Noise and Information Transmission in Promoters with Multiple Internal States

Georg Rieckh* and Gasper 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 relationship 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
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 addi-

**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. 1A. 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. 1B).

When the noise is small, it is useful to think of a regulatory function as describing the average expression level, $\bar{g}(k)$, and of the noise as inducing some random fluctuation around that average. The variance of these fluctuations, $\sigma^2_g(k) = \int d\bar{g} g^2 P(g|k)$, of the 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) output 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:

$$I(k; g) = \int dk \int dg P(k)P(g|k) \log_2 \left( \frac{P(g|k)}{P(g)} \right).$$

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, $\sigma^2_g(k)$. The result (10–14) is as follows:

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

$$Z = \int_0^{\gamma_{\text{max}}} dk \frac{d\bar{g}}{\sigma_g(k)} = \int_{\gamma_{\text{min}}}^{\gamma_{\text{max}}} \frac{d\bar{g}}{\sigma_g(\bar{g})},$$

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 $\bar{g}(k)$. This integral is graphically depicted in Fig. 1C (inset). Finally, we will use this to explore the dependence of $I^*(k; g)$ on parameters that define the promoter architecture (see Fig. 1D), 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.
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 lowest 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 \( p_a \) be the fractional occupancy of state 1 and \( k_j \) 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 \( m \):

\[
\frac{\partial p}{\partial t} = Kp + \xi, \quad \frac{\partial m}{\partial t} = rp_a - dm + \xi_m, \quad \text{with} \quad (4)
\]

\[
K = \begin{pmatrix}
-(k_{a1} + k_{a2}) & k_{a1} & k_{2a}
k_{a1} & -(k_{a1} + k_{12}) & k_{21}
k_{a2} & k_{12} & -(k_{2a} + k_{21})
\end{pmatrix}, \quad (5)
\]

and \( p = (p_a, p_1, p_2)^T \); \( \xi = (\xi_a, \xi_1, \xi_2)^T \) 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, \( d \ll k_j \), we can write the noise in a generic way:

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

where \( p_{\text{act}} \) is the occupancy of the active states (\( p_a \) or \( p_a + p_1 \)), 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 (\( \sigma_m^2 = \mathbb{E} + \ldots \) ) 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 \( k = k_{a1} \). 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 \( c k_{c1} \). 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 \( k \), such that 

\[
\sigma_k^2 = \ldots + v \left( \frac{\partial \sigma_k}{\partial k} \right)^2 k, \tag{7}
\]

where \((\ldots)\) indicate output and switching terms from Eq. 6 and \( v \) is the proportionality constant \( \sigma_k^2 \) 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 average 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

![Figure 3](https://example.com/figure3.png)

**FIGURE 3** Mean expression level and noise for different promoter models. (A) Expressions for the mean (first row, \( m \)) and the variance (second row, \( \sigma_m^2 \)) of the mRNA distribution in steady state for different promoter architectures. In the limit of \( k_{11} \to 0 \) (respectively \( k_{22}, k_{33} \to 0 \)), 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(idle), D(ouble). (B) Noise characteristics (\( p_s \) on the x axis vs. the Fano factor, \( \sigma_m^2/m \), on y axis) for different promoter models. Here, in all models \( k = k_{11} \) 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 \( k_{11} \) 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, \( \text{pad}(k) \), 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.
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. 3B). Additionally, we can also plot the noise (shown in Fig. 3B as the Fano factor, \( \frac{\sigma^2}{\mu} \)) as a function of the mean expression, \( \mu \), to get the noise characteristic of every modulation scheme. These curves, shown in Fig. 3B, 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. 3B (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. 3C, 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 reported 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 \( \bar{z}(k) = (r'/d')\bar{m}(k) \). The output and promoter switching noise contributions are affected differently, so that the protein level noise can be written as (45):

\[
\sigma^2 = \bar{z}
\left[
1 + \frac{w}{d'} p_{\text{act}} \Delta'
\right] + \text{input noise},
\]

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-a1). Here the transition into the active state is modulated by TF concentration \( c \) via \( k = k_{1a} = k_c c \), 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:

\[
\sigma^2 = \bar{z}
\left(1 + w + \frac{r r'}{d a_1} (1 - p_a)^2 + \frac{v}{k_{1a}} r'\right)
\]

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

\[
Z = \int_{\sigma_{\text{max}}}^{\bar{z}_{\text{max}}} \frac{d\bar{z}}{\sigma(\bar{z})} = \sqrt{N_{\text{max}}} \int_{0}^{p_{\text{act}}^{\text{max}}} dp_a
\times p_a^{-1/2}
\left[1 + \frac{1}{k_-} (1 - p_a)^2 + \frac{v}{k_-} (1 - p_a)^3\right]^{-1/2}
\]

\[
= \sqrt{N_{\text{max}}} Z_0.
\]
Here, $N_{\text{max}} = (rr'\varepsilon)/[(dd')(1+w)]$ and $\tilde{k}_- = k_{\text{at}}/(d'N_{\text{max}})$ is the dimensionless combination of parameters related to the off-rate for the TF dissociation from the binding site. $N_{\text{max}}$ 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, $N_{\text{max}}$ corresponds to the maximal average number of mRNA synthesized during a protein lifetime: $N_{\text{max}} = r/d' w/(1+w) \approx r/d'$. With this choice of parameters, $N_{\text{max}}$ 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 $I^* = \log_2Z_0$. 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 $2$-$1$a given by Eq. 10? Given that the dynamic range of input (e.g., TF concentration) is limited (10–13), $k \in [0, k_{\text{max}}]$, and given a choice of $\nu$ 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, $k \in [0, k_{\text{max}}]$, the information-maximizing solution chooses $k^*$ that balances the strength of binding (such that the dynamic range of expression is large), while simultaneously 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

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

and $p = (p_b,p_1,p_2)^T$, respectively $p = (p_b,p_0,p_1)^T$. 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: $k_{b1} = 2\gamma^{-1}k_-$, $k_{1a} = \rho k$, $k_{12} = k_-$, $k_{21} = 2k$.

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 4** Finding $k_-$ that optimizes information transmission in a two-state promoter. The strength of the input noise is fixed at $\nu = 2$ and the input dynamic range for $k$ is from 0 to $k_{\text{max}} = 10$. (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 $k \in [0,k_{\text{max}}]$. To see this figure in color, go online.
The occupancy of state 1 becomes negligible, and the regulation function becomes ever steeper, as is clear from Fig. 5E, 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. 5D. 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. 5E and F show the information of the optimal solutions as a function of the input noise magnitude as well as the input range, $k_{\text{max}}$. As expected, the information increases as a function of $k_{\text{max}}$ since the influence of input and switching noise can be made smaller with more input molecules. This increase saturates at high $k_{\text{max}}$ because output noise becomes limiting to the information transmission—this is why the capacity curves converge to the same maximum, the $v = 0$ 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 $k_{\text{max}}$ and $v$. It is interesting to note that increasing $k_{\text{max}}$ and decreasing $v$ 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. 6A shows a possible dual-role signaling scheme where the total concentration of dual-role TFs is fixed (at $k_{\text{max}}$), but the signal modulates the fraction of these TFs that play the activator role $(k)$ and the remaining fraction that act as repressors $(k_{\text{max}} - k)$, which compete for the same binding site. The channel capacity of this motif is depicted in Fig. 6B as a function of promoter parameters $k_R$ and $k_A$, showing that a globally optimal setting (denoted “A”) exists for these parameters; with these parameters, the input/output function, shown in Fig. 6C, is much steeper than what could be achieved with the best two-state promoter, and that is
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 $k_{\text{max}}$. This is how the dual-role regulation can escape the tradeoff faced by the two-state model 2-1a (shown in Fig. 4). 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 $k_{\text{max}}$. To see this figure in color, go online.

**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 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 $k_{\text{max}}$. At large $k_{\text{max}}$ 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.

**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 $k_{\text{max}}$, we look for the optimal choice of $k_{\text{off}} = k_{\text{on}}$. 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 $k_{\text{max}}$. To see this figure in color, go online.
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.

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

REFERENCES