
55. Gutierrez, P. S., D. Monteoliva, and L. Diambra. 2012. Cooperative binding of transcription factors promotes bimodal gene expression response. *PLoS ONE*. 7:e44812.
56. Blake, W. J., G. Balázsi, ..., J. J. Collins. 2006. Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell*. 24:853–865.
57. Blake, W. J., M. Kærn, ..., J. J. Collins. 2003. Noise in eukaryotic gene expression. *Nature*. 422:633–637.
58. Kandhavelu, M., H. Mannerström, ..., A. S. Ribeiro. 2011. In vivo kinetics of transcription initiation of the *lac* promoter in *Escherichia coli*. Evidence for a sequential mechanism with two rate-limiting steps. *BMC Syst. Biol*. 5:149.
59. Suter, D. M., N. Molina, ..., F. Naef. 2011. Mammalian genes are transcribed with widely different bursting kinetics. *Science*. 332:472–474.
60. Setty, Y., A. E. Mayo, ..., U. Alon. 2003. Detailed map of a cis-regulatory input function. *Proc. Natl. Acad. Sci. USA*. 100:7702–7707.
61. Shannon, C. 1948. A mathematical theory of communication. *Bell Syst. Tech. J*. 27:379–423, 623–56.
62. Cover, T., and J. Thomas. 2006. *Elements of Information Theory*. Wiley Interscience, New York.
63. Walczak, A. M., M. Sasai, and P. G. Wolynes. 2005. Self-consistent proteomic field theory of stochastic gene switches. *Biophys. J*. 88:828–850.
64. Gillespie, D. T. 2000. The chemical Langevin equation. *J. Chem. Phys*. 113:297.
65. Swain, P. S., M. B. Elowitz, and E. D. Siggia. 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA*. 99:12795–12800.
66. Berg, H. C., and E. M. Purcell. 1977. Physics of chemoreception. *Biophys. J*. 20:193–219.
67. Berg, O. G., and P. H. von Hippel. 1985. Diffusion-controlled macromolecular interactions. *Annu. Rev. Biophys. Biophys. Chem*. 14:131–160.
68. Endres, R. G., and N. S. Wingreen. 2008. Accuracy of direct gradient sensing by single cells. *Proc. Natl. Acad. Sci. USA*. 105:15749–15754.
69. Carey, L. B., D. van Dijk, ..., E. Segal. 2013. Promoter sequence determines the relationship between expression level and noise. *PLoS Biol*. 11:e1001528.
70. de Ronde, W. H., B. C. Daniels, ..., I. Nemenman. 2009. Mesoscopic statistical properties of multistep enzyme-mediated reactions. *IET Syst. Biol*. 3:429–437.
71. Sanchez, A., S. Choubey, and J. Kondev. 2013. Regulation of noise in gene expression. *Annu. Rev. Biophys*. 42:469–491.
72. Golding, I., J. Paulsson, ..., E. C. Cox. 2005. Real-time kinetics of gene activity in individual bacteria. *Cell*. 123:1025–1036.
73. Geertz, M., D. Shore, and S. J. Maerkl. 2012. Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proc. Natl. Acad. Sci. USA*. 109:16540–16545.
74. Viñuelas, J., G. Kaneko, ..., O. Gandrillon. 2013. Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts. *BMC Biol*. 11:15.
75. Sanchez, A., M. L. Osborne, ..., J. Gelles. 2011. Mechanism of transcriptional repression at a bacterial promoter by analysis of single molecules. *EMBO J*. 30:3940–3946.
76. Garcia, H. G., A. Sanchez, ..., R. Phillips. 2012. Operator sequence alters gene expression independently of transcription factor occupancy in bacteria. *Cell Rep*. 2:150–161.
77. Yang, H.-T., and M. S. H. Ko. 2012. Stochastic modeling for the expression of a gene regulated by competing transcription factors. *PLoS ONE*. 7:e32376.
78. Earnest, T. M., E. Roberts, ..., Z. Luthey-Schulten. 2013. DNA looping increases the range of bistability in a stochastic model of the *lac* genetic switch. *Phys. Biol*. 10:026002.
79. Mirny, L. A. 2010. Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci. USA*. 107:22534–22539.
80. Segal, E., and J. Widom. 2009. What controls nucleosome positions? *Trends Genet*. 25:335–343.
81. Li, G.-W., O. Berg, and J. Elf. 2009. Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nat. Phys*. 5:294–297.
82. Grönlund, A., P. Lötstedt, and J. Elf. 2013. Transcription factor binding kinetics constrain noise suppression via negative feedback. *Nat. Commun*. 4:1864.
83. Lehner, B. 2010. Conflict between noise and plasticity in yeast. *PLoS Genet*. 6:e1001185.
84. Bajič, D., and J. F. Poyatos. 2012. Balancing noise and plasticity in eukaryotic gene expression. *BMC Genomics*. 13:343.
85. Parker, D. S., M. A. White, ..., S. Barolo. 2011. The cis-regulatory logic of Hedgehog gradient responses: key roles for gli binding affinity, competition, and cooperativity. *Sci. Signal*. 4:ra38.
86. White, M. A., D. S. Parker, ..., B. A. Cohen. 2012. A model of spatially restricted transcription in opposing gradients of activators and repressors. *Mol. Syst. Biol*. 8:614.
87. Müller, B., and K. Basler. 2000. The repressor and activator forms of *Cubitus interruptus* control Hedgehog target genes through common generic gli-binding sites. *Development*. 127:2999–3007.
88. Larson, D. R. 2011. What do expression dynamics tell us about the mechanism of transcription? *Curr. Opin. Genet. Dev*. 21:591–599.
89. Yosef, N., and A. Regev. 2011. Impulse control: temporal dynamics in gene transcription. *Cell*. 144:886–896.
90. Zenklusen, D., D. R. Larson, and R. H. Singer. 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol*. 15:1263–1271.
91. Dessaud, E., A. P. McMahon, and J. Briscoe. 2008. Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. *Development*. 135:2489–2503.
92. Pedraza, J. M., and J. Paulsson. 2008. Effects of molecular memory and bursting on fluctuations in gene expression. *Science*. 319:339–343.
93. Aquino, G., and R. G. Endres. 2010. Increased accuracy of ligand sensing by receptor internalization. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys*. 81:021909.

Supplement to “Noise and information transmission in  
promoters with multiple internal states”

Georg Rieckh<sup>1</sup>  
IST Austria,  
Klosterneuburg, Austria

Gašper Tkačik  
IST Austria,  
Klosterneuburg, Austria

January 7, 2014

<sup>1</sup>Corresponding author. Address: Institute of Science and Technology Austria, Am Campus 1,  
A-3400 Klosterneuburg, Austria Tel.: +43(650)8415095



## 1 Experimentally measured promoter switching rates

Direct measurements of switching rates are rare since they require live imaging. Examples include the relative measurements of on-, off- and mRNA-production rates in *E. coli* (44) using the MS2-GFP system (72), reporting 2 – 10 fold higher on- than off-rates, and mRNA production rates an order of magnitude higher than the on-rates; original bursting reported in (72) finds the on-time duration to be roughly 6 and the off time 37 minutes in a synthetic *E. coli* reporter system. Recently, on-rates of  $\sim 3 \cdot 10^{-2} \text{ min}^{-1}$ , roughly ten-fold higher off-rates, and mRNA production rates ranging from 0 – 5  $\text{min}^{-1}$  have been reported in mammalian cells using the luciferase reporter system (59). Using new high-throughput microfluidic methods, it is now possible to measure TF binding and unbinding times directly: Ref (73) reports mouse and yeast in vitro transcription factor dissociation rates between  $\sim 10 \text{ s}^{-1}$  and  $10^{-2} \text{ s}^{-1}$ , as well as the range of the corresponding association rates; it is, however, less clear if these can be unambiguously identified with switching rates in functional models.

A larger body of work extracts the rates of the two-state model from the noise characteristics (which are the primary measurement), *assuming* the two-state model without diffusion noise is applicable. The reported Fano factors for mRNA counts vary, but are of the order of 1 – 10. The typical values for kinetic parameters extracted for a range of *E. coli* promoters are  $10^{-3} - 10^{-2} \text{ s}^{-1}$  for the on-rate,  $10^{-1} - 1 \text{ s}^{-1}$  for the mRNA production rate when induced, and a variable off-rate that depends strongly on the induction level (44). Using a similar technique in mammalian cells, Raj et al (47) extracted two-state parameters and found the on-rate normalized by mRNA decay time to be roughly of order unity, while the ratio of mRNA production rate to the off-rate varied from  $\sim 10 - 400$ , depending on the system and the induction level.

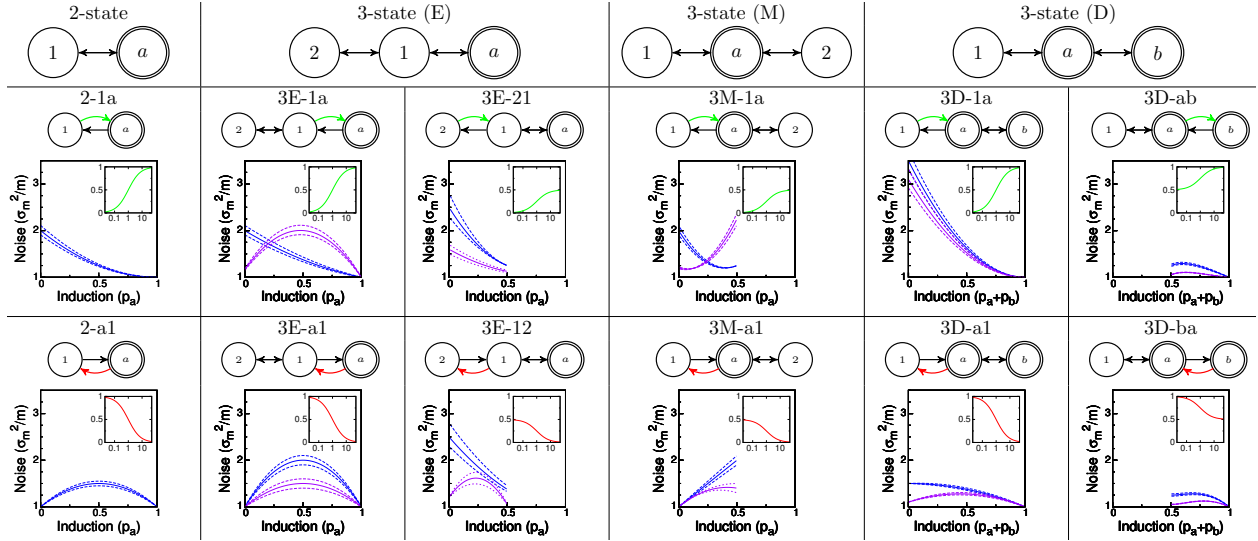


Figure S1: **Promoter architectures and interpretations.** **Scheme 2-1a:** (i) State  $a$  is the empty promoter (available for transcription), and state 1 corresponds to it being occupied by a repressing protein (e.g. a specific TF or a nucleosome). Mechanisms that change the rate of switching from state 1 into  $a$  (i.e. scheme 2-1a) are well documented for eukaryotic cells (69, 74). (ii) Simple activation where state 1 is the empty promoter and state  $a$  has an activating TF bound to it; changing TF concentration modulates  $k_{1a}$  (41, 52). **Scheme 2-a1:** The bacterial *lac*-promoter, where the binding of a specific TF represses expression. A change in the concentration of TFs now corresponds to modulating the rate  $k_{a1}$  and the rate  $k_{1a}$  is determined by the interaction energy between TF and its binding site (75, 76). **Scheme 3E:** (i) A promoter that has overlapping binding sites for both, an activator and a repressor (dual regulation) (48, 50, 77). Changing the concentration of the activator (repressor) leads to a change in rate  $k_{1a}$  ( $k_{12}$ ), which means it can be modeled as a 3E-1a (3E-12) scheme. (ii) Eukaryotic promoters with a TATA box can be modeled as 3E-21 (71). Here, state 2 is the DNA in a state not available for transcription (closed chromatin), state 1 is the conformation where the promoter (and the TATA box) is exposed and state  $a$  corresponds to the active configuration of the DNA with the pre-initiation complex assembled. Changing the concentration of chromatin remodelers now influences the rate  $k_{21}$  (similar to the  $k_{1a}$ -modulation in the scheme 2-1a mentioned above), which yields scheme 3E-21. (iii) A coarse-grained model of DNA-looping in the *lac*-operon (78). **Scheme 3M:** A nucleosome and a specific, repressing TF compete for a promoter; changes in the input TF concentration correspond to 3M-a1, while changes in factors decreasing nucleosome occupancy correspond to 3M-1a. **Scheme 3D:** (i) State 1 is the closed chromatin formation, state  $a$  is the empty promoter and in state  $b$  a TF is bound to the DNA in such a way that it prohibits the closing of the chromatin (but still permits transcription) (79, 80). In this way, even though the input molecule does not necessarily interact with the RNA polymerase directly, it can act as an activator (or rather as a de-repressor), yielding scheme 3D-ab. (ii) A scheme used to describe any promoter where the basal expression does not follow a Poisson process (optionally with different rates of expression from  $a$  and  $b$ ). (iii) Promoter with a TATA box and a competing nucleosome (cf. 3E-21) if there is a significant amount of expression from the basal state  $a$ . (iv) A scheme used as a phenomenological model with unidentified third state to explain universal noise behavior in bacterial gene expression (44, 71). Cited references use similar models.

## 2 Multi-state promoters as state-transition graphs

In this section we describe the general method used to derive the behavior of noise and mean for different promoter architectures, followed by a calculation for one example architecture.

### 2.1 Translating a state transition diagram into dynamic equations

Let  $\{a, b, \dots, M\}$  denote the states of the promoter that produce mRNA at a fixed rate  $r$  and  $\{1, 2, \dots, N\}$  denote states without production. For  $\mathcal{S} = \{a, b, \dots, M, 1, 2, \dots, N\} \ni i, j$ , let  $k_{ij} \geq 0, i \neq j$  be the rate with which the promoter switches from state  $i$  to state  $j$ ,  $d$  be the rate of mRNA-degradation, and  $p_i$  be the fractional occupancy of state  $i$ . For simplicity, we will only treat the case  $M = 1$  here.

**Deterministic equations.** The list of (non-zero) rates fully defines the state-transition graph, i.e. the promoter model. This directly translates into a linear system of equations that describes the dynamics of the system:

$$\partial_t \mathbf{p} = \mathbf{K} \mathbf{p} \quad , \quad \text{with} \quad (1)$$

$$\mathbf{K} = \begin{bmatrix} -\sum_{j \in \mathcal{S}} k_{aj} & k_{1a} & \cdots & k_{Na} \\ k_{a1} & -\sum_{j \in \mathcal{S}} k_{1j} & \cdots & k_{N1} \\ \vdots & \vdots & \ddots & \vdots \\ k_{aN} & k_{1N} & \cdots & -\sum_{j \in \mathcal{S}} k_{Nj} \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} p_a \\ p_1 \\ \vdots \\ p_N \end{bmatrix}, \quad (2)$$

subject to the normalization constraint  $\sum_{i \in \mathcal{S}} p_i = 1$ .

The dynamics of mRNA are described by linking them to the activity of the promoter:

$$\partial_t m = r p_a - d m \quad . \quad (3)$$

To compute the average amount of mRNA  $\bar{m}$  in steady state, we set the time derivatives to 0 and solve the linear set of equations

$$\mathbf{K} \bar{\mathbf{p}} = 0 \quad , \quad (4)$$

$$\bar{m} = \frac{r}{d} p_a \quad . \quad (5)$$

As the occupancy of the active state  $p_a$  is a function of the rates in  $\mathbf{K}$ , we can obtain the dependence of  $\bar{m}$  on any rate of interest, i.e. we can obtain the regulation function.

**Langevin approach to calculate noise behavior.** For the noise behavior, we linearize Eqs (4,5) of the main text around the mean:

$$\mathbf{p}(t) = \bar{\mathbf{p}} + \delta \mathbf{p}(t), \quad (6)$$

$$m(t) = \bar{m} + \delta m(t) \quad (7)$$

and introduce the Fourier-transformed variables

$$\delta p_i(t) = (2\pi)^{-1} \int d\omega \delta \hat{p}_i(\omega) \exp(-i\omega t) \quad , \quad (8)$$

$$\delta m(t) = (2\pi)^{-1} \int d\omega \delta \hat{m}(\omega) \exp(-i\omega t) \quad , \quad (9)$$

so that we get the linear response to random fluctuations:

$$(-i\omega)\delta\hat{\mathbf{p}} = \mathbf{K}\delta\hat{\mathbf{p}} + \hat{\boldsymbol{\xi}} \quad , \quad (10)$$

$$(-i\omega)\delta\hat{m} = r\delta\hat{p}_a - d\delta\hat{m} + \hat{\xi}_m \quad . \quad (11)$$

The statistics of the Langevin forces are given by:

$$\langle \hat{\xi}_i^* \hat{\xi}_j \rangle = -(\hat{p}_i K_{ij} + \hat{p}_j K_{ji}) \quad , \quad (12)$$

$$\langle \hat{\xi}_m^* \hat{\xi}_m \rangle = 2d\bar{m} \quad ; \quad (13)$$

to see this for the variances, consider  $\langle \xi_i \xi_i^* \rangle = -2\hat{p}_i K_{ii} = 2\hat{p}_i \sum_j k_{ij}$ , since all entries in the diagonal of  $\mathbf{K}$  are negative. This is two times the rate of leaving state  $i$ . Similarly, for  $\langle \xi_m^* \xi_m \rangle$  the variance is two times the rate of degrading a molecule. The factor of two comes from the fact that we consider a system at steady state, so the rates of entering and leaving a state (or creating an destroying a molecule) must be equal. For the covariances  $\langle \xi_i^* \xi_j \rangle$  ( $i \neq j$ ), the two Langevin forces are anti-correlated, since leaving one state means entering another. The rate of changing between the two states is the probability of being in state  $i$  ( $p_i$ ) times the rate of transition from that state into the other ( $k_{ij} = K_{ij}$ ) – and this holds for both directions between the pair of states. Also, since we assume that production of mRNA and promoter switching are independent,  $\langle \xi_i^* \xi_m \rangle = 0$  for all states  $i$ .

To get the variance in mRNA, we compute  $\sigma_m^2 = (2\pi)^{-1} \int d\omega |\delta\hat{m}(\omega)|^2$ , where  $\delta\hat{m}(\omega)$  is obtained from Eq (11) as

$$\langle \delta\hat{m}^* \delta\hat{m} \rangle = \frac{2d\bar{m}}{d^2 + \omega^2} + \frac{r^2}{d^2 + \omega^2} \langle \delta\hat{p}_a^* \delta\hat{p}_a \rangle \quad , \quad (14)$$

where  $\langle \delta\hat{p}_a^* \delta\hat{p}_a \rangle$  is calculated by solving Eq (10) and using the Langevin noise magnitudes from Eqs (12,13).

With the assumption  $d \ll k_{ij}$ , Eq (10) becomes

$$0 = \mathbf{K}\delta\hat{\mathbf{p}} + \hat{\boldsymbol{\xi}} \quad , \quad (15)$$

$$\sum_i \delta\hat{p}_i = 0 \quad , \quad (16)$$

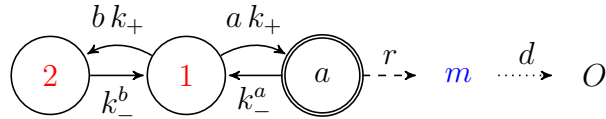
which simplifies the expressions for the  $\delta\hat{p}_i$ . This is because the terms with  $(-i\omega)$  in the denominator (as seen in the next section in Eqs (36,37)) would give an additional, multiplicative term of the form  $1/(k_{ij}^2 + \omega^2)$  in Eq (14). The  $\omega$ -dependence of these terms can be neglected for the integration, since for  $d \ll k_{ij}$  we have

$$\frac{1}{k_{ij}^2 + \omega^2} \frac{1}{d^2 + \omega^2} \approx \frac{1}{k_{ij}^2} \frac{1}{d^2 + \omega^2} \quad . \quad (17)$$



## 2.2 Example: Dual regulation (3E-1a)

We are interested in a system where the promoter of a gene can either be occupied by an activator (present at concentration  $a$ ) or a repressor (present at concentration  $b$ ). If it is in the active state, it produces mRNA at a constant rate  $r$ , which is later degraded at rate  $d$ .



**Deterministic equations.** Following the setup from the last section, we translate the state transition diagram into a matrix that describes the dynamics at the promoter:

$$\begin{bmatrix} \partial_t p_a \\ \partial_t p_1 \\ \partial_t p_2 \end{bmatrix} = \begin{bmatrix} -k_-^a & ak_+ & 0 \\ k_-^a & -(ak_+ + bk_+) & k_-^b \\ 0 & bk_+ & -k_-^b \end{bmatrix} \cdot \begin{bmatrix} p_a \\ p_1 \\ p_2 \end{bmatrix} . \quad (18)$$

This is then the basis for a description of the dynamics of the output (here mRNA):

$$\partial_t m = rp_a - dm \quad , \quad (19)$$

$$\partial_t p_a = ak_+p_1 - k_-^a p_a \quad , \quad (20)$$

$$\partial_t p_2 = bk_+p_1 - k_-^b p_2 \quad , \quad (21)$$

$$p_a + p_1 + p_2 = 1 \quad . \quad (22)$$

With the definitions  $A = \frac{ak_+}{k_-^a}$ ,  $B = \frac{bk_+}{k_-^b}$ ,  $S = 1 + A + B$  and  $R = \frac{r}{d}$  we get for the steady state:

$$\bar{p}_a = A/S, \quad \bar{p}_1 = 1/S, \quad \bar{p}_2 = B/S \quad , \quad (23)$$

$$\bar{m} = R\bar{p}_a = RA/S \quad . \quad (24)$$

**Langevin approach.** To see how the dynamics of the promoter influence the statistics of mRNA we perturb the systems with Langevin forces (while still keeping the gene copy number constant):

$$\partial_t m = rp_a - dm + \xi_m \quad , \quad (25)$$

$$\partial_t p_a = ak_+p_1 - k_-^a p_a + \xi_a \quad , \quad (26)$$

$$\partial_t p_2 = bk_+p_1 - k_-^b p_2 + \xi_2 \quad , \quad (27)$$

$$p_a + p_1 + p_2 = 1 \quad . \quad (28)$$

The mean of the Langevin forces is zero ( $\langle \xi_i(t) \rangle = 0$ ) and they are uncorrelated in time ( $\langle \xi_i(t)\xi_i(t') \rangle \propto \delta(t - t')$ ).

We linearize around the mean, where deviations from the mean are denoted by  $\delta$ :

$$m(t) = \bar{m} + \delta m(t) \quad , \quad (29)$$

$$p_a(t) = \bar{p}_a + \delta p_a(t) \quad , \quad (30)$$

$$p_2(t) = \bar{p}_2 + \delta p_2(t) \quad , \quad (31)$$

$$\delta p_1(t) = -\delta p_a(t) - \delta p_2(t) \quad . \quad (32)$$

After inserting the linearized equations into the Langevin approach we perform a Fourier transform:

$$-i\omega\delta\hat{m}(\omega) = r\delta\hat{p}_a - d\delta\hat{m} + \hat{\xi}_m \quad , \quad (33)$$

$$-i\omega\delta\hat{p}_a(\omega) = ak_+\delta\hat{p}_1 - k_-^a\delta\hat{p}_a + \hat{\xi}_a \quad , \quad (34)$$

$$-i\omega\delta\hat{p}_2(\omega) = bk_+\delta\hat{p}_1 - k_-^b\delta\hat{p}_2 + \hat{\xi}_2 \quad . \quad (35)$$

Starting with the equations for the occupancies, we rewrite Eqs (34,35) and use the approximation that  $d$  is significantly slower than the other rates to get:

$$\delta\hat{p}_a(\omega) = \frac{ak_+\delta\hat{p}_1 + \hat{\xi}_a}{k_-^a - i\omega} \approx A\delta\hat{p}_1 + \frac{\hat{\xi}_a}{k_-^a} \quad , \quad (36)$$

$$\delta\hat{p}_2(\omega) = \frac{bk_+\delta\hat{p}_1 + \hat{\xi}_2}{k_-^b - i\omega} \approx B\delta\hat{p}_1 + \frac{\hat{\xi}_2}{k_-^b} \quad , \quad \text{or} \quad (37)$$

$$\delta\hat{p}_a = -\delta\hat{p}_2 \frac{A}{(1+A)} + \frac{\hat{\xi}_a}{k_-^a} \frac{1}{(1+A)} \quad , \quad (38)$$

$$\delta\hat{p}_2 = -\delta\hat{p}_a \frac{B}{(1+B)} + \frac{\hat{\xi}_2}{k_-^b} \frac{1}{(1+B)} \quad . \quad (39)$$

Solving this system yields:

$$\delta\hat{p}_a = -\frac{\hat{\xi}_2}{k_-^b} \bar{p}_a + \frac{\hat{\xi}_a}{k_-^a} (\bar{p}_1 + \bar{p}_2) \quad . \quad (40)$$

The variances of the Langevin forces are:

$$\langle \hat{\xi}_a^* \hat{\xi}_a \rangle = 2k_-^a \bar{p}_a \quad , \quad (41)$$

$$\langle \hat{\xi}_2^* \hat{\xi}_2 \rangle = 2k_-^b \bar{p}_2 \quad , \quad (42)$$

$$\langle \hat{\xi}_m^* \hat{\xi}_m \rangle = 2d\bar{m} \quad , \quad (43)$$

and their covariances vanish, since the direct transition from state  $a$  to state 2 is not allowed. From Eqs (33,40) we get:

$$\langle \delta\hat{p}_a^* \delta\hat{p}_a \rangle = 2\frac{\bar{p}_2}{k_-^b} \bar{p}_a^2 + 2\frac{\bar{p}_a}{k_-^a} (\bar{p}_1 + \bar{p}_2)^2 \quad , \quad (44)$$

$$\langle \delta\hat{m}^* \delta\hat{m} \rangle = \frac{2d\bar{m}}{d^2 + \omega^2} + \frac{r^2}{d^2 + \omega^2} \langle \delta\hat{p}_a^* \delta\hat{p}_a \rangle \quad . \quad (45)$$

Finally, with  $\frac{1}{2\pi} \int_{-\infty}^{\infty} 2\frac{1}{x^2 + \omega^2} d\omega = \frac{1}{x}$  we get:

$$\begin{aligned} \sigma_m^2 &= \frac{d\bar{m}}{d} + \frac{r^2}{d} \left( \frac{p_2}{k_-^b} p_a^2 + \frac{p_a}{k_-^a} (p_1 + p_2)^2 \right) = \\ &= \bar{m} \left[ 1 + r \left( \frac{p_2}{k_-^b} p_a + \frac{1}{k_-^a} (1 - p_a)^2 \right) \right] \quad . \end{aligned} \quad (46)$$

This is one description of noise in the 3E architecture. To get the noise characteristics for modulation scheme 3E-1a, we need to express  $p_2$  in terms of  $p_a$  (not shown). From Eq (46) we can see that in the absence of repressors ( $p_2 = 0$ ) and also for very fast unbinding of the repressors ( $k_-^b \rightarrow \infty$ ) the noise shows the quadratic dependence on the occupation of the promoter that we see in the corresponding two-state model 2-1a.

### 2.3 Comparison to other methods

The results obtained with the Langevin approach were compared against two other methods: (i) the exact numerical solution of the chemical master equation and (ii) results from stochastic simulation using the Gillespie algorithm. Two kinds of comparisons are relevant: first, how well the gaussian distribution approximates the true distribution of mRNA levels; and second, how the Langevin-derived expressions for the noise characteristics compare to the exact values.

Fig S2A compares the distribution of mRNA levels obtained from the numerical solution of the chemical master equation to the gaussian approximation for the dual regulation architecture discussed in the last section.

The stochastic simulation algorithm is time consuming and offers no special benefit for the simple systems studied here, but we have nevertheless checked a few example architectures against simulation results. The results for dual regulation are shown in Fig S2B. Values for  $ak_+$  and  $k_-^a$  were chosen from a grid. This makes it possible to show the agreement with the Langevin-derived noise characteristics in two different modulation schemes (cf. inset in Fig S2B).

Another way to obtain analytical expressions for the mean and variance of the mRNA-distributions is the method of partial moments (e.g., (48, 52)). While this method can also be used to derive higher moments, a minor advantage of the Langevin method for the purposes here is that the approximation  $d \ll k_{ij}$  can be used earlier in the derivations, leading to simpler expressions.

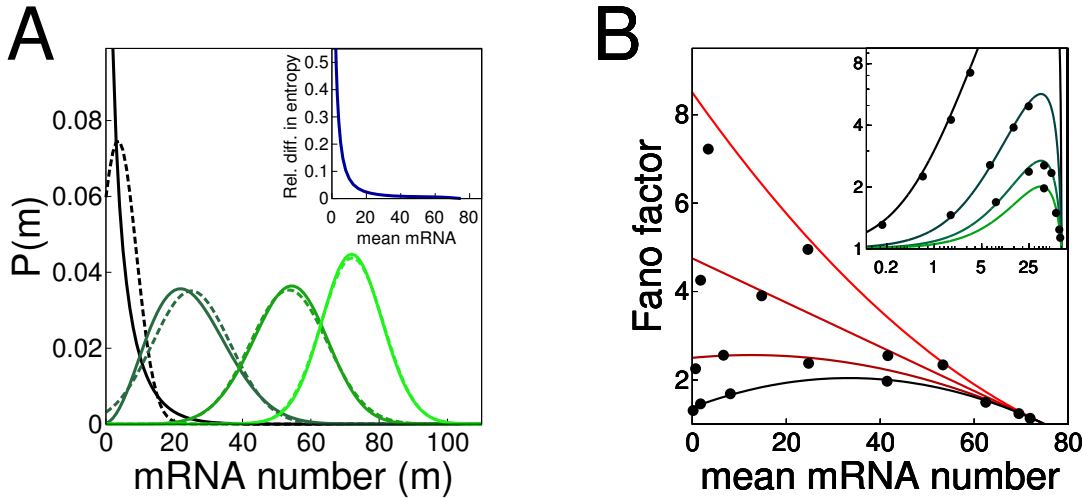


Figure S2: **Comparison of Langevin approach to other methods.** (A) Probability distribution of the model 3E for different values of  $ak_+$  (thus treating it as 3E-1a). Solid lines are the numerical solution of the master equation and dashed lines are a gaussian approximation using the analytical expressions for mean and variance described in the main text. Different colors are for different values of  $ak_+$ . Parameters for this plot:  $ak_+ = [1, 10, 50, 500]$ ,  $k_-^a = 10$ ,  $bk_+ = 10$ ,  $k_-^b = 10$ ,  $r = 75$  and  $d = 1$ . Inset: The relative difference in entropy between the full distribution and the gaussian approximation used in this study (i.e. using the results for mean and variance from the Langevin method). The error drops to  $< 5\%$  very fast, with the main difference at slow rates stemming mainly from non-gaussianity at low expression levels. Importantly, for information calculations this non-gaussianity is mitigated by the protein averaging; even without the averaging, the effect on the comparison between architectures is minor. (B) Comparison of Fano factors for different mean expression levels for the 3E-1a model (black data points = Gillespie simulation, solid lines = Langevin method). Different colors correspond to different values for  $k_-^a$ . For small values of  $k_-^a$  and  $ak_+$  we start to see deviations since the approximation  $d \ll k_{ij}$  no longer holds. Parameters for the Gillespie simulations:  $k_-^a = [10, 20, 50, 200]$ ,  $ak_+ = [1, 10, 50, 500]$ ,  $bk_+ = 10$ ,  $k_-^b = 10$ ,  $r = 75$  and  $d = 1$ . The Fano factor was calculated from 10000 runs. Error bars from resampling are smaller than the symbols. Inset: The same results from the Gillespie simulations replotted (on a log-log scale) and compared to the noise characteristics of the 3E-a1 scheme, i.e. the rate  $k_-^a$  is modulated to obtain different mean mRNA levels. The solid lines are the noise characteristics calculated with the Langevin methods for different (fixed) values of  $ak_+ = [1, 10, 50, 500]$ .



### 3 Magnitude of input noise

Our noise model includes the possibility of upstream noise propagating through a promoter. Two sources for this noise that is already present when the signal enters the system are considered here. The Poisson like fluctuations stemming from the production of the input molecules was already mentioned in the main text. It led to the model where input noise is propagated to the output variance in proportion to the “susceptibility” of the regulatory element:

$$\sigma_g^2 = \dots + v \left( \frac{\partial \bar{g}}{\partial k} \right)^2 k, \quad (47)$$

where (...) indicate output and switching terms (see main text).

The second, fundamental noise source at the input side is related to the fact that the regulatory proteins need to find their interaction partners by diffusion. This puts a lower bound on the variance of the *local concentration* at the regulatory site. This diffusion limit, first formulated for the case of bacterial chemotaxis by Berg and Purcell (66), has been subsequently derived for the general case of biochemical signaling (33, 34): the lower bound on the variance in local concentration obeys  $\sigma_c^2 \propto cd'/D\ell$ , where  $D$  is the diffusion constant of the TF molecules,  $\ell$  is the linear size of the binding site, and  $1/d'$  is the noise averaging time (here the lifetime of the gene product). Analyses of high-precision measurements in gene expression noise during early fruit fly development have shown that this diffusion noise represents a substantial contribution to the total (35, 45). Thus, also this biophysical limit set by diffusion yields a variance that is proportional to the input itself.

To see which values the constant  $v$  can take, note that  $\sigma_k^2 = k_+^2 \sigma_c^2 \propto k_+^2 cd'/D\ell = kk_+d'/D\ell$ . As an example, consider diffusion-limited association, where  $k_+ = 4\pi D\ell$  (67). Depending on the accuracy and the geometry of the sensing mechanism we now get different values for  $\tilde{v} = v/d'$ , but in general  $\tilde{v}$  is expected to be of order unity. For example, the perfect absorbing sphere has  $\sigma_c^2 = cd'/(4\pi D\ell)$  and therefore  $\tilde{v} = 1$ ; the perfect monitoring sphere in the Berg–Purcell limit has  $\sigma_c^2 = 3cd'/(5\pi D\ell)$  and therefore  $\tilde{v} = 2.4$  (66, 68).

## References

33. Bialek, W., and S. Setayeshgar, 2005. Physical limits to biochemical signaling. *Proc Natl Acad Sci USA* 102:10040–5.
34. Bialek, W., and S. Setayeshgar, 2008. Cooperativity, sensitivity, and noise in biochemical signaling. *Phys Rev Lett* 100:258101.
35. Gregor, T., D. Tank, E. Wieschaus, and W. Bialek, 2007. Probing the limits to positional information. *Cell* 130:153–64.
41. Kepler, T., and T. Elston, 2001. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys J* 81:3116–36.
44. So, L.-H., A. Ghosh, C. Zong, L. Sepulveda, R. Segev, and I. Golding, 2011. General properties of transcriptional time series in *Escherichia coli*. *Nat Genet* 43:554–60.
45. Tkačik, G., T. Gregor, and W. Bialek, 2008. The role of input noise in transcriptional regulation. *PLoS ONE* 3:e2774.
47. Raj, A., C. Peskin, D. Tranchina, D. Vargas, and S. Tyagi, 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* 4:e309.
48. Sanchez, A., and J. Kondev, 2008. Transcriptional control of noise in gene expression. *Proc Natl Acad Sci USA* 105:5081–6.
50. Karmakar, R., 2010. Conversion of graded to binary response in an activator-repressor system. *Phys Rev E* 81:021905.
52. Sanchez, A., H. Garcia, D. Jones, R. Phillips, and J. Kondev, 2011. Effect of promoter architecture on the cell-to-cell variability in gene expression. *PLoS Comput Biol* 7:e1001100.
59. Suter, D., N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef, 2011. Mammalian genes are transcribed with widely different bursting kinetics. *Science* 332:472–4.
66. Berg, H., and E. Purcell, 1977. Physics of chemoreception. *Biophys J* 20:193–219.
67. Berg, O. G., and P. H. von Hippel, 1985. Diffusion-controlled macromolecular interactions. *Annu Rev Biophys Biophys Chem* 14:131–58.
68. Endres, R. G., and N. S. Wingreen, 2008. Accuracy of direct gradient sensing by single cells. *Proc Natl Acad Sci USA* 105:15749–54.
69. Carey, L. B., D. van Dijk, P. M. Sloom, J. A. Kaandorp, and E. Segal, 2013. Promoter sequence determines the relationship between expression level and noise. *PLoS Biol* 11:e1001528.
71. Sanchez, A., S. Choubey, and J. Kondev, 2013. Regulation of noise in gene expression. *Annu Rev Biophys* 42:469–91.

72. Golding, I., J. Paulsson, S. Zawilski, and E. Cox, 2005. Real-time kinetics of gene activity in individual bacteria. *Cell* 123:1025–36.
73. Geertz, M., D. Shore, and S. J. Maerkl, 2012. Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proc Natl Acad Sci USA* 109:16540–5.
74. Vinuelas, J., G. Kaneko, A. Coulon, E. Vallin, V. Morin, C. Mejia-Pous, J.-J. Kupiec, G. Beslon, and O. Gandrillon, 2013. Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts. *BMC Biol* 11:15.
75. Sanchez, A., M. L. Osborne, L. J. Friedman, J. Kondev, and J. Gelles, 2011. Mechanism of transcriptional repression at a bacterial promoter by analysis of single molecules. *EMBO J* 30:3940–6.
76. Garcia, H., A. Sanchez, J. Boedicker, M. Osborne, J. Gelles, J. Kondev, and R. Phillips, 2012. Operator sequence alters gene expression independently of transcription factor occupancy in bacteria. *Cell Rep* 2:150–61.
77. Yang, H.-T., and M. S. H. Ko, 2012. Stochastic modeling for the expression of a gene regulated by competing transcription factors. *PLoS ONE* 7:e32376.
78. Earnest, T. M., E. Roberts, M. Assaf, K. Dahmen, and Z. Luthey-Schulten, 2013. DNA looping increases the range of bistability in a stochastic model of the lac genetic switch. *Phys Biol* 10:026002.
79. Mirny, L., 2010. Nucleosome-mediated cooperativity between transcription factors. *Proc Natl Acad Sci USA* 107:22534–9.
80. Segal, E., and J. Widom, 2009. What controls nucleosome positions. *Trends Genet* 25:335–42.