Biophysical Journal
Article

Biophysical Society

Distribution of Initiation Times Reveals Mechanisms of Transcriptional Regulation in Single Cells

Sandeep Choubey,[1] Jane Kondev,[1] and Alvaro Sanchez[2,3,*]
[1]Department of Physics, Brandeis University, Waltham, Massachusetts; [2]Rowland Institute at Harvard, Harvard University, Cambridge, Massachusetts; and [3]Department of Ecology and Evolutionary Biology, Microbial Sciences Institute, Yale University, New Haven, Connecticut

ABSTRACT Transcription is the dominant point of control of gene expression. Biochemical studies have revealed key molecular components of transcription and their interactions, but the dynamics of transcription initiation in cells is still poorly understood. This state of affairs is being remedied with experiments that observe transcriptional dynamics in single cells using fluorescent reporters. Quantitative information about transcription initiation dynamics can also be extracted from experiments that use electron micrographs of RNA polymerases caught in the act of transcribing a gene (Miller spreads). Inspired by these data, we analyze a general stochastic model of transcription initiation and elongation and compute the distribution of transcription initiation times. We show that different mechanisms of initiation leave distinct signatures in the distribution of initiation times that can be compared to experiments. We analyze published data from micrographs of RNA polymerases transcribing ribosomal RNA genes in *Escherichia coli* and compare the observed distributions of interpolymerase distances with the predictions from previously hypothesized mechanisms for the regulation of these genes. Our analysis demonstrates the potential of measuring the distribution of time intervals between initiation events as a probe for dissecting mechanisms of transcription initiation in live cells.

INTRODUCTION

One of the key findings of the genomic era is the unexpectedly high similarity between the genomes of different organisms (1). As the number of genomes being sequenced is increasing, it is becoming clear that the biggest difference among organisms is not to be found in their protein-coding sequences, but in the ways in which their genes are regulated (2–5). This is putting the spotlight on the parts of the genome that are responsible for gene regulation and prompting the question: how do changes in regulatory sequences alter the way in which cells respond to intra- and extracellular signals (6)?

More often than not, genetic regulation occurs at the level of transcription, by which cells control the amount of messenger RNA of each gene they express (7). Regulation of transcription is commonly achieved by the integration of multiple intracellular signals at the regions of DNA upstream from and in proximity to the gene's coding region. This "promoter region" consists of a collection of transcription factor binding sites and, in eukaryotes, nucleosome positioning sites. Together, these binding sites dictate the binding and unbinding of specific transcription factors, cofactors, and chromatin remodeling factors, which, in turn, either promote or inhibit the assembly of the transcriptional machinery at the gene.

The collection of transcription factor binding sites (which includes enhancer regions in eukaryotes), their position, and affinity for transcription factor proteins constitutes the promoter architecture. Considerable effort has been directed to elucidating how promoter architecture determines measurable quantities like the average transcriptional response (8–11) of cells to a given stimulus, as well as the population-wide fluctuations of that response (12–15). Although we have witnessed considerable progress on this front in recent years, many mysteries remain even in simple organisms such as bacteria. In particular, how promoter architecture affects the dynamics of transcription initiation in single cells remains poorly understood. To answer this question, experiments are being done in which the number of RNA molecules from a gene of interest is measured at a single-cell level in a population of isogenic cells (12,16–18). The measured distribution of mRNA numbers in the cell population can then be used to test the predictions of different models of transcription initiation in the hope that some of these are supported by the data (19–30). This approach has led to the discovery of bursty mechanisms of transcription initiation (12,16). However, this method of inferring the

Submitted November 27, 2017, and accepted for publication March 29, 2018.

*Correspondence: alvaro.sanchez@yale.edu

Editor: Anatoly Kolomeisky.
https://doi.org/10.1016/j.bpj.2018.03.031

kinetics of transcription is limited by the fact that the mRNA copy number reflects additional processes downstream of transcription, such as the nonlinear degradation of mRNA and proteins (31), maturation time of fluorescent reporters (32), mRNA transport (33), mRNA splicing (34), and small RNA regulation (35). The stochastic nature of these processes may introduce fluctuations in the number of mRNAs that masks the contribution of the transcriptional dynamics (36–38).

In contrast, experiments that catch RNA polymerase (RNAP) molecules in the process of transcribing a gene provide a more direct readout of transcription initiation dynamics. Techniques developed by Miller and his group in a series of landmark papers over several decades rely on imaging actively transcribed genes in recently lysed cells by electron microscopy (39–44). In these images, the positions of transcribing polymerases along a gene can be determined (Fig. 1 *A*). Interpolymerase distance distributions can be extracted from these positions, and, with a few reasonable assumptions (36), these can be used to extract information about the distribution of times between successive initiation events (Fig. 1 *B*). The information contained in these distributions of interpolymerase distances is akin to that obtained in live cells by fluorescently labeling nascent RNAs to observe transcription initiation events in real time at the single-cell level (19,20,45–49), as illustrated in Fig. 1 *B*.

Here, we calculate the distribution of times between successive initiation events to quantitatively test mechanistic models of transcription initiation in cells. To accomplish this, we introduce a stochastic model of transcription that incorporates both initiation and elongation kinetics. Using the derived analytical results in conjunction with simulations, we show that the kinetics of initiation leaves a signature in the distribution of transcription initiation times that can be used to discern different models of transcription initiation. To showcase the power of this approach, we have reanalyzed a set of micrographs of *Escherichia coli* genomic DNA that provided evidence of transcriptional bursting, a phenomenon that was later found to be widespread across all organisms (50). In fact, to our knowledge, McKnight et al. used Miller spreads to provide the first evidence of transcriptional bursting (51). We show that by filtering the information contained in these micrographs through our theoretical framework, we can test different models of regulation of ribosomal promoters in *E. coli* that have been proposed to account for the cell's response to an increase in rrn operon copy number. We find that some of these previously proposed models produce distributions of interpolymerase distances that are inconsistent with the Miller spread data, whereas others are in excellent agreement.

METHODS

Data analysis and parameter estimation

We evaluate the utility of our theoretical framework by employing it to gain mechanistic insights into the regulation of ribosomal genes in *E. coli*. To this end, we have reanalyzed data obtained from images of elongating RNAP molecules on ribosomal RNA (rRNA) genes in *E. coli*. Images of RNAPs were obtained from electron micrographs of fixed cells using the Miller spread technique by Voulgaris et al. (39), and the distances between neighboring RNAPs were measured by them. To extract the data from the published interpolymerase distance distribution plots in Fig. 3 *B* of (39), we used DigitizeIt, a free online tool for digitizing data plots.

Voulgaris et al. (39) increased the number of rrn operons in *E. coli* cells by inserting an rrn operon on a multicopy plasmid. They observed that the rate of rRNA transcription per operon is reduced to maintain a constant number of rRNA in the cell. In fact, electron micrograph (EM) images showed that fewer RNAP molecules were engaged in transcribing each rrn gene, consistent with previous studies (52). Moreover, RNAP molecules formed bunches separated by gaps along the genes. Although the authors ruled out transcription elongation or termination as origins of these bunches, they suggested that the bunches are caused by stochastic interruptions of initiation, or promoter-proximal elongation events. Using our theory, we analyze the distributions of intrabunch distances between RNAP.

In this study, we consider three models of transcriptional regulation of ribosomal genes: the dead-end complex model, the cooperative recruitment model, and the "ppGpp" model. For the dead-end complex model, after the loading of polymerase molecules to the promoter at a rate k_{LOAD}, each RNAP molecule either escapes the promoter at a rate k_{ESC} and starts transcribing the gene or forms a dead-end complex at the promoter at a rate k_{DEAD}. These dead-end complexes are unproductive and are removed at a rate k_{OFF}.

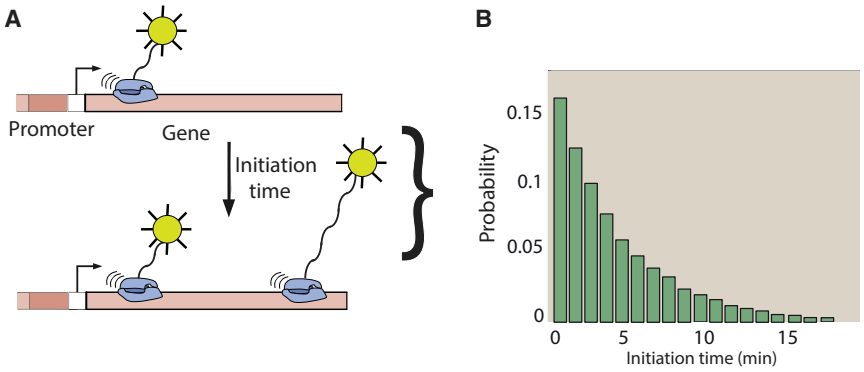

FIGURE 1 Positions of transcribing RNAP carry the signature of transcription initiation dynamics. A schematic of the key idea of this work is shown. (*A*) The times between successive transcription initiation events ("initiation time") can be extracted at the single-cell level using fluorescent reporters for nascent RNA molecules (19,20,45–49), or from electron micrograph (EM) images of RNAP caught in the process of transcribing a gene (39–44). Native elongating transcript sequencing (85) can obtain the same quantitative information as EM images. (*B*) The distribution of times between individual transcription initiation events can be extracted from experiments and compared to theoretical predictions based on stochastic models of transcription initiation. To see this figure in color, go online.

For the cooperative recruitment of RNAP by DNA supercoiling, RNAP molecules are loaded onto the promoter at a rate $k_{LOAD}{}^{LOW}$. After RNAP initiates transcription at a rate k_{ESC}, it leaves the promoter DNA in a supercoiled state, and subsequent loading of polymerases occurs at the promoter at a faster rate, $k_{LOAD}{}^{HIGH}$. The rate of relaxation of the supercoiled state is k_{RELAX}.

For the third model, the production of "control molecules" (e.g., ppGpp) reduces the initiation rate by regulating the initiation process by converting the active promoter-RNAP complexes into inactive ones. It is described by the same kinetic scheme as the dead-end complex model. However, the rate of inactivation of RNAP-DNA complex is given by k_{ON}, with every other rate remaining the same. Each of the three models considered above has five parameters.

To test these proposed models based on the experimentally observed transcriptional bunching data for rrn genes, we first extract all the model parameters by fitting the predicted inter-RNAP distance distributions to the data from wild-type *E. coli* cells with seven rrn operons, as reported in (39). The authors in (39) analyzed the interpolymerase distance distributions extracted from the micrograph images by defining a "transcriptional bunch" as a group of RNAPs separated by less than 240 bp from each other (39). The distribution of distances greater than 240 bp is referred to as interbunch distribution. Evidently, the intrabunch distribution is dictated by the biochemical rates k_{ESC} (rate of promoter escape), k_{LOAD} (rate of RNAP loading onto the promoter, which is called $k_{LOAD}{}^{HIGH}$ for the supercoiling mediated recruitment model), and τ_{clear} (time for an RNAP to clear the promoter). In Fig. 3 *C*, we show the interpolymerase distribution for the seven rrn promoters for the ribosomal genes in wild-type *E. coli* cells. Using Eq. 4 we find the probability distribution of interpolymerase distances as a function of these different rates is given by

$$p_1(x) = \frac{1}{v}\frac{k_{LOAD}k_{ESC}}{k_{LOAD}+k_{ESC}}\left[\exp\left(-k_{ESC}\left(\frac{x}{v}-\frac{30}{v}\right)\right) - \exp\left(-k_{LOAD}\left(\frac{x}{v}-\frac{30}{v}\right)\right)\right]. \quad (1)$$

Here, 30 bps is roughly the size of an RNAP molecule, and v is the rate of transcription elongation of the RNAP molecules along a gene. By comparing the data from the experiments for interpolymerase distances within a bunch and the prediction from the model, we extract k_{LOAD}, k_{ESC}, and τ_{CLEAR}. By fitting the model (see Fig. 3 *C*), we extract the rates $k_{ESC} \approx 3$/s, $k_{LOAD} \approx 3$/s, and $\tau_{clear} \approx 0.3$ s ($\tau_{clear} = (30/v)$); v is the rate of elongation, which we take to be 78 bps/s (39). These three parameters are common to all three models examined.

Next, we extract the remaining two parameters for each of these three models (see Fig. 4, *A–C*) by making use of the intrabunch distances (39). The mean intrabunch distance allows us to extract the average time the promoter spends in the inactive state, which does not lead to initiation. From Fig. 3 *A* of (39), we extract the mean gap between RNAP bunches to be ~5 s. Hence, for the three bursting models, we take k_{OFF} (dead-end complex model) = k_{OFF} (ppGpp model) = $k_{LOAD}{}^{LOW}$ (cooperative recruitment model) ≈ 0.2/s. To obtain the fifth parameter of these models, we assume that the addition of extra copies of ribosomal genes to the *E. coli* cell reduces the average transcription rate per gene to keep the level of rRNA in the cell constant. In other words, for the models considered above, the total initiation rate remains constant, or $nI = Constant$, where n is the number of ribosomal genes and I is the initiation rate on one of the genes. The initiation rate for a wild-type strain with $n = 7$ genes is $I = 1$ initiation/s (39). Using the formulas for the rate of average initiation for each of the models of transcription initiation (see the Supporting Material), we find for the cooperative recruitment model $k_{RELAX} \approx 0.055$/s and for the k_{DEAD} (dead-end complex) = k_{ON} (ppGpp model) ≈ 0.047/s.

Estimate of error bars

In the published data plots that were digitized to obtain our Fig. 3 *C*, no error bars were given. To obtain estimates for error bars on the measured frequencies of interpolymerase distances for ribosomal genes in wild-type *E. coli* cells (Fig. 3 *C*), we consider the binomial counting error. Let $i \in 1,\ldots,I$ be the histogram bins that the measured distances fall in. The probability (frequency) of a distance measurement falling in bin i is p_i. This is just a binomial trial, i.e., the data point is either in the bin or not, and therefore the variance of the measured frequency for this bin is $p_i\,(1-p_i)/N$, where N is the total number of distance measurements. The error bars in Fig. 3 *C* are the standard deviations, i.e., square root of the variance, for each bin. Based on the total number of distance measurements reported, which is roughly 1000, the relative errors on the data points are around 10%.

Limitations of the EM data

Interpolymerase distances extracted from EM images of RNAP molecules transcribing a gene have known limitations. The possibility of RNAP molecules bound to the gene of interest being lost during preparation of samples for EM imaging cannot be ruled out (53). This does open up the possibility that the observed fluctuation in the interpolymerase distances is an artifact of the experiment. Furthermore, the resolution of the EM images is typically 1.8 nm/pixel, and therefore any distance less than 5 bps cannot be resolved (54).

Limitations of the modeling framework

A crucial assumption our model makes is that interpolymerase distances along a gene of interest are governed by the initiation dynamics and not the elongation dynamics. Although for the data set analyzed in this study, this assumption is valid (as outlined in the Discussion; also see the Supporting Material), one needs to be cautious when using this assumption for other genes. Ideally, any kind of analysis of interpolymerase distances should to be done on a gene-by-gene basis. However, because of the simplicity of our modeling framework, we believe it serves as a good starting point for analyzing interpolymerase distance distribution data. Deviations from the model predictions could point to other transcriptional processes such as elongation and termination.

RESULTS

Distribution of initiation times for an arbitrary mechanism of transcription initiation can be computed from a master equation

Transcription initiation is typically regulated by transcription factors and cofactors that bind to the regulatory DNA sequences and either inhibit or aid the binding of RNAP molecules to the promoter. To connect mechanisms of transcription initiation with measured times between successive initiation events, we consider a stochastic model of transcription with a general initiation mechanism, in which the promoter can be in an arbitrary number of states defined by different constellations of bound transcription factors and cofactors. Using a chemical master equation approach (22,55,56), we show that the distribution of times between two initiation events and its moments can be computed analytically for any mechanism of transcription initiation. These equations allow us to discriminate between different

mechanisms of initiation by comparing the predicted distributions to experimental distributions of transcription initiation times.

To compute the distribution of times between successive initiation events, we assume that the promoter can be in N different discrete biochemical states and that transitions between different states occur as different transcription factors bind and fall off their respective binding sites. The rate of transition from the m-th to the n-th promoter state is $k_{m,n}$, and the rate at which an RNAP molecule initiates transcription from the m-th promoter state is $k_{m,esc}$. The assumption that the transitions between these states are random Poisson processes characterized by rate constants leads to a chemical master equation that describes the time evolution of $P_m(t)$, the probability that the promoter is in the m-th state ($m = 1, 2,.., N$) at time t:

$$\frac{dP_m}{dt} = \sum_{n=1}^{N}[k_{\mathrm{n,m}}P_n - k_{\mathrm{m,n}}P_m] - k_{m,esc}P_m. \tag{2}$$

Solving this chemical master equation for all the states m from which the promoter initiates transcription with rate $k_{m,esc}$, leads to a general formula for the probability distribution $q_1(t)$ of the time intervals between successive transcription initiation events (details of the calculation can be found in the Supporting Material):

$$q_1(t) = \sum_{m=1}^{N} k_{m,esc}P_m(t). \tag{3}$$

Assuming a uniform elongation rate v (37,57) along the gene, the distribution $q_1(t)$ of time intervals between successive initiation events directly translates into a distribution of distances $p_1(x)$ between RNAPs along the gene. In other words, the interpolymerase distance distribution along a gene is given by

$$p_1(x) = q_1\left(\frac{x}{v}\right)\frac{1}{v}. \tag{4}$$

In the ensuing investigation of regulation of ribosomal genes, Eq. 4 forms the basis of our analysis of positions of RNAP molecules along a gene at a given moment in time, which provides a quantitative test for different models of transcription initiation. Although transcription elongation of ribosomal genes is typically more complicated and involves pausing and backtracking of polymerases along the gene (58), here we assume that transcriptional pausing happens on timescales that are negligible compared to the times between transcription initiation events (59). For a detailed discussion of this model assumption, see the Supporting Material.

The distribution of transcription initiation times can be used to discern between different models of initiation

To illustrate how the distribution of times between successive initiation events can be used to extract mechanistic insights about the process of transcription initiation, we consider three different models of initiation as case studies (see Fig. 2).

Poisson (single rate-limiting step) model

The Poisson model is the null model of initiation, which is usually associated with constitutive promoters (12). In this model, initiation happens with a constant probability of k_{LOAD} per unit time, as shown in Fig. 2 *A*. In bacteria, this step could, at the molecular scale, represent the rate of loading of RNAP molecules to the promoter DNA, whereas for eukaryotes, this step could correspond to the formation of the preinitiation complex. As obtained from Eq. 3, the one-state model is characterized by exponentially distributed times between successive initiation events. One of the key properties of an exponential distribution is that its mean and standard deviation are equal. Therefore, the CV^2 (defined as the ratio of the variance to the square of the mean) is always equal to one, independent of the rate k_{LOAD}, as shown in Fig. 2 *A*.

Two-limiting-steps model

Next, we consider a model in which initiation happens in two sequential rate-limiting steps. This is the situation in which two steps in the sequence of events leading to initiation are of comparable duration. For example, in bacteria, the first step could correspond to an RNAP molecule binding to the promoter with a rate k_{LOAD}. In eukaryotes, this step could represent the loading of the transcriptional machinery at the promoter (23,48). In the second step, the promoter-bound RNAP molecule escapes the promoter at a rate k_{ESC} and starts transcribing the gene. For several promoters in yeast (36) and *E. coli* (60), it has been reported that initiation proceeds through two sequential steps.

For this case, using Eq. 3, we find that the waiting-time distribution between successive initiation events is gamma distributed. This result agrees with previous theoretical studies (25) and leads to the following relationships between the kinetic rates associated with this mechanism (k_{LOAD} and k_{ESC}) and the mean and the coefficient of variation of the waiting-time distribution:

$$\begin{aligned} Mean &= \frac{k_{ESC} + k_{LOAD}}{k_{ESC}k_{LOAD}}, \\ CV^2 &= \frac{k_{ESC}^2 + k_{LOAD}^2}{(k_{ESC} + k_{LOAD})^2}. \end{aligned} \tag{5}$$

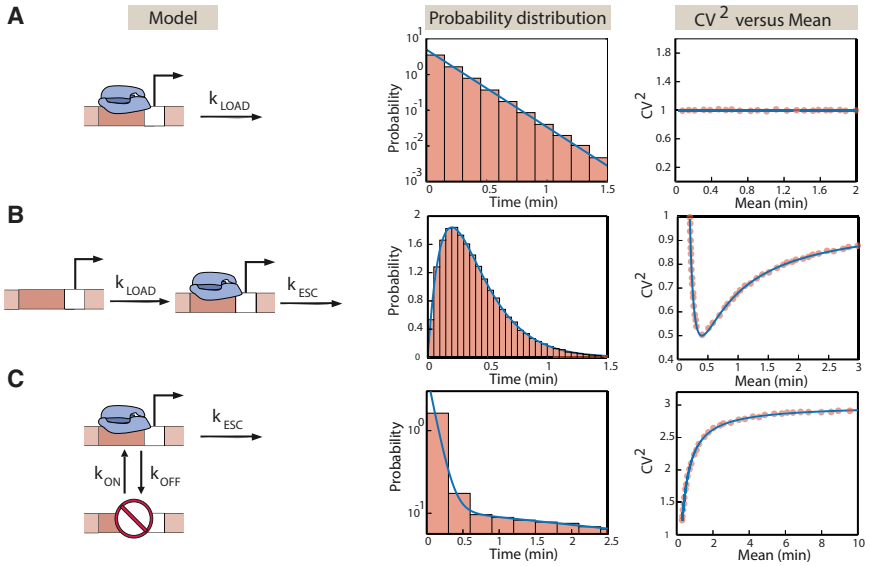

FIGURE 2 Different models of transcriptional regulation lead to distinct signatures in the initiation times. (*A*) The one-step model of transcription initiation is depicted. Initiation happens at a constant rate k_{LOAD}. The times between successive initiation events are exponentially distributed. The square of the coefficient of variation is plotted as a function of the mean, in which we change the mean by changing the rate of initiation, k_{LOAD}. We confirm the analytical results using Gillespie simulations (65). The histograms and closed circles represent simulation results. (*B*) The two-step model of transcription initiation is depicted. Initiation happens in two sequential steps: the rate of RNAP loading onto the promoter occurs with rate k_{LOAD}, followed by RNAP escaping the promoter, leading to transcript elongation at a rate k_{ESC}. The distribution of times between successive initiation events and the square of the coefficient of variation of the distribution as a function of the mean are shown. To change the mean, we change the rate of loading of RNAP polymerase molecules on the promoter, k_{LOAD}. As in (*A*), simulation results are compared to the analytical results. (*C*) The ON-OFF model is depicted. The promoter switches between two states: an active and an inactive one. The rate of switching from the active state to the inactive state is k_{OFF} and from the inactive to the active state is k_{ON}. From the active state, transcription initiation proceeds with a probability per unit time, k_{ESC}. The distribution of times between initiation events and the square of the coefficient of variation as a function of the mean are shown. Results from Gillespie simulations (65) are shown for comparison. To change the mean, we tune the rate k_{ON} of switching from the inactive to the active state. To illustrate the distinctive impact of the different initiation models on the distribution and moments of the times between successive initiation events, we use the following parameters: $k_{OFF} = 5$/min, $k_{ON} = 0.435$/min, $k_{LOAD} = 0.14$/min, and $k_{ESC} = 0.14$/min, which are characteristic of yeast promoters (36). To see this figure in color, go online.

As shown in Fig. 2 *B*, when we tune either one of the two rates of the model while keeping the other one constant, the coefficient of variation initially decreases as a function of the mean, develops a minimum when the two rates become equal, and then asymptotically goes to one. In the limit of one rate being much slower than the other one, the waiting-time distribution becomes exponential, leading to a coefficient of variation of one.

ON-OFF promoter

The third scenario we consider is the ON-OFF model of initiation. This model of initiation has been established as the canonical model of transcriptional regulation for both bacteria (19) and eukaryotes (17,21,61–64). In this model, the promoter switches between two states: an active state from which transcription initiation can occur, and an inactive state from which initiation does not occur. The two states might correspond to a free promoter and one bound by a repressor protein, or a promoter covered by nucleosomes. The rate of switching from the active to the inactive state is k_{OFF} and from inactive to the active state is k_{ON}. The rate of initiation from the active state is k_{ESC}.

In this case, we find that the waiting-time distribution between successive initiation events is given by a sum of two exponentials, as shown in Fig. 2 *C*. Thus, it can be distinguished from a single exponential expected from the one-state promoter on the condition that the decay constants of the two exponentials are well-separated in magnitude. The mean and the coefficient of variation as functions of the different biochemical rates are given by

$$\begin{aligned} Mean &= \frac{k_{ON} + k_{OFF}}{k_{ON} k_{ESC}}, \\ CV^2 &= 1 + \frac{2 k_{OFF} k_{ESC}}{(k_{ON} + k_{OFF})^2}. \end{aligned} \tag{6}$$

When we tune the rate k_{ON}, the CV^2 increases as a function of the mean and eventually saturates, as shown in Fig. 2 *C*.

We compare our analytical results for the three models described above against Gillespie simulations (65). This allows us to numerically generate multiple time traces of initiation events. From these different time traces, we obtain the distribution of initiation times as well as the corresponding values of the mean and variance. The histograms for the times between initiation events for all the three models are shown in Fig. 2, *A–C*. We also show the coefficient of variation as a function of the mean for these models as we tune the relevant rates in Fig. 2, *A–C*. These results imply that we can discern these three different models of initiation based on the predictions they make for the waiting-time distribution of consecutive initiation events as a function of the different experimentally tunable parameters.

The dynamics of transcription initiation of ribosomal genes in *E. coli* can be extracted from images of transcribing polymerases in fixed cells

To demonstrate how the distribution of interpolymerase distances along a gene can be used to extract dynamical information about the process of transcription initiation in vivo, we have reanalyzed data obtained from images of elongating RNAP molecules on rRNA genes in *E. coli*, which were obtained using the Miller spread technique by Voulgaris et al. (39). In Fig. 3 *C*, we reproduce the interpolymerase distance distribution for the seven ribosomal genes in wild-type *E. coli* cells (39). A remarkable feature of this distribution is the presence of a peak at small distances. This is inconsistent with a Poisson initiation mechanism (19) described above. Indeed, the presence of a maximum in the probability at intermediate distances suggests the two-limiting-steps model of initiation (Fig. 3 *B*), in which, for example, the polymerase first binds to the promoter and then escapes the promoter, leading to elongation, and in which the two steps occur with comparable rates. Recent in vivo studies in yeast have shown that initiation can proceed in multiple sequential steps, in which the rates involving these steps have comparable magnitude (36). In addition, when analyzing this data, we consider the time it takes for the polymerase to clear the promoter by elongating through it.

To test the hypothesis of two sequential steps leading to initiation, we fit Eq. 4 to the experimental distribution obtained from the images and assume an elongation speed of 78 bp/s (as measured elsewhere (39)). We find that the two-limiting-steps model is in good agreement with the data. Furthermore, the fit provides estimates for the rates of promoter escape, the rate of RNAP loading onto the promoter, and time to clear the promoter ($k_{LOAD} \approx 3$/s, $k_{ESC} \approx 3$/s, and $\tau_{CLEAR} \approx 0.3$ s), all of which are in good agreement with previous measurements (66) (see the Methods).

Bursting accompanies the downregulation of the transcription of rrn operons in the presence of additional copies of the rrn genes

In a second set of experiments, EM images were used to shed light on a previously reported effect (52), namely that the transcriptional activity of individual rrn genes is inversely proportional to the copy number of these genes in such a way that the net transcriptional output of rrn genes in the cell is kept constant. Surprisingly, the EM showed that when the copy number of ribosomal genes was altered by placing extra rrn genes on plasmids, a very different pattern of polymerase occupancy of the rrn genes emerged. Besides the mean number of RNAP molecules along each gene decreasing with increased gene-copy number, it was also observed that RNAP molecules were then grouped in bunches along the gene, which is indicative of transcriptional bursting. Moreover, it was observed that the mean and the variance of the interpolymerase distances within a bunch remained unchanged as the copy number varied (39). To explain the observed reduction in the mean number of RNAP molecules along each gene, Bremer et al. (67) focused on the fact that a significant fraction of RNAPs in the cell are engaged in transcribing the ribosomal genes. In that case, changing the gene-copy number will significantly alter the concentration of free polymerases available for transcription initiation. This will, in turn, decrease the rate of transcription initiation, which is assumed to be proportional to the free RNAP concentration. This "free RNAP hypothesis" (67), along with the two-step model of transcription initiation (Fig. 1 *B*), predicts that the mean number of RNAPs per gene will decrease in response to an increase in the number of genes, as observed experimentally, but it cannot account for the observed bunching of polymerases, as was shown in a previous theoretical study (25). Therefore, to account for the bunching of RNAPs seen in experiments by Voulgaris et al. (39), it is necessary to consider models of initiation with promoter states that are off-pathway to elongation. Indeed, several such models have been proposed to explain the downregulation of individual rrn genes in response to an increase in the gene-copy number (39). These models can be broadly classified into three different classes. Using our mathematical framework, we put to the test these different classes of models with the goal of gaining insight into the unresolved question of how the transcription of ribosomal genes is regulated in response to an increase in the gene-copy number.

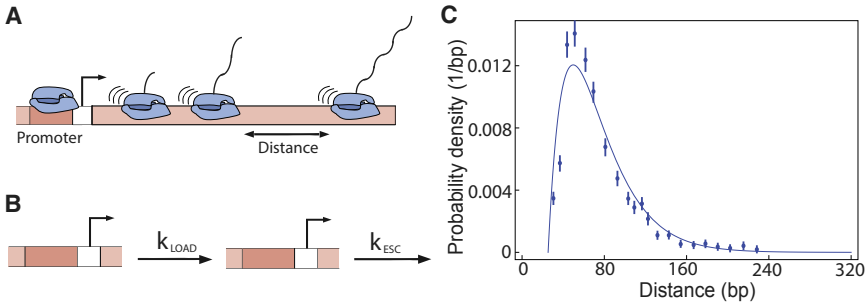

FIGURE 3 Initiation of transcription of ribosomal genes in *E. coli*. (*A*) The positions of RNAP molecules transcribing a gene at a given instant in time can be obtained from electron microscopy images or native elongating transcript sequencing (85). (*B*) The two-step model of transcription initiation is shown. (*C*) The fit (*line*) of the two-step model to the interpolymerase distance distribution data (points) obtained by Voulgaris et al. (39) for ribosomal genes in *E. coli* is given. The different biochemical rates we extract from the fit are as follows: k_{ESC} (rate of promoter escape) ≈ 3/s, k_{LOAD} (rate of RNAP loading onto the promoter) ≈ 3/s, and τ_{clear} (time for an RNAP to clear the promoter) ≈ 0.3 s, taking the elongation speed $v = 78$ bps/s, as reported in experiments (39). To see this figure in color, go online.

The first proposed class of models considers the formation of long-lived, nonproductive initiation complexes at the promoter (68–70). For example, nonproductive complexes that cannot exit the abortive initiation state into productive elongation have been observed in vitro (25). The second class of models assumes cooperative recruitment of RNAP molecules to the promoter by an RNAP molecule already present on the promoter (71–73). For instance, when an RNAP molecule initiates transcription, it can leave the promoter DNA in a supercoiled state, as illustrated in Fig. 3 *B*. In the supercoiled state, the energy barrier for melting a strand of DNA to make a transcriptional bubble is lowered, leading to an increased rate of RNAP loading onto the promoter (25). Both these classes of models can lead to bunching of RNAP molecules on the gene, as observed in the experiments (39). However, a common feature of both classes of models is that they incorporate the "free RNAP" hypothesis, in that an increase in the number of rrn genes in the cells leads to a lower concentration of free RNAP and thus to a reduction in the rate of RNAP loading onto the promoter (*red arrows* in Fig. 4, *A* and *B*). Consequently, for both classes of models, the mean and variance of the intrapolymerase distance distribution within a bunch will increase as the rrn gene-copy number is increased, which is in contrast to the experimental observations as shown in Fig. 4, *A* and *B*. It must be noted that the interpolymerase distance distribution within a bunch is dictated by the biochemical rates k_{ESC} (rate of promoter escape), k_{LOAD} (rate of RNAP loading onto the promoter, which we denote as k_{LOAD}^{HIGH} in the supercoiling mediated recruitment model), and τ_{clear} (time for an RNAP to clear the promoter). Hence, any change in one of these rates would lead to a change in the interpolymerase distance distribution within a bunch, contrary to what is seen in experiments.

The third proposed class of regulatory models for the transcription of ribosomal genes considers a secondary molecular messenger, the abundance of which in the cell is regulated in response to the changing number of ribosomal genes. These molecules either inhibit or activate the transcription of rrn genes (Fig. 3 *C*) (74). One such model is based on the alarmone nucleotide molecule ppGpp, which is capable of inactivating the promoter-RNAP complex upon binding to the polymerase (74,75). Formation of such inactive complexes can lead to bursty transcription initiation dynamics caused by the inactivated polymerase

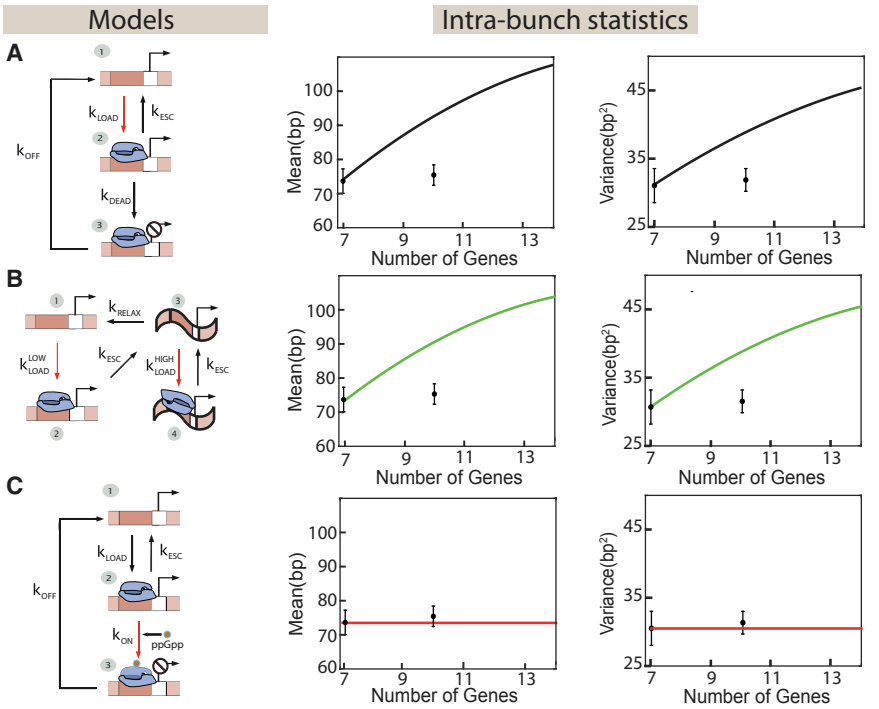

FIGURE 4 Different models of transcriptional regulation of ribosomal genes can be tested by tuning the gene-copy number. (*A* and *B*) Models of transcription initiation that rely solely on the interaction of RNAP with promoter DNA are shown. (*A*) This class of model considers the formation of long-lived nonproductive initiation complexes at the promoter by RNAP molecules (25,70). After binding the promoter at a rate k_{LOAD}, each RNAP can initiate transcription at a rate k_{ESC} or make a dead-end complex at the promoter at a rate k_{DEAD}. These dead-end complexes are unproductive and are removed at a rate k_{OFF}. The change in gene-copy number affects the binding rate of RNAP molecules to the promoter because of a change in the free RNAP concentration, as indicated by the red arrow. Theory predicts that the mean and variance of distances between RNAPs within a bunch increase with the gene number, contrary to experiments on ribosomal genes in *E. coli*. (*B*) The cooperative recruitment of RNAP by DNA supercoiling is shown. RNAP molecules are loaded onto the promoter at a rate k_{LOAD}^{LOW}. After RNAP initiates transcription at a rate k_{ESC}, it leaves the promoter DNA in a supercoiled state, and subsequent loading of RNAP polymerases at the promoter happens at a faster rate, k_{LOAD}^{HIGH}. The rate of relaxation of the supercoiled state is k_{RELAX}. The change in gene-copy number affects both polymerase loading rates (*red arrows*) because of the change in free RNAP concentration. The model predicts that the mean and variance of the intrabunch RNAP distances increase with the gene-copy number, contrary to measurements in *E. coli*. (*C*) As the number of genes increases, the rate of rRNA production increases. This triggers the production of control molecules (e.g., ppGpp), which then reduce the initiation rate by modulating the promoter-RNAP interactions. ppGpp regulates the initiation process by converting the active promoter-RNAP complexes into inactive ones. It is described by the same kinetic scheme as the dead-end complex model (*A*) with a critical difference: in this case, it is the rate of inactivation of RNAP-promoter complex (k_{ON}) because of ppGpp binding to the complex that is tuned as the copy number of ribosomal genes is increased (*red arrow*), whereas the rate of RNAP loading onto the promoter is unchanged. The mean and variance of distances between RNAPs within a bunch are predicted to remain constant, as observed in experiments. In all the plots, the two data points shown are taken from (39). To see this figure in color, go online.

blocking the promoter. Another set of experiments has shown that the number of ppGpp molecules increases with increasing ribosomal gene-copy number (76). Consistent with these observations, we propose a kinetic model of transcriptional regulation of ribosomal genes in which RNAP-promoter complexes are inactivated by ppGpp molecules that increase in concentration as the rrn gene-copy number is increased (shown in Fig. 4 *C*). The key assumption of this class of models that distinguishes it from the first two is that an increase in the number of rrn genes does not significantly change the number of free RNAPs in the cell (74,75).

The mean and variance of the interpolymerase distance distributions within a bunch for the secondary messenger mechanism are consistent with the experimental results. Because the key assumption of this model is that an increase in the number of genes does not significantly change the number of free RNAPs in the cell (74,75), the rate of RNAP loading onto the promoter remains unchanged. This has the effect of keeping the distribution of distances between RNAPs within the bunch unchanged (Fig. 4 *C*). At the same time, the proposed mechanism generates the bursting kinetics that is observed in EM, suggesting that it may be a candidate mechanism for regulating the transcription of rrn genes. We note that any mechanism based on stochastically and transiently preventing promoter escape by a bound polymerase would have the same outcome. Although ppGpp-based mechanism is a plausible candidate, as this mode of its action has been documented before, whether it is in fact responsible for the changes in transcription initiation that accompany the increase in rrn gene-copy number is still an open question.

DISCUSSION

The dynamics of transcription in live cells is poorly understood. Because of the difficulties in directly imaging the process of transcription (19,47,48,50,77), experimental methods for counting the products of transcription (such as RNA and protein molecules) in single cells have been developed over the past years. The protein and mRNA distributions carry the signature of the dynamics of transcription and hence can be exploited to decipher the underlying mechanisms of transcriptional regulation (12,16,47,78). However, both mRNA and protein counts are affected by noisy processes other than transcription, such as mRNA processing, binomial partitioning, nonlinear degradation of mRNA molecules, etc (31,33,34,79–82), which can potentially mask the signature of transcription on protein and mRNA distributions.

EM images of RNAP molecules engaged in the process of transcribing a gene at a given instant in time (also known as "Miller spreads") provide information about the position of polymerase molecules on the gene (39–44). Similar information can be extracted by observing transcription initiation events in real time using fluorescent reporters (19,20,45–49). These measurements are not affected by posttranscriptional processes and are therefore more direct readouts of transcription compared to mRNA and protein counting (83). In this study, we have derived mathematical equations that allow us to interpret and analyze the interpolymerase distance distribution, or equivalently, the waiting-time distribution between successive initiation events across a population of isogenic cells. To demonstrate the potential utility of our analytical results, we fit the interpolymerase distance distribution for ribosomal genes in *E. coli* (acquired from EM) to a theoretical distribution computed for a two-step model of initiation. The model fits the data well, allowing us to extract the rates that characterize transcription initiation dynamics.

It must be noted that other mechanisms, like fluctuations in the promoter clearing time, could also generate a peaked distribution of interpolymerase distances in Fig. 3 *C*. A multistep initiation model, in which initiation happens in more than two sequential steps, could also account for a peaked interpolymerase distribution. Although we cannot rule out these and other possibilities, we make use of the two-step initiation model as a simple scenario that explains the data well and provides mechanistic insight into the dynamics of initiation that leads to experimentally testable predictions.

We also reanalyze data from images of RNAP transcribing ribosomal genes in wild-type and mutant strains of *E. coli*. We show that previously proposed mechanisms, based on the effect that extra rrn gene copies might have on the concentration of free RNAPs in the cell, are inconsistent with the observed interpolymerase distance distributions. In contrast, we find that an alternative possibility (74), in which the alarmone nucleotide ppGpp may interact with promoter-bound RNAP and prevent promoter escape, produces bursting transcription initiation kinetics that is consistent with experimental observations.

Our model assumes that interpolymerase distance distributions are governed by transcription initiation dynamics. However, stochastic transcription elongation dynamics can also impact interpolymerase distances. Depending on the timescales of the two processes, they will affect the distribution differently (see Supporting Material for details). If the initiation timescales are much longer than the elongation timescales, the distribution of RNAP distances will remain largely unaffected by the elongation dynamics, as in the case of most of the mRNA promoters in *E. coli* (36). Because the ribosomal genes are very highly transcribed, the initiation and elongation timescales become comparable, and hence the effect of elongation on the RNAP distance distributions can be significant (58). To better understand this interplay of the two, we use a model of initiation and elongation proposed by Klumpp et al. (58) that incorporates ubiquitous pauses on the elongating RNAPs. Using stochastic simulations (65,84), we explore the effect of elongation dynamics on the interpolymerase distance

distributions for ribosomal genes (for a detailed discussion, see the Supporting Material). We find that the interpolymerase distribution, for rate parameters that are pertinent to experiments on rrn genes (39), is solely a result of the transcription initiation dynamics, and that elongation has little impact on it.

We believe that the approach and ideas presented here will be helpful to uncovering detailed kinetic information about the process of transcription initiation in live cells.

SUPPORTING MATERIAL

Supporting Materials and Methods, two figures, and two tables are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(18)30407-7.

AUTHOR CONTRIBUTIONS

S.C., J.K., and A.S. designed and performed the research. S.C. analyzed data. S.C., J.K., and A.S. wrote the manuscript.

ACKNOWLEDGMENTS

We wish to thank Rob Phillips, Hernan Garcia, Jeff Gelles and Timothy Harden for years of stimulating discussions and shared thoughts about transcriptional dynamics.

This work was supported by the National Science Foundation through grants DMR-1206146 and DMR-1610737 and the Simons Foundation through a Targeted Grant in the Mathematical Modeling of Living Systems.

REFERENCES

1. N. H. Bergman, ed 2007. Comparative Genomics: Volumes 1 and 2. Humana Press, Totowa, NJ.
2. Varki, A., and T. K. Altheide. 2005. Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Res.* 15:1746–1758.
3. Cheng, Y., Z. Ma, …, M. P. Snyder; mouse ENCODE Consortium. 2014. Principles of regulatory information conservation between mouse and human. *Nature.* 515:371–375.
4. Stergachis, A. B., S. Neph, …, J. A. Stamatoyannopoulos. 2014. Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature.* 515:365–370.
5. Lin, S., Y. Lin, …, M. P. Snyder. 2014. Comparison of the transcriptional landscapes between human and mouse tissues. *Proc. Natl. Acad. Sci. USA.* 111:17224–17229.
6. Loots, G. G. 2008. Genomic identification of regulatory elements by evolutionary sequence comparison and functional analysis. *Adv. Genet.* 61:269–293.
7. Alberts, B., A. Johnson, …, P. Walter. 2002. Molecular Biology of the Cell, Fourth Edition. Garland Science, New York.
8. Bintu, L., N. E. Buchler, …, R. Phillips. 2005. Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.* 15:125–135.
9. Garcia, H. G., A. Sanchez, …, R. Phillips. 2010. Transcription by the numbers redux: experiments and calculations that surprise. *Trends Cell Biol.* 20:723–733.
10. Garcia, H. G., and R. Phillips. 2011. Quantitative dissection of the simple repression input-output function. *Proc. Natl. Acad. Sci. USA.* 108:12173–12178.
11. Rydenfelt, M., R. S. Cox, III, …, R. Phillips. 2014. Statistical mechanical model of coupled transcription from multiple promoters due to transcription factor titration. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 89:012702.
12. 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.
13. Castelnuovo, M., S. Rahman, …, D. Zenklusen. 2013. Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat. Struct. Mol. Biol.* 20:851–858.
14. Raj, A., and A. van Oudenaarden. 2009. Single-molecule approaches to stochastic gene expression. *Annu. Rev. Biophys.* 38:255–270.
15. Jones, D. L., R. C. Brewster, and R. Phillips. 2014. Promoter architecture dictates cell-to-cell variability in gene expression. *Science.* 346:1533–1536.
16. Gandhi, S. J., D. Zenklusen, …, R. H. Singer. 2011. Transcription of functionally related constitutive genes is not coordinated. *Nat. Struct. Mol. Biol.* 18:27–34.
17. Raj, A., C. S. Peskin, …, S. Tyagi. 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* 4:e309.
18. Padovan-Merhar, O., G. P. Nair, …, A. Raj. 2015. Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol. Cell.* 58:339–352.
19. Golding, I., J. Paulsson, …, E. C. Cox. 2005. Real-time kinetics of gene activity in individual bacteria. *Cell.* 123:1025–1036.
20. Chubb, J. R., T. Trcek, …, R. H. Singer. 2006. Transcriptional pulsing of a developmental gene. *Curr. Biol.* 16:1018–1025.
21. Kepler, T. B., and T. C. Elston. 2001. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* 81:3116–3136.
22. Sanchez, A., S. Choubey, and J. Kondev. 2013. Stochastic models of transcription: from single molecules to single cells. *Methods.* 62:13–25.
23. Bothma, J. P., H. G. Garcia, …, M. Levine. 2014. Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. *Proc. Natl. Acad. Sci. USA.* 111:10598–10603.
24. Das, D., S. Dey, …, S. Choubey. 2017. Effect of transcription factor resource sharing on gene expression noise. *PLoS Comput. Biol.* 13:e1005491.
25. Mitarai, N., I. B. Dodd, …, K. Sneppen. 2008. The generation of promoter-mediated transcriptional noise in bacteria. *PLoS Comput. Biol.* 4:e1000109.
26. Elgart, V., T. Jia, …, R. Kulkarni. 2011. Connecting protein and mRNA burst distributions for stochastic models of gene expression. *Phys. Biol.* 8:046001.
27. Singh, A., B. Razooky, …, L. S. Weinberger. 2010. Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression. *Biophys. J.* 98:L32–L34.
28. Dar, R. D., B. S. Razooky, …, L. S. Weinberger. 2012. Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc. Natl. Acad. Sci. USA.* 109:17454–17459.
29. Kumar, N., A. Singh, and R. V. Kulkarni. 2015. Transcriptional bursting in gene expression: analytical results for general stochastic models. *PLoS Comput. Biol.* 11:e1004292.
30. Munsky, B., G. Neuert, and A. van Oudenaarden. 2012. Using gene expression noise to understand gene regulation. *Science.* 336:183–187.
31. Platini, T., T. Jia, and R. V. Kulkarni. 2011. Regulation by small RNAs via coupled degradation: mean-field and variational approaches. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 84:021928.
32. Dong, G. Q., and D. R. McMillen. 2008. Effects of protein maturation on the noise in gene expression. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 77:021908.

33. Singh, A., and P. Bokes. 2012. Consequences of mRNA transport on stochastic variability in protein levels. *Biophys. J.* 103:1087–1096.

34. Melamud, E., and J. Moult. 2009. Stochastic noise in splicing machinery. *Nucleic Acids Res.* 37:4873–4886.

35. Baker, C., T. Jia, and R. V. Kulkarni. 2012. Stochastic modeling of regulation of gene expression by multiple small RNAs. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 85:061915.

36. Choubey, S., J. Kondev, and A. Sanchez. 2015. Deciphering transcriptional dynamics in vivo by counting nascent RNA molecules. *PLoS Comput. Biol.* 11:e1004345.

37. Xu, H., S. O. Skinner, …, I. Golding. 2016. Stochastic kinetics of nascent RNA. *Phys. Rev. Lett.* 117:128101.

38. Choubey, S. 2018. Nascent RNA kinetics: transient and steady state behavior of models of transcription. *Phys. Rev. E.* 97:022402.

39. Voulgaris, J., S. French, …, C. L. Squires. 1999. Increased rrn gene dosage causes intermittent transcription of rRNA in Escherichia coli. *J. Bacteriol.* 181:4170–4175.

40. French, S. L., Y. N. Osheim, …, A. L. Beyer. 2003. In exponentially growing Saccharomyces cerevisiae cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes. *Mol. Cell. Biol.* 23:1558–1568.

41. French, S. L., M. L. Sikes, …, A. L. Beyer. 2011. Distinguishing the roles of Topoisomerases I and II in relief of transcription-induced torsional stress in yeast rRNA genes. *Mol. Cell. Biol.* 31:482–494.

42. French, S. L., and O. L. Miller, Jr. 1989. Transcription mapping of the Escherichia coli chromosome by electron microscopy. *J. Bacteriol.* 171:4207–4216.

43. Gotta, S. L., O. L. Miller, Jr., and S. L. French. 1991. rRNA transcription rate in Escherichia coli. *J. Bacteriol.* 173:6647–6649.

44. Condon, C., S. French, …, C. L. Squires. 1993. Depletion of functional ribosomal RNA operons in Escherichia coli causes increased expression of the remaining intact copies. *EMBO J.* 12:4305–4315.

45. Bertrand, E., P. Chartrand, …, R. M. Long. 1998. Localization of ASH1 mRNA particles in living yeast. *Mol. Cell.* 2:437–445.

46. Yunger, S., L. Rosenfeld, …, Y. Shav-Tal. 2010. Single-allele analysis of transcription kinetics in living mammalian cells. *Nat. Methods.* 7:631–633.

47. Larson, D. R., D. Zenklusen, …, R. H. Singer. 2011. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science.* 332:475–478.

48. Garcia, H. G., M. Tikhonov, …, T. Gregor. 2013. Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. *Curr. Biol.* 23:2140–2145.

49. Lucas, T., T. Ferraro, …, N. Dostatni. 2013. Live imaging of bicoid-dependent transcription in Drosophila embryos. *Curr. Biol.* 23:2135–2139.

50. Sanchez, A., and I. Golding. 2013. Genetic determinants and cellular constraints in noisy gene expression. *Science.* 342:1188–1193.

51. McKnight, S. L., and O. L. Miller, Jr. 1979. Post-replicative nonribosomal transcription units in D. melanogaster embryos. *Cell.* 17:551–563.

52. Baracchini, E., and H. Bremer. 1991. Control of rRNA synthesis in Escherichia coli at increased rrn gene dosage. Role of guanosine tetraphosphate and ribosome feedback. *J. Biol. Chem.* 266:11753–11760.

53. Dickinson, P., P. R. Cook, and D. A. Jackson. 1990. Active RNA polymerase I is fixed within the nucleus of HeLa cells. *EMBO J.* 9:2207–2214.

54. Osheim, Y. N., S. L. French, …, A. L. Beyer. 2009. Electron microscope visualization of RNA transcription and processing in Saccharomyces cerevisiae by Miller chromatin spreading. *Methods Mol. Biol.* 464:55–69.

55. Sanchez, A., S. Choubey, and J. Kondev. 2013. Regulation of noise in gene expression. *Annu. Rev. Biophys.* 42:469–491.

56. Sánchez, A., and J. Kondev. 2008. Transcriptional control of noise in gene expression. *Proc. Natl. Acad. Sci. USA.* 105:5081–5086.

57. Rosenfeld, L., E. Kepten, …, Y. Garini. 2015. Single-site transcription rates through fitting of ensemble-averaged data from fluorescence recovery after photobleaching: a fat-tailed distribution. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 92:032715.

58. Klumpp, S., and T. Hwa. 2008. Stochasticity and traffic jams in the transcription of ribosomal RNA: intriguing role of termination and antitermination. *Proc. Natl. Acad. Sci. USA.* 105:18159–18164.

59. Dennis, P. P., M. Ehrenberg, …, H. Bremer. 2009. Varying rate of RNA chain elongation during rrn transcription in Escherichia coli. *J. Bacteriol.* 191:3740–3746.

60. Muthukrishnan, A. B., M. Kandhavelu, …, A. S. Ribeiro. 2012. Dynamics of transcription driven by the tetA promoter, one event at a time, in live Escherichia coli cells. *Nucleic Acids Res.* 40:8472–8483.

61. Paulsson, J. 2004. Summing up the noise in gene networks. *Nature.* 427:415–418.

62. 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.

63. Peccoud, J., and B. Ycart. 1995. Markovian modeling of gene-product synthesis. *Theor. Popul. Biol.* 48:222–234.

64. Hornos, J. E., D. Schultz, …, P. G. Wolynes. 2005. Self-regulating gene: an exact solution. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 72:051907.

65. Gillespie, D. T. 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81:2340–2361.

66. Dennis, P. P., M. Ehrenberg, and H. Bremer. 2004. Control of rRNA synthesis in Escherichia coli: a systems biology approach. *Microbiol. Mol. Biol. Rev.* 68:639–668.

67. Bremer, H., P. Dennis, and M. Ehrenberg. 2003. Free RNA polymerase and modeling global transcription in Escherichia coli. *Biochimie.* 85:597–609.

68. Stepanova, E., J. Lee, …, S. Borukhov. 2007. Analysis of promoter targets for Escherichia coli transcription elongation factor GreA in vivo and in vitro. *J. Bacteriol.* 189:8772–8785.

69. Susa, M., T. Kubori, and N. Shimamoto. 2006. A pathway branching in transcription initiation in Escherichia coli. *Mol. Microbiol.* 59:1807–1817.

70. Kubori, T., and N. Shimamoto. 1996. A branched pathway in the early stage of transcription by Escherichia coli RNA polymerase. *J. Mol. Biol.* 256:449–457.

71. Liu, L. F., and J. C. Wang. 1987. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA.* 84:7024–7027.

72. Lim, H. M., D. E. Lewis, …, S. Adhya. 2003. Effect of varying the supercoiling of DNA on transcription and its regulation. *Biochemistry.* 42:10718–10725.

73. Opel, M. L., and G. W. Hatfield. 2001. DNA supercoiling-dependent transcriptional coupling between the divergently transcribed promoters of the ilvYC operon of Escherichia coli is proportional to promoter strengths and transcript lengths. *Mol. Microbiol.* 39:191–198.

74. Maitra, A., I. Shulgina, and V. J. Hernandez. 2005. Conversion of active promoter-RNA polymerase complexes into inactive promoter bound complexes in E. coli by the transcription effector, ppGpp. *Mol. Cell.* 17:817–829.

75. Barker, M. M., T. Gaal, …, R. L. Gourse. 2001. Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. *J. Mol. Biol.* 305:673–688.

76. Gourse, R. L., Y. Takebe, …, M. Nomura. 1985. Feedback regulation of rRNA and tRNA synthesis and accumulation of free ribosomes after conditional expression of rRNA genes. *Proc. Natl. Acad. Sci. USA.* 82:1069–1073.

77. So, L. H., A. Ghosh, …, I. Golding. 2011. General properties of transcriptional time series in Escherichia coli. *Nat. Genet.* 43:554–560.

78. Taniguchi, Y., P. J. Choi, …, X. S. Xie. 2010. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science.* 329:533–538.

79. Jia, T., and R. V. Kulkarni. 2010. Post-transcriptional regulation of noise in protein distributions during gene expression. *Phys. Rev. Lett.* 105:018101.

80. Huh, D., and J. Paulsson. 2011. Random partitioning of molecules at cell division. *Proc. Natl. Acad. Sci. USA.* 108:15004–15009.

81. Huh, D., and J. Paulsson. 2011. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat. Genet.* 43:95–100.

82. Schmidt, U., E. Basyuk, …, E. Bertrand. 2011. Real-time imaging of cotranscriptional splicing reveals a kinetic model that reduces noise: implications for alternative splicing regulation. *J. Cell Biol.* 193:819–829.

83. Little, S. C., M. Tikhonov, and T. Gregor. 2013. Precise developmental gene expression arises from globally stochastic transcriptional activity. *Cell.* 154:789–800.

84. Gillespie, D. T. 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* 22:403–434.

85. Churchman, L. S., and J. S. Weissman. 2012. Native elongating transcript sequencing (NET-seq). *Curr. Protoc. Mol. Biol.* Chapter 4:Unit 4.14.1–Unit 4.14.17.

Biophysical Journal, Volume *114*

Supplemental Information

Distribution of Initiation Times Reveals Mechanisms of Transcriptional Regulation in Single Cells

Sandeep Choubey, Jane Kondev, and Alvaro Sanchez

Supplementary Information

Distribution of initiation times reveals mechanisms of transcriptional regulation in single cells

Sandeep Choubey[1], Jane Kondev[1], Alvaro Sanchez[2,3]

[1]Department of Physics, Brandeis University, Waltham, Massachusetts, [2]Rowland Institute at Harvard, Harvard University, Cambridge, Massachusetts, [3]Department of Ecology and Evolutionary Biology, Microbial Sciences Institute, Yale University, New Haven;

Distribution of transcription initiation times for an arbitrary promoter architecture

In the following, we formulate a general theory of transcription initiation with an arbitrary promoter architecture to connect the distribution of times between transcription initiation events (or, equivalently, the distribution of distances between transcribing RNA polymerases along a gene) with mechanisms of transcription initiation. We assume that the promoter may exist in N different biochemical states (as defined by the different binding states of transcription factors, DNA conformations, etc.), and stochastically transitions between these states, causing fluctuations in the initiation rate. This stochastic initiation dynamics is followed by RNA elongation for which we assume that every RNA polymerase (RNAP) molecule moves along the gene at a constant speed v.

For this model we compute the inter-RNAP distance distribution along a gene of interest. In order to find the distances between adjacent RNAP molecules, we first compute the distribution of times between transcription initiation events $q_1(t)$ using a master equation approach (1–3). The rate of transitions from the m-th to the n-th state is $k_{m,n}$. The rate at which an RNAP molecules escape the m-th promoter state leading to transcription initiation is $k_{m,esc}$. To obtain a master equation for the general N-state case, we consider the probability P_m that the promoter is in the state m, at time t, given that no initiation event has taken place between 0 and t. The master equation for P_m is given by

$$\frac{dP_m}{dt} = \sum_{n=1}^{N}\left[k_{n,m}P_n - k_{m,n}P_m\right] - k_{m,esc}P_m. \qquad (1)$$

This accounts for all the ways by which the fraction of promoters in state m, which have not initiated transcription from 0 to time t, can change over the time interval (t, $t + dt$). Hence for all the promoter states the set of master equations are given by

$$\frac{d\boldsymbol{P}}{dt} = [\boldsymbol{K} - \boldsymbol{R}]\boldsymbol{P}. \qquad (2)$$

Where $\boldsymbol{P} = (P_1, P_2, \ldots, P_N)$, $\boldsymbol{K}$ and $\boldsymbol{R}$ are the transition and transcription initiation rate matrices respectively. We can solve the above set of master equations above to get,

$$\boldsymbol{P}(t) = e^{[\boldsymbol{K}-\boldsymbol{R}]t}\boldsymbol{P}(t=0). \qquad (3)$$

In order to compute the probability that two initiation events are separated by a time *t*, we compute the contribution of each promoter state to this quantity. For example, the probability that two initiation events are separated by time *t*, and the second initiation event happens while the promoter is in the state *m* is given by $k_{m,esc}P_m dt$. This is the product of the probability P_m that the promoter is in the *m*-th state at time *t* (and no other transcription events have happened before this time), and the probability $k_{m,esc}dt$ that an RNAP molecule escapes the promoter between *t* and *t* + *dt* while the promoter is in state *m*. The total probability that an initiation event happens in the time interval (*t*; *t* + *dt*) is simply the sum over all *N* states of the probability that this event occurs while the promoter is in any of these states:

$$q_1(t) = \sum_{m=1}^{N} k_{m,esc} P_m(t) \tag{4}$$

To find $P_m(t)$ from Eqn. 3 we need to specify the initial condition for $\boldsymbol{P}$ i.e. $\boldsymbol{P}(t=0)$. At time $t = 0$ there is an initiation event and that can take place from any promoter state. If we wait long enough, such that the system samples all the different promoter states, the probability that an initiation event occurs from a given promoter state should only depend on the steady state probability of being in that particular state and the rate of escape into elongation from that state. Hence it is given by

$$P_m(t=0) = \frac{k_{m,esc} P_m^{\,ss}}{\sum_{m=1}^{N} k_{m,esc} P_m^{\,ss}} \tag{5}$$

Where $P_m^{\,ss}$ is the steady state probability for the promoter to be in the *m*-th state. Steady state probabilities are given by

$$\boldsymbol{KP}=0 \tag{6}$$

Furthermore, there is the normalization condition, $\sum_{m=1}^{N} P_m^{\,ss} = 1$, and the average initiation rate is

$$I = \sum_{m=1}^{N} k_{m,esc} P_m^{\,ss} \tag{7}$$

We want to find the probability distribution $p_1(x)$ of inter-polymerase distances along the gene. Assuming a constant speed of elongation, $p_1(x)$ is essentially the imprint of the distribution of times between transcription initiation events. This allows for a very simple relationship between the inter-polymerase distance distributions and distribution of times between initiation events which is given by,

$$\begin{aligned} p_1(x)dx &= q_1(t)dt \\ p_1(x) &= q_1(t)\frac{1}{v} \end{aligned} \tag{8}$$

We use this equation to compare our theory with experiments which measure the inter-polymerase distance distributions along a gene.

Distribution of transcription initiation times for different models of initiation

We apply this general theoretical framework to obtain the probability distributions of initiation times and the inter-polymerase distance distributions (Eqns. 4 and 8), for different promoter architectures described below.

Two limiting steps model

The two limiting steps model is described in Fig. 2B. In this model, initiation happens in two sequential steps: the rate of RNAP loading on to the promoter occurs with rate k_{LOAD}, followed by RNA polymerase escaping the promoter leading to an initiation event at a rate k_{ESC} (4). The relevant master equations are in this case:

$$\begin{aligned}\frac{dP_1}{dt} &= -k_{LOAD}P_1 \\ \frac{dP_2}{dt} &= k_{LOAD}P_1 - k_{ESC}P_2.\end{aligned} \tag{9}$$

Using Eqn. 4,5 and 6, we can obtain the distribution of times between successive transcription initiation events, which is given by,

$$q_1(t) = \frac{k_{LOAD}k_{ESC}}{k_{LOAD} - k_{ESC}}\left[exp(-k_{ESC}t) - exp(-k_{LOAD}t)\right]. \tag{10}$$

Using Eqn. 8 and 10, we find the distribution of inter-RNAP initiation events to be

$$p_1(\mathrm{x}) = \frac{1}{v}\frac{k_{LOAD}k_{ESC}}{k_{LOAD} - k_{ESC}}\left[exp(-k_{ESC}(\frac{x}{v} - \frac{30}{v})) - exp(-k_{LOAD}(\frac{x}{v} - \frac{30}{v}))\right]. \tag{11}$$

where we have also taken into account the time delay $30/v$, which is the time it takes for the polymerase to clear the promoter, which is 30bp in length.

To find the initiation rate, we consider the different states of the promoter only and how they evolve in time. When an RNAP molecule initiates transcription by escaping the promoter, it goes back to state 1. Hence the master equations for the promoter states are given by

$$\begin{aligned}\frac{dP_1}{dt} &= -k_{LOAD}P_1 + k_{ESC}P_2 \\ \frac{dP_2}{dt} &= k_{LOAD}P_1 - k_{ESC}P_2.\end{aligned} \tag{12}$$

We get the steady state solution for P_1 and P_2 by setting the left sides of these two equations to zero. Then, from Eqn. 7 the transcription initiation rate is

$$\begin{aligned}I' &= P_2^{ss}k_{ESC} \\ &= \frac{k_{LOAD}k_{ESC}}{k_{LOAD} + k_{ESC}}.\end{aligned} \tag{13}$$

When we include the time to clear the promoter τ_{clear}, the effective initiation rate becomes

$$I = \frac{1}{1/I' + \tau_{CLEAR}} = \frac{k_{ESC}\, k_{LOAD}}{k_{ESC} + k_{LOAD} + k_{ESC}\, k_{LOAD}\tau_{CLEAR}}. \quad (14)$$

When all the rates are very slow τ_{CLEAR} becomes negligible.

Dead-end complex model

Dead-end complex model is described in Fig. 4A. This class of model considers the formation of a long-lived non-productive initiation complex at the promoter by RNAP (5, 6). After binding the promoter at a rate k_{LOAD}, each RNAP can initiate transcription at a rate k_{ESC} or make a dead-end complex at a rate k_{DEAD}. These dead-end complexes are unproductive and are removed at a rate k_{OFF}.

The master equations for this model are given by,

$$\begin{aligned} \frac{dP_1}{dt} &= -k_{LOAD}P_1 + k_{OFF}P_3 \\ \frac{dP_2}{dt} &= k_{LOAD}P_1 - (k_{ESC} + k_{DEAD})P_2 \\ \frac{dP_3}{dt} &= k_{DEAD}P_2 - k_{OFF}P_3. \end{aligned} \quad (15)$$

Using Eqn. 4, we find the inter-initiation time distribution, which is

$$q_1(t) = k_{ESC}P_2. \quad (16)$$

The initiation rate is then obtained using Eqn. 7 and 15

$$I' = \frac{k_{OFF}\, k_{ESC}\, k_{LOAD}}{k_{OFF}\, k_{ESC} + k_{OFF}\, k_{LOAD} + k_{OFF}\, k_{DEAD} + k_{DEAD}\, k_{LOAD}}. \quad (17)$$

When we include τ_{clear}, the average initiation rate using Eqn. 15, is given by

$$I = \frac{k_{OFF}\, k_{ESC}\, k_{LOAD}}{k_{OFF}\, k_{ESC} + k_{OFF}\, k_{LOAD} + k_{OFF}\, k_{DEAD} + k_{DEAD}\, k_{OPEN} + k_{OFF}\, k_{ESC}\, k_{LOAD}\tau_{CLEAR}}. \quad (18)$$

DNA supercoiling model

This model is described in Fig. 4B where RNAPs are recruited cooperatively in the supercoiled state of the promoter. RNAP molecules are loaded on to the promoter at a rate k_{LOAD}^{LOW}. After RNAP initiates transcription, at a rate k_{ESC}, it leaves the promoter DNA in a supercoiled state and subsequent loading of RNAP polymerases at the promoter occurs at a faster rate k_{LOAD}^{HIGH}. The rate of relaxation of the supercoiled state is k_{RELAX}. Master equations for this case are:

$$\begin{aligned}
\frac{dP_1}{dt} &= -k_{LOAD}^{LOW} P_1 + k_{RELAX} P_2 \\
\frac{dP_2}{dt} &= k_{LOAD}^{LOW} P_1 - k_{ESC} P_2 \\
\frac{dP_3}{dt} &= k_{ESC}(P_2 + P_4) - (k_{RELAX} + k_{LOAD}^{HIGH}) P_3 \\
\frac{dP_4}{dt} &= k_{LOAD}^{HIGH} P_3 - k_{ESC} P_4 .
\end{aligned} \tag{19}$$

As before, using equation Eqn. 4 we get the distribution of inter-initiation times,

$$q_1(\mathrm{t}) = k_{ESC}(P_2 + \mathrm{P}_4). \tag{20}$$

Using the same procedure as above, we find the average initiation rate to be,

$$I = \frac{k_{LOAD}^{HIGH} k_{LOAD}^{LOW} k_{ESC} + k_{RELAX} k_{LOAD}^{LOW} k_{ESC}}{k_{LOAD}^{LOW} k_{ESC} + k_{LOAD}^{HIGH} k_{LOAD}^{LOW} + k_{ESC} k_{RELAX} + k_{LOAD}^{LOW} k_{RELAX} + (k_{LOAD}^{HIGH} k_{LOAD}^{LOW} k_{ESC} + k_{RELAX} k_{LOAD}^{LOW} k_{ESC}) \tau_{CLEAR}}. \tag{21}$$

ppGpp model

The ppGpp model is described in Fig. 4C. As the number of ribosomal genes increases the rate of ribosomal RNA production goes up. This triggers the production of `control molecules' (e.g. ppGpp) which then reduce the initiation rate by modulating the promoter-RNAP interactions (7). ppGpp regulates the initiation process by converting the active promoter-RNAP complexes into inactive ones. It is described by the same kinetic scheme as the dead-end complex model but with a critical difference, namely in this case it is the rate of ppGpp binding to the RNAP-promoter complex (red arrow in Fig. 4C) and not the rate of RNAP loading on to the promoter is tuned as the copy number of ribosomal genes is changed. Here we call the rate of inactivation of the RNAP-DNA complex k_{ON}. To find the probability distribution of times between successive initiation events, as before we consider the following set of master equations,

$$\begin{aligned}
\frac{dP_1}{dt} &= -k_{LOAD} P_1 + k_{OFF} P_3 \\
\frac{dP_2}{dt} &= k_{LOAD} P_1 - (k_{ESC} + k_{ON}) P_2 \\
\frac{dP_3}{dt} &= k_{ON} P_2 - k_{OFF} P_3 .
\end{aligned} \tag{22}$$

Using Eqn. 4, we find the inter-initiation time distribution, which is given by,

$$q_1(t) = k_{ESC} P_2 . \tag{23}$$

When we include τ_{clear}, our average initiation rate, using Eqn. 7 and 22 , is given by

$$I = \frac{k_{OFF} k_{ESC} k_{LOAD}}{k_{OFF} k_{ESC} + k_{OFF} k_{LOAD} + k_{OFF} k_{ON} + k_{ON} k_{OPEN} + k_{OFF} k_{ESC} k_{LOAD} \tau_{CLEAR}}. \tag{24}$$

ON-OFF model

We consider the canonical model of transcription initiation, the ON-OFF model (1, 8). In this model, the promoter switches between two states: an ON state, from which transcription initiation can occur, and an OFF state from which initiation does not occur. The two states might correspond to a free promoter and one bound by a repressor protein. The rate of transitioning from the ON to OFF state is k_{OFF} and from OFF to ON is k_{ON}. The rate of initiation from the ON state is k_{ESC}. The probability distribution of times between successive initiation events, using Eqn. 4, is given by,

$$q_1(t) = A_1 k_1 exp(-k_1 t) + A_2 k_2 exp(-k_2 t). \tag{25}$$

where the constants k_1, k_2, A_1 and A_2 are given by

$$k_{1,2} = \frac{1}{2}\left[k_{ON} + k_{OFF} + k_{ESC} \pm \sqrt{\left(k_{ON} + k_{OFF} + k_{ESC}\right)^2 - 4k_{ESC}k_{ON}} \right], \tag{26}$$

$$A_1 = \frac{r - k_2}{k_1 - k_2}, \tag{27}$$

$$A_2 = 1 - A_1. \tag{28}$$

The mean and variance of the times between successive initiation events are given by

$$Mean = \frac{k_{ON} + k_{OFF}}{k_{ON}k_{ESC}} \tag{29}$$

$$Variance = \left[\frac{(k_{ON} + k_{OFF})^2 + 2k_{OFF}k_{ESC}}{(k_{ON}k_{ESC})^2} \right] \tag{30}$$

Effect of pausing on inter-polymerase distance distributions along the ribosomal genes

Ribosomal genes are very highly transcribed and hence even very short pauses can cause traffic jam of RNAPs along the path. This can dramatically reduce the rate of ribosomal RNA production. In order to keep the transcription rate high the anti-termination system suppresses most of the ubiquitous pauses that are observed in vitro (9). Depending on the efficiency of the anti-termination system pausing can play a vital role in the ribosomal RNA production process. Klumpp et al (9) did a systematic study of how RNAP pausing affects the process of transcription of ribosomal genes. We take their model and analyze the effect of RNAP pausing on the inter-polymerase distance distributions using the Gillespie algorithm (10, 11). This allows us to find the conditions under which the inter-polymerase distance distribution remains largely unaffected by transcriptional pausing and hence provides an imprint of the initiation dynamics.

Klumpp's model of RNAP pausing along the ribosomal genes

Klumpp et al (9) modeled (see Fig. S1A) the transcription of RNAPs in dense traffic, using a stochastic cellular automaton model. In this model the RNAP molecules are represented as extended objects moving along a one-dimensional lattice. Sites of this lattice represent the individual bases of a DNA template. Each RNAP can be in either the active or paused state, independent of the state of other RNAPs. An active RNAP transcribes along the DNA by making a single nucleotide step forward at a stepping or elongation rate k_{EL}. This can only happen when the next base is not blocked by the presence of another RNAP. During elongation, the active RNAP may switch stochastically to the paused state with a rate k_{P+}. Here the assumption is that each RNAP can pause at any site along the DNA. A paused RNAP remains at the same site and it can switch back to the active state with rate k_{P-}, so that the average duration of a pause is $1/ k_{P-}$. The rate at which the RNA polymerase molecules are loaded on the DNA template is k_{ESC}. The footprint of an RNAP molecule is L. Klumpp et al extended this model to include other effects of elongation such as backtracking of pauses or effects of rho dependent terminators. For our purposes though we don't take these into considerations as for a fully efficient anti-termination system the aforementioned factors do not matter (9).

Klumpp's model cannot explain the experimental observations when the operon number is increased

In order to explain the observed effect of Nus factor deletions on the distribution of RNAP distances, Klumpp et al. assumed that it was the pause duration along the ribosomal genes that increased (9). Similarly, to see whether the model can explain the observed features of inter-polymerase distance distributions when the operon number is increased, we change one of the parameters of the model at a time. We do this in such a way that the mean number of RNAPs along the gene, matches the number obtained in experiments (12) when the number of operons is increased. After tuning each parameter, we simulate the model to get the inter-polymerase distance distributions. Then we compare model predictions with the experimental results. In particular, we compare the intra-bunch mean distances and the average elongation rates of the RNAPs along the gene, both of which are measured to be constant in experiments (9). The parameter values are shown in Table.1. We did not change the pause frequency because no matter what value is chosen one cannot get the observed mean number of RNAPs along the gene when the operon number is increased. Our analysis shows that by changing one parameter at a time one cannot explain the experimental results as shown in Table.2. Also, we think it highly unlikely that multiple parameters of the pausing model change at the same time so as to keep both the elongation rate and intra-bunch mean fixed.

Table.1

Parameters	Seven genes	Ten genes
Elongation attempt rate (k_{EL})	100 bps/sec	160 bps/sec
Pause duration ($1/k_{P\text{-}}$)	0.23 sec	3 sec
Initiation rate (k_{ESC})	2 /sec	1.35 /sec

Effect of elongation on the inter-polymerase distance distributions

The distribution of inter-polymerase distances is determined by both the initiation and elongation dynamics. Depending on the timescales of the two processes, they affect the distribution differently. If the initiation timescales are much longer than the elongation timescales the distribution of distances will remain largely unaffected by the elongation dynamics, as in the case of most of the mRNA promoters in *E.coli*. As the ribosomal genes are very highly transcribed the initiation and elongation timescales become comparable and hence the effect of elongation on the RNAP distance distributions can be significant (9). We use the model proposed by Klumpp et al.(9) to see the extent of the effect of ubiquitous pauses on the elongating RNAPs. For a fully efficient anti-termination system most of the ubiquitous pauses are suppressed except few short lived pauses (9). Using the Gillespie algorithm (10, 11), we find the inter-polymerase distance distributions for two cases. In the first case we take the parameters proposed by Klumpp et al.(9) to find the distribution. In the second case we take the same parameters for the pausing dynamics but with lower initiation and elongation attempt rates so as to match the experimental observations (12). We compare both distributions with the ones we compute above and in the main text, which are derived only from the initiation dynamics. For the higher initiation rate the coefficient of variation (CV = 1.14±0.35) which is defined as the standard deviation divided by the mean of the distribution deviates significantly from the exponential distribution. When the initiation rate is lower the distribution obtained from the simulations matches well with the exponential distribution (CV = 1.03±0.042) as expected from the initiation dynamics. The distributions are shown in Fig. S2. We conclude that the inter-polymerase distribution as observed in the experiment (12) is a result of initiation dynamics and elongation has a very little effect on it.

Table. 2

Intra-bunch mean RNAP distance in bps	Gene copy number	Experimental results	Simulation results for different tuned parameters		
			Pause duration ($1/k_{P\text{-}}$)	Elongation attempt rate (k_{EL})	Initiation rate (k_{ESC})
	7	72±5	73.4	73.4	73.4
	10	72±4	52±3	87±4	83±3
			Simulation results for different tuned parameters		

RNAP elongation rate in bps/sec	**Gene copy number**	**Experimental results**	Pause duration ($1/k_{P\text{-}}$)	Elongation attempt rate (k_{EL})	Initiation rate (k_{ESC})
	7	78	78	78	78
	10	78	50±5	120±6	90±5

Estimating the number of ribosomal genes

It is experimentally difficult to measure the exact copy number for pBR322-based plasmids (13). Hence, we make an estimate in the following way. Voulgaris et al.(12) observed in their experiments that the total number of transcripts produced in the cell remains constant even when the number of genes is increased by inserting plasmids. The initiation rate per operon goes down to keep to the total initiation rate fixed. The number of RNAPs per gene was reduced by 34 percent. Hence the number of genes should be n =7/0.66 ~ 10. Using this number, we plot the data in Fig. 4.

References

1. Sanchez, A., S. Choubey, and J. Kondev. 2013. Stochastic models of transcription: From single molecules to single cells. Methods San Diego Calif. 62: 13–25.

2. Gopich, I.V., and A. Szabo. 2006. Theory of the statistics of kinetic transitions with application to single-molecule enzyme catalysis. J. Chem. Phys. 124: 154712.

3. 2007. Preface to the third edition. In: Kampen NGV, editor. Stochastic Processes in Physics and Chemistry (Third Edition). Amsterdam: Elsevier. p. xi.

4. Choubey, S., J. Kondev, and A. Sanchez. 2015. Deciphering Transcriptional Dynamics In Vivo by Counting Nascent RNA Molecules. PLOS Comput. Biol. 11: e1004345.

5. Kubori, T., and N. Shimamoto. 1996. A Branched Pathway in the Early Stage of Transcription byEscherichia coliRNA Polymerase. J. Mol. Biol. 256: 449–457.

6. Mitarai, N., I.B. Dodd, M.T. Crooks, and K. Sneppen. 2008. The generation of promoter-mediated transcriptional noise in bacteria. PLoS Comput. Biol. 4: e1000109.

7. Maitra, A., I. Shulgina, and V.J. Hernandez. 2005. Conversion of active promoter-RNA polymerase complexes into inactive promoter bound complexes in E. coli by the transcription effector, ppGpp. Mol. Cell. 17: 817–829.

8. Golding, I., J. Paulsson, S.M. Zawilski, and E.C. Cox. 2005. Real-Time Kinetics of Gene Activity in Individual Bacteria. Cell. 123: 1025–1036.

9. Klumpp, S., and T. Hwa. 2008. Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination. Proc. Natl. Acad. Sci. 105: 18159–18164.

10. Gillespie, D.T. 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81: 2340–2361.

11. Gillespie, D.T. 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. 22: 403–434.

12. Voulgaris, J., S. French, R.L. Gourse, C. Squires, and C.L. Squires. 1999. Increased rrn gene dosage causes intermittent transcription of rRNA in Escherichia coli. J. Bacteriol. 181: 4170–4175.

13. Heinrich, T., C. Condon, T. Pfeiffer, and R.K. Hartmann. 1995. Point mutations in the leader boxA of a plasmid-encoded Escherichia coli rrnB operon cause defective antitermination in vivo. J. Bacteriol. 177: 3793–3800.

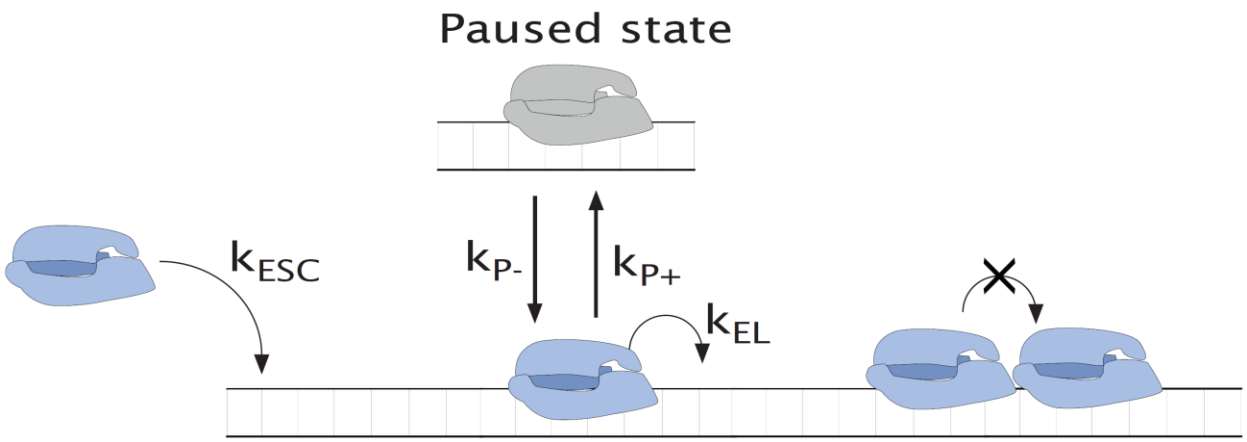

Figure S1: **Kinetic scheme of Klumpp's pausing model**: Active RNA polymerase molecules elongate along the ribosomal gene by making single-nucleotide forward steps along the DNA template. These elongation steps occur with rate k_{EL}, provided the site in front of the RNAP is not occupied by another RNAP. The rate at which RNAPs go to the paused state is k_{P+} and a paused RNAP returns to the active state with rate k_{P-}, the loading rate of RNAPs along the DNA template is k_{ESC}.

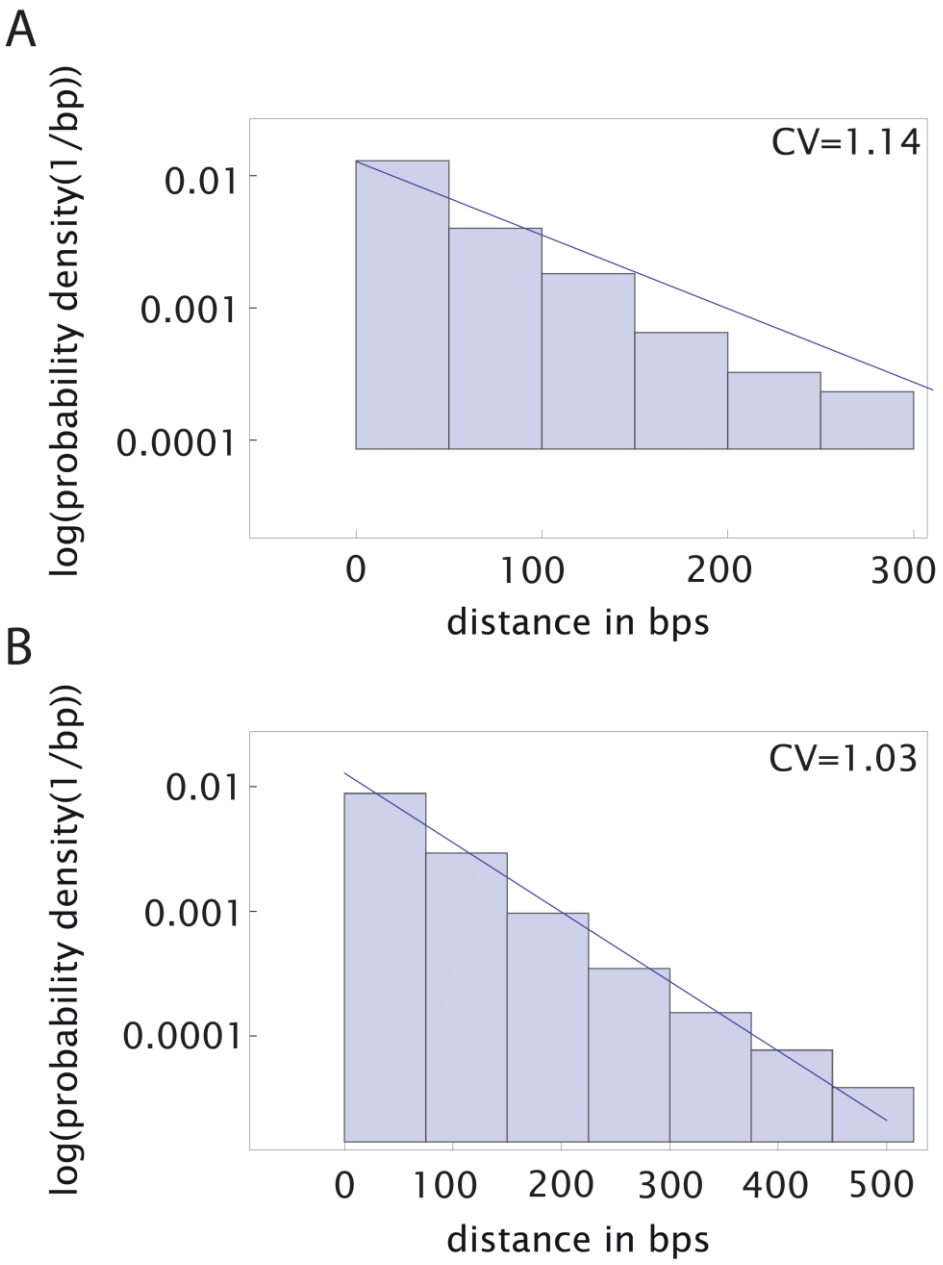

Figure S2: **Initiation versus elongation**: **A. Elongation limited regime:** Distance distribution of RNAPs along the gene gets affected substantially by the elongation process when the initiation time scale becomes comparable to elongation time scales. Using Klumpp et al.'s model (9) with the parameters they use i.e. k_{ESC} = 2/sec, k_{EL} = 100bps/sec, k_{P+} = 0.1/ sec, k_{P-}= 4.55/sec, and L = 50 bps we get the inter-polymerase distance distributions and compare it with an exponential distribution that one expects just from the initiation dynamics. It deviates quite a lot from an exponential distribution as suggested by the coefficient of variation. **B. Initiation limited regime:** With the same pausing dynamics but a lower initiation rate of k_{ESC} = 1/ sec and elongation attempt rate k_{EL} = 80/sec as observed in experiments [3], the distribution inter-polymerase distances

remains largely unaffected by the elongation dynamics. We compare it with an exponential distribution as expected from the initiation dynamics. Deviation of the distribution extracted from simulations from an exponential distribution is very small as suggested by the coefficient of variation.