Supplemental Information Appendix

Nucleosomal Proofreading of Activator-Promoter Interactions

Robert Shelansky and Hinrich Boeger

General Approach

We represent hypotheses of stochastic molecular dynamics in terms of 'directed graphs' (1-3). A graph consists of nodes and connecting directed edges. The nodes correspond to distinct states of a single promoter molecule, the directed edges to possible transitions between them. The stochastic dynamics of the system (promoter molecule) are modeled as flow of probability mass across the graph ('stochastic process').

We obtained mathematical expressions for the flux (rate of flow) of probability mass across the graph on the Markov assumption — *i.e*., the future evolution depends on the present state of the promoter alone, and not its prior history (4). (This lack-of-memory property is shared between deterministic and Markovian stochastic processes.)

We focus on stationary processes (‘steady state dynamics’), where promoter state probabilities are time-independent: for every node the flux of probability mass into and out of the node is equal. (Only in steady state are state probabilities constrained by and a reflection of the biochemical assumptions, *viz.* graph topology and transition kinetics.) A

steady state or 'stationary' distribution of probabilities always exists (see below), and is uniquely defined when the graph is strongly connected, *i.e.*, when every node can be reached from any other node *via* a string of directed edges (1). All models discussed in the sequel are represented by strongly connected graphs.

Promoter State Transitions

The stochastic promoter state dynamics are defined by the transition functions $p_{ji}(t+h,t)$, indicating the *conditional* probabilities that the promoter will be in state *j* at time $t+h$, given it is in state i at time t. We assume that the transition functions depend on the promoter's present state, i, but not its past — and on h, the distance between both time points, and not time itself; *i.e.*,

$$p_{ji}(t+h,t) = p_{ji}(h,0) \equiv p_{ji}(h)$$

(assumption of a time-homogeneous Markov process).

The perhaps deepest and far reaching implication of the Markov assumption is that it mathematically entails the second law of thermodynamics (5): the existence of a function (total entropy) whose value only increases, *viz*. during processes that violate detailed balance and therefore are irreversible, or else remains constant. Detailed balance, thus, corresponds to thermodynamic equilibrium, statistical irreversibility to entropy production (see below).

The promoter may transition from its initial state i at time 0 to state *j* at time $t+h$ *via* state k at time t. Summing over all k gives

$$p_{ji}(t+h) = \sum_k p_{jk}(h)p_{ki}(t)$$

(Chapman-Kolmogorov equation). With $P_t = \left(p_{ji}(t)\right)$ the square matrix of transition functions, the Chapman-Kolmogorov equation may be written in matrix form:

$$P_{t+h} = P_h P_t \qquad [3].$$

Although usually introduced as an independent assumption, the time-homogeneous Markov assumption implies continuous differentiability of the transition functions (6). That is, the equation

$$\frac{P_{t+h} - P_t}{h} = \frac{(P_h - P_0)P_t}{h}$$

(note that P_0 equals the identity matrix, which follows from [3] with $h = 0$) tends to

$$dP_t/dt = \mathrm{W}P_t \qquad [4]$$

for $h \to 0$, where $\mathrm{W} = dP_0/dt$. W is called the 'generator' of the process. We may refer to its elements

$$w_{ji} = \lim_{h \to 0} \frac{p_{ji}(h) - p_{ji}(0)}{h}$$

as 'transition rates' (although 'transition rate constants' would be more accurate). The diagonal elements of W are given by

$$w_{ii} = -\sum_{j \neq i} w_{ji} \qquad [5].$$

Equation 5 is a consequence of probability mass conservation (see below).

We now show that differentiability of the transition functions implies that the flux of probability mass from state i to state j is a linear function of $p_i(t)$, the probability of promoter state $j \in \{1, \ldots, n\}$ at time t.

Let $\boldsymbol{p}(t) = \big(p_1(t), \ldots, p_n(t)\big)^T$, where the superscript *T* indicates that $\boldsymbol{p}(t)$ is a column and not row vector. From the definition of the transition functions follows,

$$\boldsymbol{p}(t) = P_t \boldsymbol{p}(0) \, .$$

Differentiation and insertion of [4] into the results yields

$$\frac{d\boldsymbol{p}(t)}{dt} = \mathrm{W}\boldsymbol{p}(t) \, .$$

This set of coupled linear differential equations is called 'the master equation' of the process.

With $\mathbf{1} = (1, \ldots, 1)$, the *row* vector whose components are all 1, conservation of probability mass may be expressed as $\mathbf{1}\boldsymbol{p}(t) = 1$ for all t. Differentiation of $\mathbf{1}\boldsymbol{p}(t) = 1$, and insertion of the master equation into the result yields $\mathbf{1}\mathrm{W}\boldsymbol{p}(t) = 0$. This is true for any $\mathbf{1}\boldsymbol{p}(t) = 1$. With $\boldsymbol{p}(t) = (1,0, \ldots, 0)^T$, $\boldsymbol{p}(t) = (0,1, \ldots, 0)^T$ *etc*. now follows equation [5].

In steady state, the master equation becomes

$$\mathrm{W}\boldsymbol{\pi} = \mathbf{0} \qquad [6],$$

where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)^T$ is the vector of steady state (stationary) probabilities π_j of promoter states j, and $\mathbf{0} = (0, \ldots, 0)^T$ is the zero column vector. Equation [5] implies that

W is singular. Therefore, a non-trivial solution to equation [6], $\boldsymbol{\pi} \neq \mathbf{0}$, always exists. It is uniquely defined, provided the graph is strongly connected (7). Furthermore, $\boldsymbol{p}(t) \rightarrow \boldsymbol{\pi}$ for $t \rightarrow \infty$ (8).

Equation [6] shows that $\boldsymbol{\pi}$ is the basis vector with unit length (*i.e.*, $\mathbf{1}\boldsymbol{\pi} = 1$) of the kernel of W. We used Mathematica to obtain $\boldsymbol{\pi}$ and calculate activator fidelities, f. Alternatively, $\boldsymbol{\pi}$ may be calculated using the Matrix-Tree theorem (1, 7).

Model 1

The generator of Model 1 for correct promoter binding is

$$\mathrm{W}_1 = \begin{pmatrix} -\kappa & k_C \\ \kappa & -k_C \end{pmatrix}.$$

(The generator for incorrect binding is obtained by replacing k_C with k_i.)

$$\boldsymbol{\pi} = \frac{1}{(\kappa + k_C)} \begin{pmatrix} k_C \\ \kappa \end{pmatrix},$$

which is easily verified. The probability of the transcriptionally active state is

$$\pi_{ON} = \pi_2 = \kappa/(\kappa + k_C) \qquad [7].$$

A corresponding equation holds for incorrect promoter binding. On the assumption that the rate of transcription linearly depends on π_{ON}, we obtain for the activator fidelity

$$f_1 = \frac{\kappa + k_i}{\kappa + k_C}.$$

Model 2

The generator for Model 2 is

$$\mathrm{W}_2 = \begin{pmatrix} -(\kappa + \lambda) & k_C & 0 & \beta \\ \kappa & -(k_C + \alpha) & \beta & 0 \\ 0 & \alpha & -(k_C + \beta) & \kappa \\ \lambda & k_C & k_C & -(\beta + \kappa) \end{pmatrix}.$$

From the kernel vectors of W_2 for correct and incorrect DNA binding we obtain

$$f_2 = f_1 \frac{k_i(\beta + \lambda) + M(\beta + \kappa + \lambda)}{k_C(\beta + \lambda) + M(\beta + \kappa + \lambda)} \frac{k_C\lambda + \alpha(\beta + \kappa + \lambda)}{k_i\lambda + \alpha(\beta + \kappa + \lambda)},$$

where $M = \alpha + \beta$, α and λ are the rate constants for nucleosome removal in the presence and absence of the activator, respectively, and β is the rate constant for nucleosome

removal. Since (by definition) $k_i > k_C$, f_2 increases for decreasing λ, M, and κ. When all three tend to zero

$$f_2 \to f_0^2 .$$

In contrast, for $M \to \infty$,

$$f_2 \to f_1 .$$

From the kernel vector of W_2 with $\alpha = \lambda$, we obtain for $\pi_{ON} = \pi_3$

$$\pi_{ON} = (\alpha/M)(\kappa/(\kappa + k_C) \qquad [10].$$

The analogous equation holds for incorrect DNA binding ($k_C \to k_i$), from which it follows that $f_2 = f_1$.

Model 3

For Model 3 (activation by nucleosome removal; Fig. 3A), the probability of the active state is $\pi_{ON} = \pi_3 + \pi_4$. In the following, we assume $\lambda = 0$, which simplifies the algebra and increases activator fidelity, f_3. From the generator of Model 3, which is the same as for Model 2, we obtain

$$f_3 = f_1 \frac{L + Qk_i}{L + k_i} \frac{L + k_C}{L + Qk_C} \qquad [11],$$

where $L = \kappa + \beta$ and $Q = \beta/M$. By contradiction it may be proved that

$$f_3 \leq f_1 .$$

Suppose there exist L, k_i, and $Q \leq 1$, such that $f_3/f_1 > 1$. With equation [11] it then follows that

$$(L + Qk_i)(L + k_C) > (L + Qk_C)(L + k_i).$$

Multiplying out the brackets, subtracting equal terms from both sides of the inequality and dividing by L yields $k_C + Qk_i > k_i + Qk_C$ and thus $Q(k_i - k_C) > k_i - k_C$. Hence $Q > 1$ for $k_i \neq k_C$, in contradiction to $Q \leq 1$. Thus $f_3 \leq f_1$ for all L and all k_i.

Model 4

With rate constant as indicated in Fig. 3D, the generator of Model 4 is

$$\mathrm{W}_4 = \begin{pmatrix} -\kappa & k_C & 0 & \beta & 0 & 0 \\ \kappa & -(k_C+\alpha) & \beta & 0 & 0 & 0 \\ 0 & \alpha & -(k_C+\beta+\zeta) & \kappa & z & 0 \\ \lambda & 0 & k_C & -(\beta+\kappa) & 0 & z \\ 0 & 0 & \zeta & 0 & -(k_C+z) & \kappa \\ 0 & 0 & 0 & \eta & k_C & -(z+\kappa) \end{pmatrix}.$$

To simplify the algebra, we assume $\lambda, \eta = 0$ (the rates of nucleosome removal and TBP binding in the activator-unbound promoter state), which increases activator fidelity, f_4. We thus obtain from the kernel of W_4:

$$f_4 = f_1 \frac{k_i^2 Q + \big(1+(1-Q)R\big)(z+\kappa)L + k_i\big(L+Q(N+\kappa)\big)}{k_C^2 Q + \big(1+(1-Q)R\big)(z+\kappa)L + k_C\big(L+Q(N+\kappa)\big)},$$

where $R = \zeta/z$ and $N = \zeta + z$, and, as above, $L = \beta + \kappa$ and $Q = \beta/M$.

Maximal fidelity is reached as the activator on-rate, κ, the nucleosome kinetics, M, and TBP-DNA binding kinetic, N, become infinitely slow; *i.e.*, for $M, N, \kappa \to 0$ and constant Q and R,

$$f_4 \to \frac{k_i^3}{k_C^3} = f_0^3 \qquad [12].$$

Detailed Balance and Reversibility

A stochastic process is said to be in '*detailed balance'* if and only if

$$w_{ji} p_i(t) = w_{ij} p_j(t),$$

for all i, j, and all t. That is, the flux of probability mass between any two nodes is balanced.

Detailed balance implies steady state (stationarity), $d\boldsymbol{p}(t)/dt = \mathbf{0}$. This is easily proved by observing that

$$\frac{dp_j(t)}{dt} = \sum_i w_{ji} p_i(t) - \left(\sum_i w_{ij}\right) p_j(t) = \sum_i \big(w_{ji} p_i(t) - w_{ij} p_j(t)\big) = 0\,,$$

where the first equality follows from the master equation and the generator property

$\sum_i w_{ij} = 0$, and the last equality from detailed balance. The reverse, however, is not true: steady state does not guarantee detailed balance (*i.e.*, equilibrium) — *e.g.* Model 2 (Fig. 1B) with $\alpha > \lambda$.

In detailed balance, there is no way to determine the direction of time, for there is no net flux of probability mass between any two nodes. Both forward and backward direction of any sequence of events are stochastically indistinguishable (8). A process in detailed balance is therefore called '*reversible*'; a process that violates detailed balance is called '*irreversible*'.

Criterion for Detailed Balance

Whether a stationary process is in detailed balance may be inferred from its generator; for the following can be proved (8):

A stationary process on a directed graph is in detailed balance if and only if for any closed loop of transitions, multiplication of the rate constants (generator elements, edge labels) going around the loop yields the same product regardless of direction (clockwise and counterclockwise). This is called Kolmogorov's criterion or 'cycle condition' for detailed balance (1, 8).

The probability of a clockwise transition cycle in Model 2 (Fig. 2A) is given by

$$\frac{\kappa}{(\lambda+\kappa)}\frac{\alpha}{(k_C+\alpha)}\frac{k_C}{(\beta+k_C)}\frac{\beta}{(\kappa+\beta)},$$

and of an anticlockwise cycle by

$$\frac{\lambda}{(\lambda+\kappa)}\frac{\alpha}{(\kappa+\beta)}\frac{\beta}{(\beta+k_C)}\frac{k_C}{(k_C+\alpha)}.$$

The two probabilities are equal if and only if $\alpha = \lambda$. Detailed balance follows with Kolmogorov's criterion.

We note that stationary processes on 'trees' — graphs without closed loops — are necessarily in detailed balance; for the absence of loops trivially ensures that Kolmogorov's criterion is fulfilled.

Entropy Production

By introducing the master equation for Markov processes (see above) into the time-derivative of Gibbs' entropy function, $S = -k_B \sum_j p_j ln(p_j)$, it is found that dS/dt may

be decomposed into the sum of two contributions: entropy production, d_iS/dt, due to irreversibility, and entropy flow, d_eS/dt, due to heat exchange with the environment (5, 9). For Model 2, entropy production is given by

$$\frac{d_iS(t)}{dt} = (\kappa\pi_1 - k_C\pi_2)k_B ln\left(\frac{\kappa\pi_1}{k_C\pi_2}\right) + (\alpha\pi_2 - \beta\pi_3)k_B ln\left(\frac{\alpha\pi_2}{\beta\pi_3}\right)$$
$$+(k_C\pi_3 - \kappa\pi_4)k_B ln\left(\frac{k_C\pi_3}{\kappa\pi_4}\right) + (\beta\pi_4 - \lambda\pi_1)k_B ln\left(\frac{\beta\pi_4}{\lambda\pi_1}\right).$$

where $\pi_1, \ldots, \pi_4$ are the stationary (steady state) probabilities of promoter states $1, \ldots, 4$, respectively; k_B is the Boltzmann constant. All terms on the right side are equal to zero if the system is in detailed balance; otherwise $d_iS/dt > 0$. (The subscript *i*, here, refers to the change in entropy due to irreversibility — *i.e.*, entropy production — and should not be confused with *i* for incorrect DNA binding, above.)

Fano Factor and Transcriptional Bursting

As a measure of transcription noise, we use the *Fano factor*, $\mathbb{F}_R$, *i.e.*, the variance of mRNA abundance divided by its mean. The assumption of a time-homogeneous process implies that calculated noise is intrinsic, rather than total, noise.

The nodes of the graph, now, represent pairs of numbers (j, m), where j indicates the promoter state and *m* the number of mRNA molecules (3, 10). The latter, m, runs from 0 to infinity; the number of nodes is countably infinite, and the master equation encompasses infinitely many equations. A simple tool, the probability generating function (11), allows us to reduce this infinite set of differential equations to a finite set amenable to linear algebraic operations (3, 10), as detailed below.

At any given moment in time, one of the following changes may occur: (*a*) the promoter 'jumps' into another state, $(i, m) \rightarrow (j, m)$, with transition probability per unit time w_{ji}; (*b*) an mRNA molecule is produced, $(i, m) \rightarrow (j, m + 1)$, with transition probability per unit time μ_i for promoter state *i*; or (*c*) an mRNA molecule is degraded, $(i, m) \rightarrow (j, m - 1)$, with transition probability per time $m\delta$, where δ is a constant.

Let $\boldsymbol{p_m}(t) = \big(p(1, m, t), \ldots, p(N, m, t)\big)^T$, where N is the total number of promoter states and $p(j, m, t)$ is the probability of state (j, m) at time t; $E = \big(\delta_{ij}\mu_i\big)$, and $I = \big(\delta_{ij}\big)$, where δ_{ij} is the Kronecker symbol (*i.e.*, $\delta_{ij} = 1$ for $i = j$, and $\delta_{ij} = 0$ for $i \neq j$).

Both E and I are square matrices. For instance, for the two-state promoter model of Fig. 1A,

$$E = \begin{pmatrix} 0 & 0 \\ 0 & \mu \end{pmatrix} \quad \text{and} \qquad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Furthermore, let L_{-1} and L_{+1} be the 'step operators' with $L_{-1}\boldsymbol{p}_m(t) = \boldsymbol{p}_{m-1}(t)$ and $L_{+1}\boldsymbol{p}_m(t) = \boldsymbol{p}_{m+1}(t)$. The master equation of the process, then, is

$$\frac{d\boldsymbol{p}_m}{dt}(t) = \left[\mathrm{W} - m\delta I - E + EL_{-1} + (m+1)\delta L_{+1}\right]\boldsymbol{p}_m(t),$$

where W, as above, is the generator for the promoter state transitions (12). (The index m runs from 0 to infinity.)

In steady state, $d\boldsymbol{p_m}(t)/dt = \boldsymbol{0}$, and the master equation, after some rearrangements, becomes

$$\left[m\delta I - \mathrm{W} + E\right]\boldsymbol{\pi}_m = \delta(m+1)\boldsymbol{\pi}_{m+1} + E\boldsymbol{\pi}_{m-1} \qquad [14],$$

where $\boldsymbol{\pi_m}$ is

$$\boldsymbol{\pi}_m = \left(\pi(1,m),...,\pi(N,m)\right)^T, \quad \text{with } \sum_m \boldsymbol{\pi}_m = \boldsymbol{\pi}.$$

In the following we derive the first two moments of the probability distribution for the stochastic mRNA abundance (stochastic variable R), which are required to calculate the noise of mRNA expression, using vector-valued generating functions (10, 13). The approach has been to develop to calculate all moments (14); and there are also methods for obtaining the entire distribution (15).

Let $\boldsymbol{G}(x) \equiv \sum_m x^m \boldsymbol{\pi}_m$, where $x \in [0,1]$. $\boldsymbol{G}$ is called the 'generating function' of the stationary distribution $\{\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, ...\}$ (3, 10, 11). Multiplying both sides of the master equation [4] for steady with x^m and summing over all m gives

$$\delta x\frac{d\boldsymbol{G}}{dx}(x) - \mathrm{W}\boldsymbol{G}(x) + E\boldsymbol{G}(x) = \delta\frac{d\boldsymbol{G}}{dx}(x) + Ex\boldsymbol{G}(x),$$

and, after some rearrangements,

$$\left[(1-x)E - \mathrm{W}\right]\boldsymbol{G}(x) = \delta(1-x)\frac{d\boldsymbol{G}}{dx}(x) \qquad [15].$$

With $x = 1$, we obtain

$$\mathrm{W}\boldsymbol{G}(1) = \mathbf{0};$$

thus, $\boldsymbol{G}(1) = \boldsymbol{\pi}$ (which also directly follows from the definition of $\boldsymbol{G}$). Differentiation of equation 15 yields

$$-E\boldsymbol{G}(x) + \left[(1-x)E - \mathrm{W}\right]\frac{d\boldsymbol{G}}{dx}(x) = -\delta\frac{d\boldsymbol{G}}{dx}(x) + \delta(1-x)\frac{d^2\boldsymbol{G}}{dx^2}(x) \qquad [16].$$

For $x = 1$, we obtain

$$(\delta I - \mathrm{W})\frac{d\boldsymbol{G}}{dx}(1) = E\boldsymbol{\pi},$$

and with $\boldsymbol{r} \equiv d\boldsymbol{G}(1)/dx$,

$$(\delta I - \mathrm{W})\boldsymbol{r} = E\boldsymbol{\pi} \qquad [17].$$

Note that the components of $\boldsymbol{r}$, are

$$r_j = \sum\nolimits_m m\pi(j,m).$$

Thus, r_j is the average number of mRNAs in steady state, given the promoter is in state j. Summing over all j, provides the expectation, $\mathbb{E}(R)$, of the random variable R: number of mRNAs. That is, the average number of transcripts across a population of (haploid) cells is

$$\mathbb{E}(R) = \mathbf{1}\boldsymbol{r}.$$

It can be shown that for all real eigenvalues γ of W, $\gamma \leq 0$ (7). Thus, since $\delta > 0$ by definition, there is no $\boldsymbol{u} \neq \mathbf{0}$ with $(\delta I - \mathrm{W})\boldsymbol{u} = \mathbf{0}$. Therefore, $(\delta I - \mathrm{W})$ is regular (one-to-one) with inverse $(\delta I - \mathrm{W})^{-1}$, and $\boldsymbol{r}$ is uniquely determined. From equation 17 we thus obtain

$$\boldsymbol{r} = (\delta I - \mathrm{W})^{-1}E\boldsymbol{\pi} \qquad [18].$$

Differentiation of equation (16) yields

$$-2E\frac{d\boldsymbol{G}}{dx}(x) + \left[(1-x)E - \mathrm{W}\right]\frac{d^2\boldsymbol{G}}{dx^2}(x) = -2\delta\frac{d^2\boldsymbol{G}}{dx^2}(x) + \delta(1-x)\frac{d^3\boldsymbol{G}}{dx^3}(x).$$

For $x = 1$, and with $\boldsymbol{v} \equiv d^2\boldsymbol{G}(1)/dx^2$, the last equation becomes

$$(2\delta I - \mathrm{W})\boldsymbol{v} = 2E\boldsymbol{r}.$$

Again, $\boldsymbol{v}$ is uniquely determined, for $(2\delta I - \mathrm{W})$ is regular:

$$\boldsymbol{v} = (2\delta I - \Gamma)^{-1}(2E\boldsymbol{r}) \qquad [19].$$

The components of $\boldsymbol{v}$, are

$$v_j = \sum_m m(m-1)\pi(j,m),$$

and therefore,

$$\mathbf{1}\boldsymbol{v} = \mathbb{E}\big(R(R-1)\big) = \mathbb{E}(R^2) - \mathbf{1}\boldsymbol{r}.$$

Since, $var(R)$, the variance of R, is

$$var(R) = \mathbb{E}(R^2) - (\mathbf{1}\boldsymbol{r})^2$$

we obtain

$$var(R) = \mathbf{1}\boldsymbol{v} + \mathbf{1}\boldsymbol{r} - (\mathbf{1}\boldsymbol{r})^2 \quad .$$

The Fano factor of RNA expression, therefore, is given by

$$\mathbb{F}_R = \frac{\mathbf{1}\boldsymbol{v} + \mathbf{1}\boldsymbol{r} - (\mathbf{1}\boldsymbol{r})^2}{\mathbf{1}\boldsymbol{r}}.$$

Solutions for $\boldsymbol{r}$ and $\boldsymbol{v}$ (equations [18] and [19]), and thus $\mathbb{F}_R$, were calculated using Mathematica.

For Model 1 (Fig. 1A) we thus obtain

$$\mathbb{F}_R = 1 + \frac{\mu k_C}{(\kappa + k_C)(\kappa + k_C + \delta)},$$

where δ is the rate constant for mRNA degradation. (It is often claimed that the Fano factor equals burst size, μ/k_C. The equation above shows that this is incorrect. Only for $\kappa, \delta \ll k_C \ll \mu$, does the Fano factor approximately equal burst size.) The second term captures the deviation from Poissonian expectation. When $\pi_{ON} = 1$, that is $k_C = 0$, the gene is conducive to transcription at all times, and therefore not transcribed in bursts. Bursting becomes increasingly manifest, and $\mathbb{F}_R$ increases, with increasing μ and decreasing κ — as pauses between transcriptional bursts become longer and pauses between transcription events within bursts become shorter.

The last equation may also be written as

$$\mathbb{F}_R(\kappa) = 1 + \frac{[1-\pi_{ON}(\kappa)]\nu_C}{(\kappa + k_C + \delta)\pi_{ON}(\kappa)}.$$

The graph of this function is plotted in Fig. 1D. For $\kappa \to 0$, when $f_1 \to f_0$ and $\pi_{ON} \to 0$, $\mathbb{F}_R \to \infty$ for constant ν_C.

Stochastic simulations of sample paths were calculated in Python.

Supplemental Figures

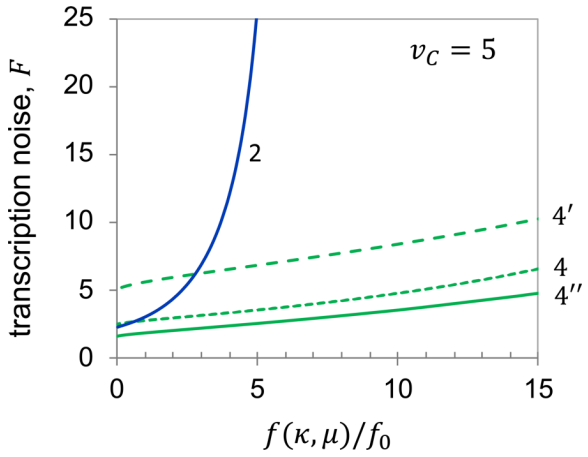

Fig. 1S. *Additional proofreading steps further attenuate noise*. Transcription noise as a function of relative activator fidelity, $f(\kappa,\mu)/f_0$, for Model 4 (green) with different parameter values for TBP binding $(4, 4', 4'')$ and Model 2 (blue, 2). All parameter values were as indicated in the legends to Figs. 2 and 3, except $\zeta = 2$, $z = 2$ $(4')$; $\zeta = 2$, $z = 10$ $(4'')$. As in Fig. 3, $\zeta = 10$, $z = 10$ (4).

References

1. Gunawardena J (2012) A linear framework for time-scale separation in nonlinear biochemical systems. *PLoS One* 7(5):e36321.
2. Boeger H (2014) Nucleosomes, transcription, and probability. *Molecular biology of the cell* 25(22):3451-3455.
3. Boeger H, Griesenbeck J, & Kornberg RD (2008) Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription. *Cell* 133(4):716-726.

4. Grimmet G & Stirzaker D (2001) *Probability and random processes* (Oxford university press) third Ed.
5. Van den Broeck C (2013) Stochastic thermodynamics: A brief introduction. Proceedings of the International School of Physics "Enrico Fermi"(Physics of Complex Colloids):155-193.
6. Cinlar E (2013) *Introduction to stochastic processes* (Dover Publications, Mineola, New York).
7. Mirzaev I & Gunawardena J (2013) Laplacian dynamics on general graphs. *Bulletin of mathematical biology* 75(11):2118-2149.
8. Kelly FP (1979) *Reversibility and stochastic networks* (John Wiley and Sons Ltd).
9. Wong F, Amir A, & Gunawardena J (2018) Energy-speed-accuracy relation in complex networks for biological discrimination. *Phys Rev E* 98(1-1):012420.
10. Mao C*, et al.* (2010) Quantitative analysis of the transcription control mechanism. *Mol Syst Biol* 6:431.
11. Feller W (1971) *An Introduction to Probability Theory* (John Wiley & Sons).
12. Brown CR, Mao C, Falkovskaia E, Jurica MS, & Boeger H (2013) Linking stochastic fluctuations in chromatin structure and gene expression. *PLoS Biol* 11(8):e1001621.
13. Boeger H, Griesenbeck J, Strattan JS, & Kornberg RD (2003) Nucleosomes unfold completely at a transcriptionally active promoter. *Molecular cell* 11(6):1587-1598.
14. Zhang J, Nie Q, He M, & Zhou T (2013) An effective method for computing the noise in biochemical networks. *The Journal of chemical physics* 138(8):084106.
15. Mugler A, Walczak AM, & Wiggins CH (2009) Spectral solutions to stochastic models of gene expression with bursts and regulation. *Physical review. E, Statistical, nonlinear, and soft matter physics* 80(4 Pt 1):041921.