Elongation dynamics shape bursty transcription and translation

Maciej Dobrzyński[a] and Frank J. Bruggeman[a,b,1]

[a]Centrum Wiskunde & Informatica, Kruislaan 413, 1098 SJ, Amsterdam, The Netherlands; and [b]Netherlands Institute for Systems Biology, Kruislaan 318, 1098 SM, Amsterdam, The Netherlands

Edited by Peter G. Wolynes, University of California at San Diego, La Jolla, CA, and approved December 19, 2008 (received for review April 10, 2008)

Cells in isogenic populations may differ substantially in their molecular make up because of the stochastic nature of molecular processes. Stochastic bursts in process activity have a great potential for generating molecular noise. They are characterized by (short) periods of high process activity followed by (long) periods of process silence causing different cells to experience activity periods varying in size, duration, and timing. We present an analytically solvable model of bursts in molecular networks, originally developed for the analysis of telecommunication networks. We define general measures for model-independent characterization of bursts (burst size, significance, and duration) from stochastic time series. Inspired by the discovery of bursts in mRNA and protein production by others, we use those indices to investigate the role of stochastic motion of motor proteins along biopolymer chains in determining burst properties. Collisions between neighboring motor proteins can attenuate bursts introduced at the initiation site on the chain. Pausing of motor proteins can give rise to bursts. We investigate how these effects are modulated by the length of the biopolymer chain and the kinetic properties of motion. We discuss the consequences of those results for transcription and translation.

bursts | interrupted Poisson process | transcriptional pausing | waiting times

The stochasticity of molecular processes contributes to heterogeneity in populations of isogenic cells. Cellular heterogeneity is manifested by differences in the copy numbers of molecules and in the timing and duration of processes. Recent advances in single-cell measurement have facilitated the quantification of stochastic phenomena (1–4) (reviewed in refs. 5 and 6). Together with models and theory much insight has been obtained into the sources of noise and how particular network designs contribute to noise suppression and amplification (7–10).

Stochasticity of gene expression has been described by distributions of macromolecules in a population of cells (1, 2). Whether averaging over a population captures the entire spectrum of molecular fluctuations a particular cell experiences over one generation, depends on magnitudes and rates of fluctuations. If these are slow but high in amplitude, the required averaging duration may extend over a generation span (11). Then, a single cell may not even be able to reach protein states accessible to other members, thus rendering cell–cell protein level distributions uninformative with respect to behavior of genetic circuits (12). In such cases, waiting times for individual birth and death events need to be monitored to assess physiological constraints on a single-cell level. This stochastic nature of waiting times will be our focus. Little analytical theory has been developed to deal with this phenomenon despite its relevance for single-cell behavior.

The waiting times in a first-order process with rate constant k follow an exponential distribution; the mean waiting time between events and its standard deviation are equal to $1/k$. The waiting time for an event is no longer exponentially distributed if it is regulated by another process. This mechanism underlies bursts in synthetic activity. The interrupted Poisson process (IPP) was introduced to study bursts in queuing and telecommunication theory (13). In an IPP, a stochastic switch modulates a process with exponentially distributed waiting times. Depending on the time scale separation between the process and the switch, multiple timescales may appear in waiting times for production events.

Bursts have received ample attention in the biophysics literature (5, 10, 14–17). These studies tend to focus predominantly on the protein number distributions, but do not analyze the distributions for waiting times in much depth. We show that such statistics are relevant for burst characterization and the molecular mechanisms giving rise to bursts.

Bursts have been experimentally observed for synthesis of mRNA and protein (3, 4, 18–22). They are characterized by rapid productions of a number of mRNA or protein molecules during short time intervals. Periods of synthetic silence occur between bursts. Bursts may give rise to significant disturbances of cellular physiology depending on burst size and the duration of synthetic silence and activity. Even though the benefit of bursts needs to be analyzed further, they could be beneficial for cells living in rapidly fluctuating environments (23). Bursts may give rise to a bimodal distribution of protein expression across cell populations (24). Thereby, 2 subpopulations could emerge having different adaptive potentials.

We apply the analytical theory of IPPs to a molecular mechanism for bursts. To identify and characterize bursts we derive 3 new indices: burst size, duration, and significance. We demonstrate how motor protein trafficking along biopolymer chains (such as mRNA polymerase and DNA polymerase along DNA, ribosomes along mRNA, and cargo-carrying dynein along microtubuli) can generate bursts depending on the length of the biopolymer and stochasticity of initiation and motion. We show that motor proteins can generate bursts by pausing or by memory of initiation bursts.

Analytical Expression of the Waiting Time Distribution

In this section, we study a small network to gain insight into burst-generating mechanisms. This will allow us to derive general indices for the characterization of burst properties. These indices will be applied to characterize biological mechanisms.

The network consists of a source switching between an inactive *OFF* and an active *ON* state according to a Poisson process (Fig. 1*A*). *OFF* and *ON* periods are defined on the level of the switch (see refs. 15 and 17 for the discussion of mechanisms giving rise to genetic switches). In the active state, production of P, e.g., mRNA or protein, occurs at exponentially distributed intervals of length $\tau_{ini} = 1/k_{ini}$. The average *ON* period lasts for $\tau_{on} = 1/k_{sw}^{-}$. Production periods are interrupted by transitions to the *OFF* state. Each rate constant corresponds to the inverse of the mean first passage time for a complex kinetic mechanism. We assume that it follows an exponential waiting time distribution.

Author contributions: M.D. and F.J.B. designed research, performed research, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: frank.bruggeman@sysbio.nl.

This article contains supporting information online at www.pnas.org/cgi/content/full/0803507106/DCSupplemental.

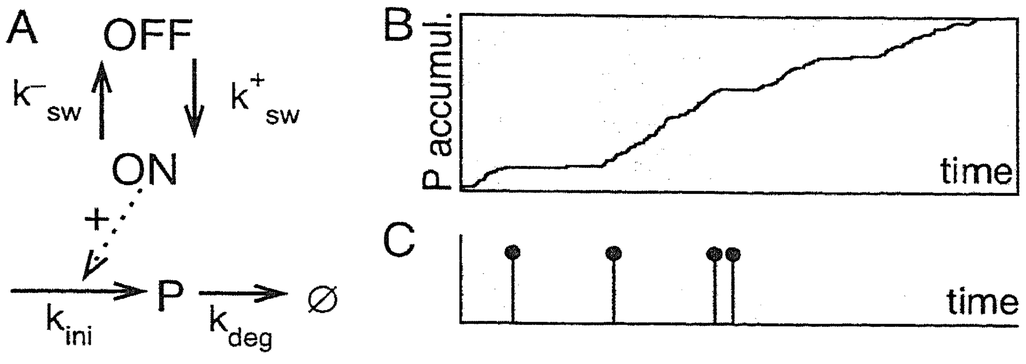

Fig. 1. Bursts generated by the minimal model. (*A*) The network consists of a switching source and a Poissonian product generator. Full arrows denote reactions. Product *P* is synthesized only in the *ON* state. k^+_{sw}, k^-_{sw}, k_{ini}, and k_{deg} denote the *ON* switching, *OFF* switching, production, and degradation rate constant, respectively. (*B*) Simulation of bursty accumulation of *P*. Two timescales correspond to uninterrupted and interrupted production events. Bars denote *OFF* (white) and *ON* (gray) state. (*C*) Waiting times for nonbursty production events (vertical lines). On average, 1 *P* is produced during the *ON* state. The resulting intervals between production events correlate weakly with *OFF* and *ON* states.

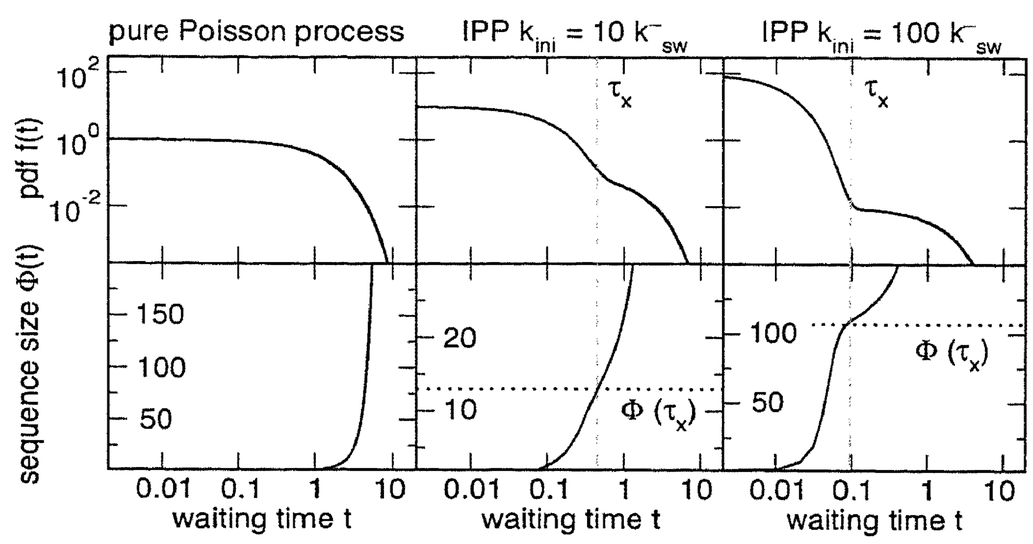

Fig. 2. Theoretical analysis of the minimal burst model (Fig. 1). Columns correspond to different parameterizations. (*Upper*) The waiting time PDF for a pure Poisson process and for 2 IPPs with different burst characteristics (Eq. **1**). The vertical dashed lines indicate the threshold of timescale separation τ_X. (*Lower*) Sequence size function (Eq. **4**). The Φ evaluated at τ_X yields the burst size β (horizontal dotted lines).

Product generation is bursty if it occurs many times during one *ON* state. In addition, the duration of the *OFF* state ($\tau_{off} = 1/k^+_{sw}$) should be longer than or comparable to the *ON* period. Under these conditions, waiting times display 2 timescales (Fig. 1*B*). Because our interest is the statistics of intervals between production, the degradation of *P* does not play a role.

The mechanism discussed here specifies an interrupted Poisson process (IPP) investigated in the field of queuing theory (13). An IPP is a Poisson process for event occurrence (*arrivals*) modulated by a random switch. In this framework a gene that switches between an *ON* and *OFF* state as function of a transcription factor would be considered the source. Arrivals would, for instance, correspond to initiations of transcription giving rise to elongating RNA polymerases. The IPP theory provides the probability density function (PDF) for waiting times between production events, $f_X(t)$, with the stochastic variable *X* as the waiting time with value *t*. It is instructive to realize that the PDF can become >1 (it is *not* a probability) and that $\int_0^\infty f_X(t)dt = 1$. The probability of an interval between consecutive events being within $[t, t + dt]$ equals $f_X(t)\,dt$. The PDF of an IPP is a weighted sum of 2 exponential distributions,

$$fx(t) = P[X \in (t, t + dt)]/dt = w_1 r_1 e^{-r_1 t} + w_2 r_2 e^{-r_2 t},$$

$$r_{1,2} = (K \pm \sqrt{K^2 - 4k_{ini}k^+_{sw}})/2, \qquad r_1 > r_2, \qquad [1]$$

where $K = k^+_{sw} + k^-_{sw} + k_{ini}$, and the weight factors $w_1 = 1 - w_2 = (k_{ini} - r_2)/(r_1 - r_2) \in (0, 1)$; derived in the supporting information (SI) *Appendix*. The PDF can reveal the presence of 2 times scales in a stochastic time series (Fig. 2 *Upper*).

The length of intervals between production events follows from the superposition of 2 independent processes: production during a single *ON* state, and periods of synthetic inactivity. The latter may result from multiple switches between *ON* and *OFF* states without producing any *P*. This is the case if the mean number of productions per *ON* state is small, i.e., $k_{ini} \approx k^-_{sw}$ (Fig. 1*C*). Accordingly, synthetic activity and silence periods (at the level of *P* production) do not strictly overlap with *ON* and *OFF* states of the switch.

Characteristic timescales of the fast and the slow process appear in Eq. **1** as rates r_1 and r_2. Weight factors w_1 and w_2 are the probabilities to observe the short period (mean duration $1/r_1$) and the long period (mean $1/r_2$) between *P* productions, respectively. For large timescale separation, i.e., when k^+_{sw} and k^-_{sw} are much smaller than k_{ini} (Fig. 2 *Right*), the rates become $r_1 \approx k_{ini}$ and $r_2 \approx k^+_{sw}$. In this regime, the *expected burst size*, β_e, equals the number of initiations per *ON* state, i.e., k_{ini}/k^-_{sw}.

The point of timescale separation we refer to as τ_X (Fig. 2). At this interval the probability to observe a waiting time resulting from the fast and the slow process is identical (two terms of Eq. **1** are equal). Two timescales are observable if $k_{ini} > k^+_{sw} - k^-_{sw}$; only then $\tau_X > 0$ (*SI Appendix*).

The mean waiting time determined from Eq. **1** corresponds to the macroscopic estimate (*SI Appendix*),

$$\langle t \rangle = \int_0^\infty t\, fx(t)dt = \frac{w_1}{r_1} + \frac{w_2}{r_2} = \tau_{ini}\left(\frac{\tau_{on}}{\tau_{on} + \tau_{off}}\right)^{-1} \qquad [2]$$

The inverse of this equation equals the mean arrival rate, $k_{ini}\langle ON\rangle$. It has the interpretation of the expected burst size divided by the duration of a single switch cycle. The noise in the waiting time is given by

$$\eta_t^2 = VAR/AVG^2 = \sigma_t^2/\langle t\rangle^2 = 1 + 2\beta_e\langle OFF\rangle^2. \qquad [3]$$

The second term expresses the deviation of IPP from a Poisson process. It is small if either the *ON* state is short-lived (β_e decreases) or silence periods are negligible.

Measures for Characterization of Bursts

The size of a burst and the burst duration are relevant burst properties. For biological applications they need to be determined on the basis of a stochastic time series as the mechanism underlying bursts is typically unknown.

The *burst size* β is defined as the mean number of production events not interrupted by a long inactivity period. Inactivity periods (interruptions) occur as often as bursts. The total number of production events divided by the number of interruptions yields the burst size. To determine the timescale of interruptions and hence their number, we define a *sequence size* function $\Phi(\vartheta)$,

$$\Phi(\vartheta) = \frac{\#\text{ total events}}{\#\text{ intervals longer than }\vartheta} = \frac{n_a}{n_a\, P[X > \vartheta]} = \frac{1}{1 - Fx\ (\vartheta)} \qquad [4]$$

where $F_X(\vartheta)$ is the cumulated distribution function (CDF), i.e., $F_X(\vartheta) = \int_0^\vartheta f_X(t)\,dt$. For a given threshold ϑ, the function yields the sequence size such that events are grouped into sequences interrupted by intervals longer than ϑ. Because of the timescale separation, there is a specific interval ϑ_b for which $\Phi(\vartheta)$ equals the burst size β. The value of ϑ_b can be determined on the basis of the functional dependence of $\Phi(\vartheta)$ as illustrated in Fig. 3.

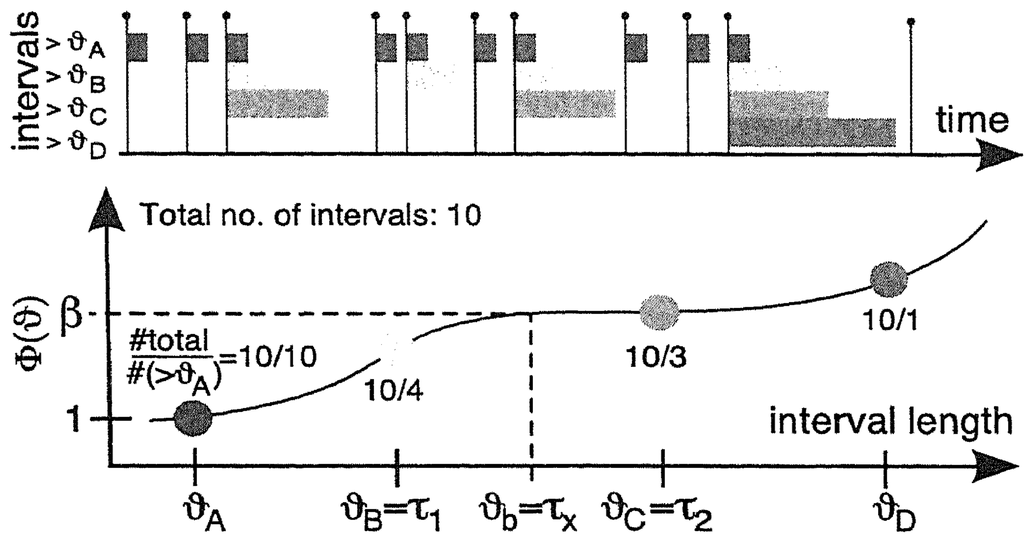

Fig. 3. Diagram of a sequence size function Φ. (*Upper*) Time series of production events. Horizontal bars denote intervals longer than thresholds ϑ_{A-D}. (*Bottom*) For a given ϑ, Φ is constructed by dividing the total number of intervals by the number of intervals longer than ϑ. Timescale separation introduces a regime where ϑ is longer than intervals within bursts but shorter than interruptions between them; a plateau appears. The point of timescale separation τ_X lies in the middle of 2 inflection points $\tau_{1,2}$ determined from the second derivative of Φ. The value of Φ at τ_X is the burst size β.

As a measure for burst size β we evaluate the sequence size function at the interval of timescale separation, τ_X, which has a straightforward interpretation for the minimal burster from Fig. 1. This point lies in the middle of 2 intervals τ_1 and τ_2 ($\tau_1 < \tau_2$); they correspond to the change of $\Phi(\vartheta)$ from convex to concave to convex as function of ϑ, respectively (*SI Appendix*).

The burst size β is greater than the expected burst size β_e because the β_e excludes the possibility of an *ON* state without a production event. Both measures are approximately equal for a large timescale separation: $\beta = \beta_e + O(\log k_{ini})$, if $k_{ini} \gg k_{sw}^+$ and $k_{sw}^- \approx k_{sw}^+$. Fig. 2 illustrates the PDF and $\Phi(\vartheta)$ for different parameter regimes of the minimal model. Bursts become more pronounced for high k_{ini} over k_{sw}^- ratios (increased timescale separation). The behavior of the sequence size function for nonexponential waiting times is explored in the *SI Appendix*. In short, a gamma-distributed waiting time for the *OFF* to *ON* transition increases the timescale separation. The applicability of the indices is not affected.

Once the burst size is known, the *duration of a burst*, τ_β, can be obtained by multiplying β by the mean waiting time within a burst ($1/r_1$ in the minimal model). Whether the interval is part of a burst can be deduced by using the threshold of timescale separation, τ_X (determined on the basis of the waiting time PDF). In addition to burst size, burst significance is important. Bursts lose significance if the interruption period becomes comparable to intervals within a burst. To quantify this we introduce a dimensionless *significance coefficient* $\xi = 1 - \tau_1/\tau_2$, $\xi \in (0, 1)$.

We use all 3 measures, β, τ_β, and ξ, to analyze stochastic time series for more complex schemes. The advantage is that the indices are mechanism-independent. In addition to their unbiased nature, they have a clear mechanistic interpretation for the minimal burster. This property facilitates interpretation of yet unidentified mechanisms giving rise to bursts.

Motor-Protein Traffic Jams along Biopolymer Chains

Bursts have been observed experimentally for single-cell synthesis of mRNA and protein (3, 4, 18–20) (review, ref. 6). For such cases, free-energy-driven motion of a catalytic motor protein along a biopolymer template is required. Here, we investigate the role of the stochasticity in the initiation of motion and in the motion itself for the observation of bursts at the end of the chain. We will consider different lengths of the polymer and kinetics of initiation and transport.

Fig. 4 shows a canonical 1D macromolecular trafficking model. It contains the switching source, as described in the previous

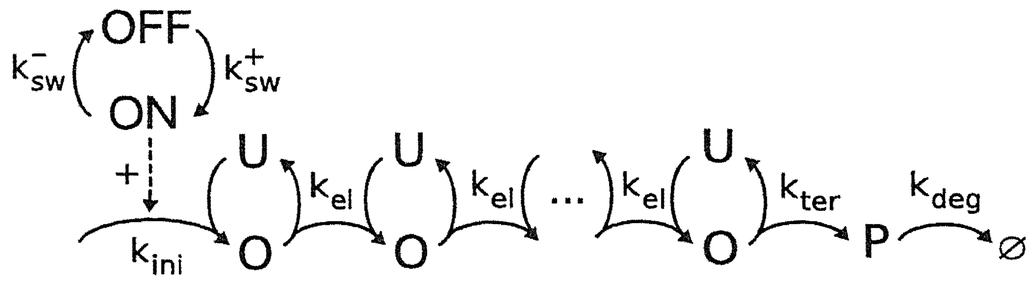

Fig. 4. Canonical model of macromolecular trafficking along a biopolymer. In the *ON* state of the switch, proteins initiate elongation with a rate constant k_{ini}. Elongation occurs with a rate constant k_{el}. "O" and "U" denote occupied and unoccupied state of the site, respectively. Motors leave the chain with a rate constant k_{ter} and accumulate a product *P*. The product is degraded with a rate constant k_{deg}.

section, followed by sites on the polymer. Each site can be occupied by a single motor, moving forward only, with the elongation rate constant k_{el} (sites per time). Evidently, motors cannot pass each other and shall collide. We also consider the motor occupying more than one site (*SI Appendix*). Previously we discussed conditions for bursts to emerge at the start of the polymer. Whether bursts are preserved at the end of the chain depends on the characteristics of the stochastic motion.

In Fig. 5*A* we consider a polymer of length 100 sites, preceded by a *bursty* switch with 100 initiations per *ON* state, on average. We plot the mean occupancy of sites at the beginning and at the end of the polymer as function of a dimensionless ratio, k_{el}/k_{ini}. If a motor protein travels many sites between initiations during a single *ON* state ($k_{el}/k_{ini} \gg 1$), its progression is not hampered by collisions. A traffic jam arises at the beginning of the chain if the number of traversed sites during consecutive initiations is small ($k_{el}/k_{ini} \ll 1$). This results in high mean site occupancy at the initial segment of the polymer. Congestion weakens at the end of the polymer because motors have fewer partners ahead of them. As shown in Fig. 5*B*, the net occupancy gradient along the chain increases with the polymer length.

If collisions are significant, the timescale separation generated at the initiation stage is disrupted. The frequency of *P* production becomes exclusively determined by the motor protein progression at the end of the chain. The effect intensifies as the length of the polymer increases as illustrated in Fig. 6*A*.

Because of collisions, the rate at which motors leave the polymer becomes smaller than k_{ini}. The mean interval between productions (the inverse of the macroscopic flux) becomes longer for longer chains (Fig. 6*B*). Additionally, the standard deviation becomes comparable to the mean waiting time. The process becomes exponentially distributed and the memory of the state of the switch is lost entirely.

Motor protein collisions disrupt inactivity periods of initiation bursts. As a result, intervals between production events become comparable. In this regime, the burst size β measured at the end of a long polymer is seemingly larger than for a short one (Fig. 6*C*).

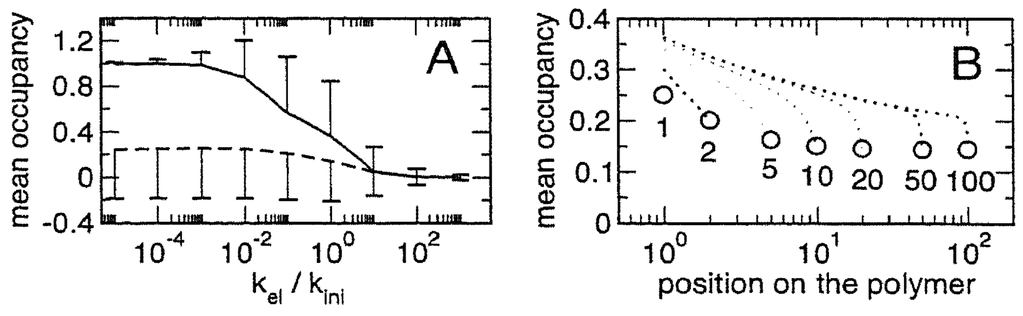

Fig. 5. Mean site occupancy. (*A*) At the beginning (solid line) and at the end (dashed line) of a 100-site polymer ($k_{sw}^+ = k_{sw}^- = 1$ [$1/T$], $k_{ini} = 100\ k_{sw}^-$); error bars, standard deviation; (*B*) along 1- to 100-site polymers (dotted), numbers indicate the length, circles mark mean occupancy at the polymer's end; parameters as in *A*, and $k_{el} = k_{ini}$.

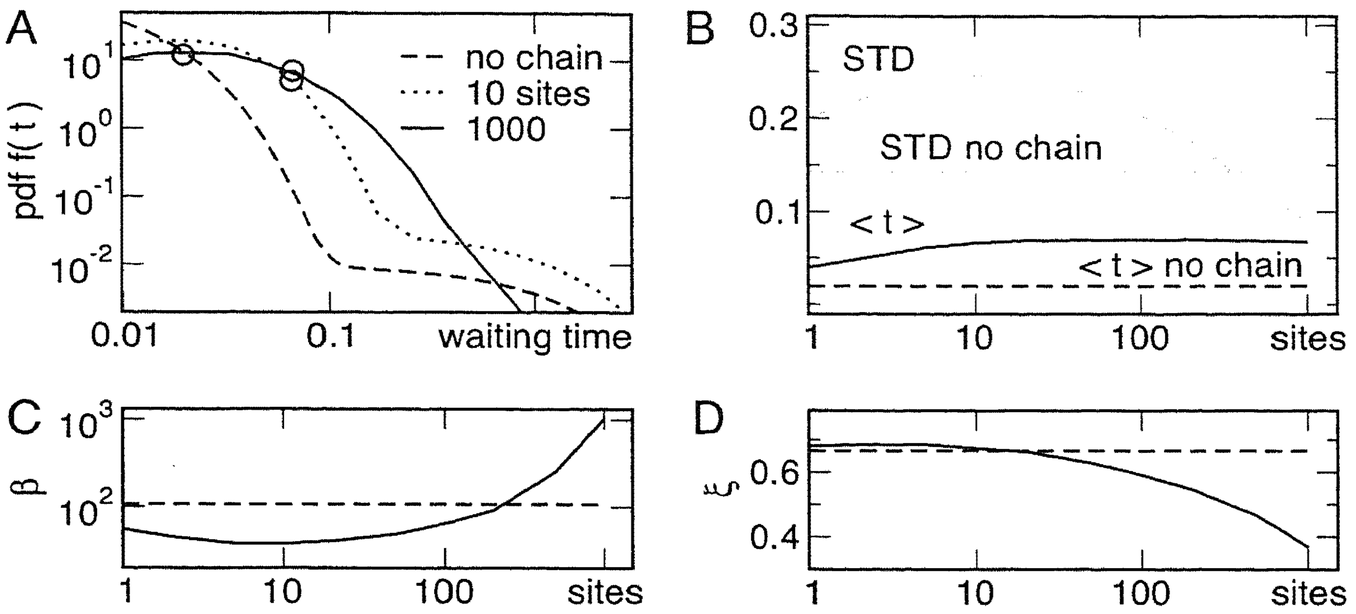

Fig. 6. Analysis of the canonical model of macromolecular trafficking. Gillespie simulations of at least 1E-05 events, $k_{sw}^{+} = k_{sw}^{-} = 1$ [1/*T*], $k_{el} = k_{ini} = 100\ k_{sw}^{-}$. (*A–D*) The dashed line marks the minimal burst model (no elongation). (*A*) The waiting time PDF for 10- and 1,000-site polymers. Circles, the mean waiting time. (*B–D*) The mean waiting time and its standard deviation (*B*), burst size (*C*), and burst significance (*D*) for different chain lengths.

Therefore, it is necessary to aid the measurement of bursts with the significance index, ξ. As the length of the polymer increases, the significance of bursts diminishes (Fig. 6*D*). Below, we introduce a mechanism that can recover bursts at the output even if no bursts occur at the input.

Pausing of Motor Proteins Can Generate Bursts

We will consider a polymer without a switching source that modulates initiation. Initiation takes place at a fixed rate constant k_{ini}. Such a model has only one timescale, thereby no bursts at initiation can occur. We consider motor protein pausing along the chain as a potential burst-generating mechanism. At every site, a motor can switch at a rate k_p^+ to a paused state that lasts $1/k_p^-$ (Fig. 7). This mechanism is known to occur for RNA polymerase (25–29) and ribosomes (30–34).

Pausing of a single motor causes congestion due to its collision with consecutive motors during its dwell time. This allows for the buildup of a burst packet. The packet can survive until the end of the chain only if the pausing frequency is low for a given chain length L. If this is not the case, there is a high probability that proteins within the potential burst will also pause and thus divide the packet (Fig. 8*A*, curves for $k_p^+ = 100$ [1/*T*]). Another requirement for bursting concerns the lifetime of the paused state. If too short, compared with the initiation rate k_{ini} and the elongation rate k_{el}, the consecutive proteins will not catch up (Fig. 8*A*, curves for $k_p^- = 100$ [1/*T*]). Timescale separation in the waiting times at the end of the chain, and hence bursts, appear only when motors do not pause too frequently during the elongation, and when the paused state is sufficiently long-lived (Fig. 8*A*, solid curve for $k_p^+ = k_p^- = 1$ [1/*T*]).

The length of the biopolymer chain affects the statistics of bursts as well. The addition of sites increases the probability that a single motor pauses a number of times during its progression. Thereby, such a motor destroys the burst it was part of. The effect is equivalent to an increase in k_p^+ at a fixed L. As a result, the waiting times lack the short timescale originating from frequent product initiation. Instead, they are dominated by the lifetime of the paused state. The mean waiting time $\langle t \rangle$ and its standard deviation increase with increasing L, and the timescale separation becomes less pronounced (Fig. 8 *B* and *C*). Because burst size β decreases, bursts tend to disappear for longer chains. They can always be recovered by decreasing the probability of a single motor protein to pause many times during its progression. As an illustration we will change the pausing rate for the longest chain considered, $L = 500$. If we set k_p^+ 10 times smaller than the value used for $L = 50$, the timescale separation is recovered and β increases (Fig. 8 *B* and *D*, dashed line). This indicates that pausing may prevent or promote bursts depending on the properties of the biopolymer.

Aggregative Behavior of Multiple Burst Generators

Statistics of bursts change when they arise from the simultaneous activity of a number of independent burst-generating mechanisms. In biological terms, this superposition may describe the transcription of an mRNA from independent copies of a gene or the translation of protein from a number of mRNAs. Here, we focus on the extension of the simple model of bursts to a superposition of many independent interrupted Poisson processes.

We choose parameters such that a single burster initiates 1 product per *ON* state ($k_{sw}^{+} = 0.1$ [1/*T*], $k_{sw}^{-} = 1$ [1/*T*], $k_{ini} = k_{sw}^{-}$). The resulting burst size is small, $\beta \approx 2.3$. The waiting time PDF almost completely loses its double-exponential character for >8 sources (Fig. 9*A*). Although the burst size increases for many sources (Fig. 9*B*), their significance ξ diminishes until the sequence size function Φ becomes always convex and ξ can no longer be evaluated (Fig. 9*B Inset*). This indicates that the clustering of events into bursts no longer occurs; waiting times follow a single-exponential distribution. Analytical results for the pooled mechanism, the PDF, and the Poissonian behavior in the limit of many independent IPPs, can be found in the *SI Appendix*. These results are particularly insightful for the conditions for bursts in the protein level. A burst of a few mRNA transcripts does not necessarily cause a protein burst. It depends on temporal correlation between translation that occurs independently on each transcript.

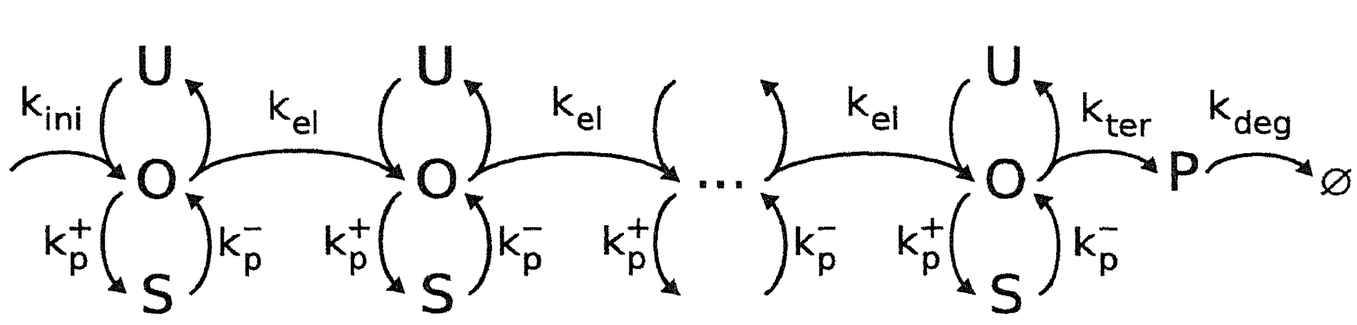

Fig. 7. Model of macromolecular trafficking with pausing motor proteins. Proteins initiate motion with a fixed rate constant k_{ini}. A site can be unoccupied (U), occupied (O) by a motor, or occupied by a motor in a paused state (S). Motors can pause at a rate k_p^+ at every site. The lifetime of the paused state is $1/k_p^-$. Other parameters are the same as in Fig. 4.

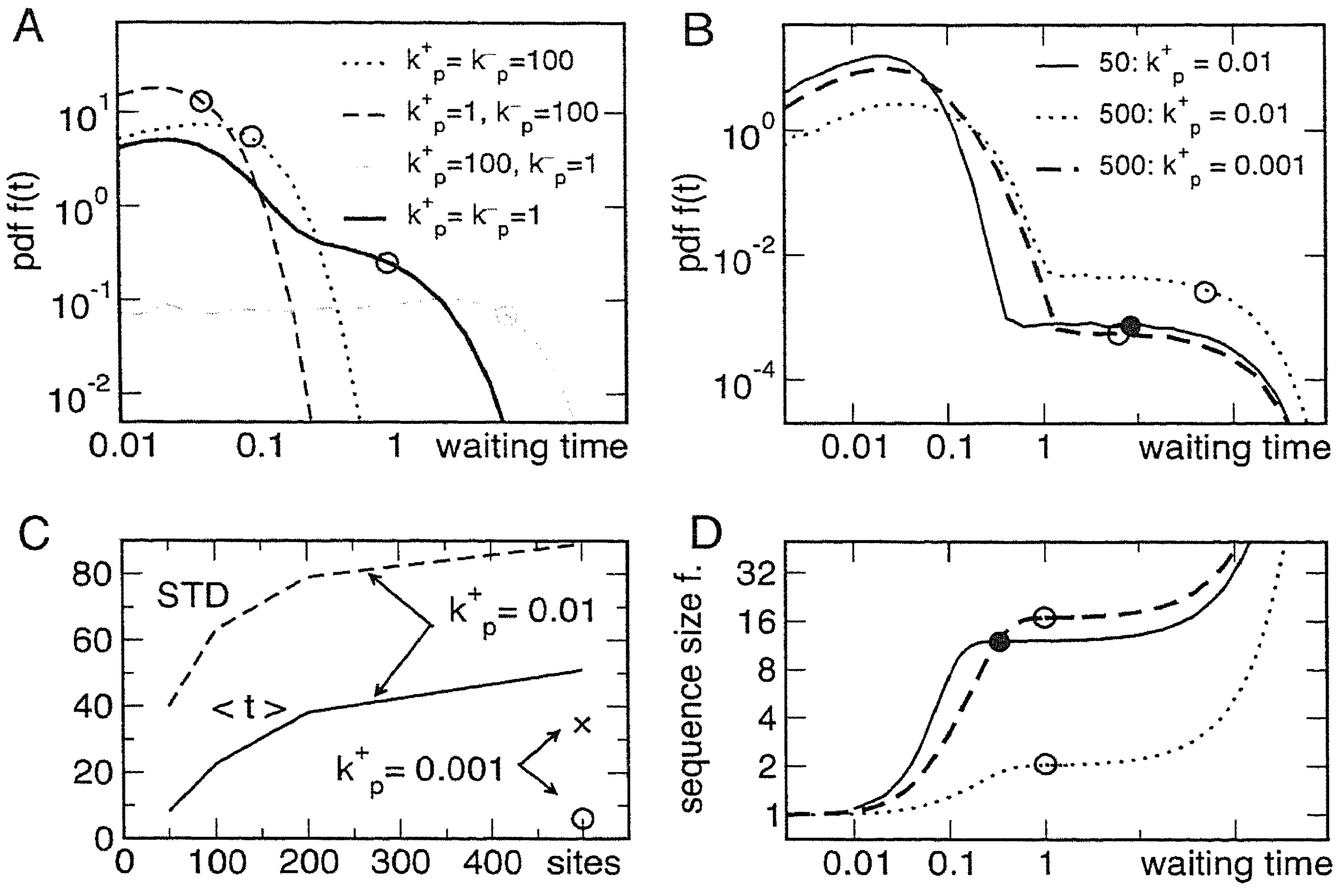

Fig. 8. The effect of chain length and pausing parameters on waiting time statistics for the model with pausing. Data obtained from Gillespie simulations of 1E-06 events. Common parameters, $k_{ini} = 100$ [$1/T$], $k_{el} = k_{ter} = k_{ini}$. (*A*) The appearance of the timescale separation due to the pausing of proteins. Chain length, 100 sites. Bursts arise when (*i*) k_p^+ allows for only few pauses during the elongation and (*ii*) $1/k_p^-$ is long enough for many initiations to occur (solid line). Circles, the mean waiting time. (*B*) Waiting time PDF for chain lengths 50 and 500 sites. The lifetime of the paused state is fixed: $1/k_p^- = 100$ [T]. At a pausing rate constant $k_p^+ = 0.01$ [$1/T$], the increase in the chain length increases the probability of multiple pauses during the progression: bursts disappear (dotted line). Reduction of k_p^+ to 0.001 [$1/T$] recovers bursts (dashed line). Circles, $\langle t \rangle$. (*C*) The mean waiting time and its standard deviation as function of chain length. Lines, $k_p^+ = k_p^- = 0.01$ [$1/T$]. Symbols at $L = 500$ correspond to the dashed line in *B*: $k_p^+ = 0.001$ [$1/T$]. (*D*) Sequence size function for polymers as in *B*. Circles, β.

Discussion

For most applications the exact burst-generating mechanism is not known or too complex to handle analytically. To overcome this problem we defined 3 measures for the characterization of bursts: *burst size*, *duration*, and *significance*. We stress the usefulness of the significance measure, because large bursts can arise at a negligible timescale separation. The indices have a transparent interpretation for the system in Fig. 1 and allow for model-independent analysis of more complicated mechanisms. Additionally, we offer a rigorous method to obtain the indices from stochastic time series. We applied those measures to investigate the influence of the stochastic motion of motor proteins along a biopolymer on bursts of product release at the end of the chain, e.g., of protein, cargo-vesicles, or mRNA. Our study was inspired by the experimental discovery of bursts in transcription and translation (3, 4, 18–20). We found that bursts at the input of the chain tend to be smoothed out by longer chains because of the congestion of motor proteins. Because of collisions, timescales within and between bursts become comparable causing the burst size to be larger but of less significance. Hence, the stationary output flux decreases with the length of the chain. At a fixed initiation rate, bursts can emerge because of the pausing of motor proteins.

We discussed 2 mechanisms that could give rise to bursts in production. Bursty transcription initiation was the first one. The second considered motor protein pausing as a source of bursts. How can these 2 mechanisms be distinguished experimentally and how can existing data be interpreted in this light? We will consider this in more detail for transcription. Promising mechanisms for initiation-induced bursts in mRNA production are genes controlled by strong repressors occasionally leaving the promotor to allow a few RNA polymerases to initiate transcription as suggested in refs. 3 and 4. On average, every τ_{ini} minutes an RNA polymerase initiates elongation if the gene is in the *ON* state, i.e., in the presence of an activating transcription factor or in the absence of a repressor. During this time, polymerase traverses $k_{el}\ \tau_{ini}$ nucleotides. If no significant congestion occurs along the DNA, the mean waiting time $\langle t \rangle$ for polymerases at the end of the chain is proportional to τ_{ini} (Eq. **2**); the timescale of initiation bursts (if present) is preserved at the end of the chain. If motors collide during their progression this relationship is destroyed. Fig. 10 illustrates this for switch param-

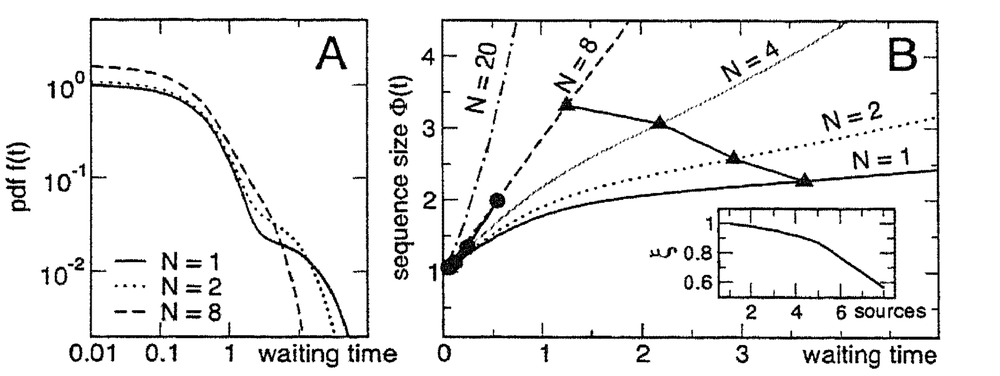

Fig. 9. Superposition of independent bursters ($k^+_{sw} = 0.1$ [$1/T$], $k^-_{sw} = 1$ [$1/T$], $k_{ini} = k^-_{sw}$). (*A*) Analytical waiting time PDF as a function of the number of burst sources. The curve becomes almost exponential for 8 sources. (*B*) Analytical sequence size function for 1–20 sources. The 2 roots of its second derivative (filled circles and triangles) vanish for >8 sources. (*Inset*) Burst significance as function of the number of sources.

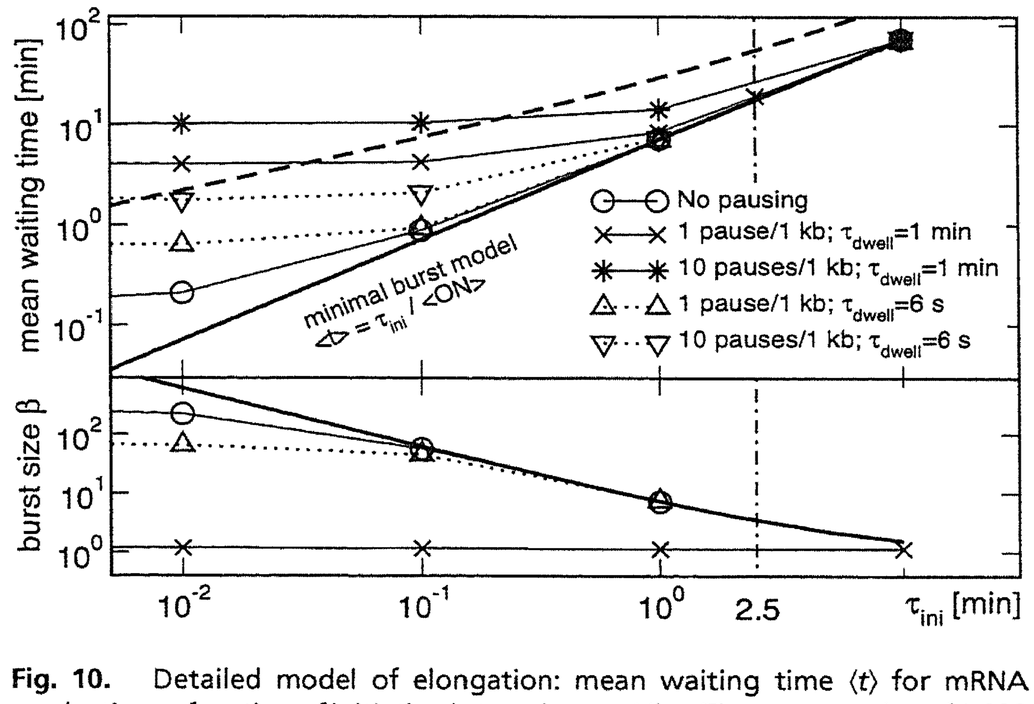

Fig. 10. Detailed model of elongation: mean waiting time $\langle t \rangle$ for mRNA production as function of initiation intervals $\tau_{ini} = 1/k_{ini}$. The gene consists of 1,000 nt, RNA polymerase occupies 50 nt (35). Parameters of the initiation switch are $\tau_{on} = 6$ and $\tau_{off} = 37$ min (3). Elongation occurs at 50 nt/s (28). Means were obtained from Gillespie simulations of at least 1E-04 events. Solid line without symbols denotes the mean, and the dashed line denotes one standard deviation plus the mean for the minimal burst model (no elongation). The vertical line indicates the mean waiting time of mRNA in the experiment of Golding *et al.* (3). Deviation from the solid line results from collisions of RNA polymerases.

eters corresponding to measurements by Golding *et al.* (3); $\tau_{on} = 6$, $\tau_{off} = 37$ min. This figure displays the dependency of $\langle t \rangle$ on the initiation time τ_{ini} for the minimal, canonical trafficking, and detailed model (with initiation switch and pausing). Assuming that the experimental system studied by Golding *et al.* (3) is in the regime where collisions do not disrupt the proportionality between $\langle t \rangle$ and τ_{ini}, their measured waiting time within a burst corresponds to τ_{ini} = 2.5 min. For these parameters, Fig. 10 indicates that most of the control on the waiting time is exerted by the initiation; pausing properties do not affect $\langle t \rangle$. The calculated burst size (3.6) is close to the experimental result ($\approx$2.2).

How would pausing of RNA polymerases and altered initiation rates (different genes) change this picture? In Fig. 10, we investigate this question by using realistic pausing parameters. The panel for burst size shows that initiation bursts are reduced by pausing. Spontaneous collisions and hence deviation of $\langle t \rangle$ from linear dependence (Eq. **3**) occur only at high initiation rates. The discrepancy is larger if pausing is considered. Therefore, pausing is a more plausible mechanism for such deviations in real biological systems where τ_{ini} exceeds 0.1 min. This makes pausing a potent target for regulation, e.g., by NusA and NusG, in accordance to recent experiments (26–29). Voliotis *et al.* (35) have postulated collision-induced bursts caused by backtracking of RNA polymerases. Fig. 10 offers a convenient method to determine whether collisions induced by pausing or spontaneous are additional controlling processes besides initiation.

Which genes are likely to generate bursts? Strong repressors could induce bursts according to bursty initiation as described above. In experimental studies this mechanism yielded a small burst size (3, 4). If the lifetime of mRNAs and proteins is shorter than the *OFF* period, transient bursts ("puffs") are produced. Savageau's demand principle (36) predicts that infrequently used genes are regulated by repressors to prevent them from accumulating mutations. Such systems should be susceptible to bursts, e.g., repressor-regulated operons of prokaryotic signaling networks.

Highly activated genes under the control of an activating transcription factor can give rise to bursts by the pausing mechanism (Fig. 8). Such systems are predominantly in the *ON* state (e.g., by way of enhancer dependence) and have an approximately constant and high initiation rate. This should make them prone to bursts induced by pausing. Such a mechanism has not been experimentally observed so far. *Escherichia coli*'s rrn genes (coding for ribosomal RNA) would be likely candidates, for they are among the genes with the highest expression activity (37). In mammalian systems, poised and paused polymerases occur often and could underlie the experimentally observed bursts (20, 38, 39).

Typically, products are not synthesized in a single macromolecular process. We showed that if bursts are generated independently, the resultant burst tends to lose significance with an increasing number of sources. Hence, mRNA bursts are more likely to occur than protein bursts, because protein is typically produced from a few transcripts simultaneously. Depending on the significance and size of the burst generated by a single catalytic process, this rules out significance of bursts in, for instance, metabolism where the copy number of catalytic proteins is thousands.

Bursts are a powerful mechanism to generate cellular heterogeneity. Key processes such as transcription and translation are particularly prone to generate bursts. How biological systems manage to function reliably in the presence of bursts, whether they actively suppress them or control their characteristics, remains to be experimentally shown. Exact burst properties and their functional consequences depend on the relative timescales of initiation, elongation, and termination processes. The analytical theory we presented gives insight into generic burst properties. The proposed burst indices allow for quantitative studies of specific systems and their comparison.

ACKNOWLEDGMENTS. We thank J. G. Blom, J. Vidal Rodríguez, H. Westerhoff, and P. R. ten Wolde for helpful input and the anonymous reviewers for their very insightful comments. This work was supported by Netherlands Organisation for Scientific Research (NWO) Computational Life Sciences Project NWO-CLS-635.100.007 and by the Netherlands Institute for Systems Biology and NWO (F.J.B.).

1. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.
2. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. *Nat Genet* 31:69–73.
3. Golding I, Paulsson J, Zawilski SM, Cox EC (2005) Real-time kinetics of gene activity in individual bacteria. *Cell* 123:1025–1036.
4. Yu J, Xiao J, Ren X, Lao K, Xie XS (2006) Probing gene expression in live cells, one protein molecule at a time. *Science* 311:1600–1603.
5. Kærn M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: From theories to phenotypes. *Nat Rev Genet* 6:451–464.
6. Kaufmann BB, van Oudenaarden A (2007) Stochastic gene expression: From single molecules to the proteome. *Curr Opin Genet Dev* 17:107–112.
7. Elf J, Ehrenberg M (2003) Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Res* 13:2475–2484.
8. Simpson ML, Cox CD, Sayler GS (2003) Frequency domain analysis of noise in autoregulated gene circuits. *Proc Natl Acad Sci USA* 100:4551–4556.
9. Paulsson J (2004) Summing up the noise in gene networks. *Nature* 427:415–418.
10. Pedraza JM, Paulsson J (2008) Effects of molecular memory and bursting on fluctuations in gene expression. *Science* 319:339–343.
11. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) Gene regulation at the single-cell level. *Science* 307:1962–1965.
12. Sigal A, *et al.* (2006) Variability and memory of protein levels in human cells. *Nature* 444:643–646.
13. Kuczura A (1973) The interrupted Poisson process as an overflow process. *Bell Syst Tech J* 52:437–448.
14. Colquhoun D, Hawkes AG (1982) On the stochastic properties of bursts of single ion channel openings and of clusters of bursts. *Philos Trans R Soc Lond B Biol Sci* 300:1–59.
15. Walczak AM, Sasai M, Wolynes PG (2005) Self-consistent proteomic field theory of stochastic gene switches. *Biophys J* 88:828–850.
16. van Zon JS, Morelli MJ, Tănase-Nicola S, ten Wolde PR (2006) Diffusion of transcription factors can drastically enhance the noise in gene expression. *Biophys J* 91:4350–4367.
17. Mitarai N, Dodd IB, Crooks MT, Sneppen K (2008) The generation of promoter-mediated transcriptional noise in bacteria. *PLoS Comput Biol* 4:e1000109.
18. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* 4:1707–1719.
19. Cai L, Friedman N, Xie XS (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature* 440:358–362.
20. Chubb JR, Trcek T, Shenoy SM, Singer RH (2006) Transcriptional pulsing of a developmental gene. *Curr Biol* 10:1018–1025.
21. Newman JRS, *et al.* (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* 441:840–846.
22. Bar-Even A, *et al.* (2006) Noise in protein expression scales with natural protein abundance. *Nat Genet* 38:636–643.
23. Acar M, Mettetal JT, van Oudenaarden A (2008) Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet* 40:471–475.
24. Friedman N, Cai L, Xie S (2006) Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys Rev Lett* 97:168302.
25. Yin H, Artsimovitch I, Landick R, Gelles J (1999) Nonequilibrium mechanism of transcription termination from observations of single RNA polymerase molecules. *Proc Natl Acad Sci USA* 96:13124–13129.
26. Artsimovitch I, Landick R (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci USA* 97:7090–7095.
27. Bar-Nahum G, *et al.* (2005) A ratchet mechanism of transcription elongation and its control. *Cell* 120:183–193.
28. Greive SJ, von Hippel PH (2005) Thinking quantitatively about transcriptional. regulation. *Nat Rev* 6:221–232.
29. Hatoum A, Roberts J (2008) Prevalence of RNA polymerase stalling at Escherichia coli promoters after open complex formation. *Mol Microbiol* 68:17–28.
30. Hayes CS, Sauer RT (2003) Cleavage of the A site mRNA codon during ribosome pausing provides a mechanism for translational quality control. *Mol Cell* 12:903–911.
31. Sunohara T, Jojima K, Tagami H, Inada T, Aiba H (2004) Ribosome stalling during translation elongation induces cleavage of mRNA being translated in Escherichia coli. *J Biol Chem* 279:15368–15375.
32. Buchan JR, Stansfield I (2007) Halting a cellular production line: Responses to ribosomal pausing during translation. *Biol Cell* 99:475–487.
33. Galburt EA, *et al.* (2007) Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* 446:820–823.
34. Wen J, *et al.* (2008) Following translation by single ribosomes one codon at a time. *Nature* 452:598–603.
35. Voliotis M, Cohen N, Molina-Paris C, Liverpool TB (2008) Fluctuations, pauses, and backtracking in DNA transcription. *Biophys J* 94:334–348.
36. Savageau MA (1998) Demand theory of gene regulation. I. Quantitative development of the theory. *Genetics* 149:1665–1676.
37. Bremer H, Dennis P, Ehrenberg M (2003) Free RNA polymerase and modeling global transcription in Escherichia coli. *Biochimie* 85:597–609.
38. Wolf DA, Strobl LJ, Pullner A, Eick D (1995) Variable pause positions of RNA polymerase II lie proximal to the c-myc promoter irrespective of transcriptional activity. *Nucleic Acids Res* 23:3373–3379.
39. Price DH (2008) Poised polymerases: On your mark... Get set... Go! *Mol Cell* 30:7–10.