Effects of Molecular Memory and Bursting on Fluctuations in Gene Expression

Juan M. Pedraza, *et al.*

Science **319**, 339 (2008);
DOI: 10.1126/science.1144331

generation? Its clonal expansion, coupled with its persistence since birth, is highly suggestive of self-renewal. Its clonal relation to more differentiated cell types both in the healthy and in the leukemic twins implies differentiation potential. The balance of evidence thus favors the notion that this cell may itself function as a preleukemic stem cell. This proposal is supported by our xenograft modeling studies, which further suggest that *TEL-AML1* may be sufficient to generate this population of preleukemic stem cells.

Our results suggest that a hierarchical structure, which has been demonstrated in frank leukemia (*2*, *4*), is also a feature of "early" or preleukemic populations. Understanding the nature of the preleukemic hierarchy is fundamental to understanding the function of the first-hit mutation and how it predisposes to leukemic transformation. Our studies therefore have implications for disease etiology, and the xenograft model presented may provide a tool for examining the biological role of genetic alterations that cooperate with the *TEL-AML1* fusion gene. Our studies may also be relevant to cancer therapy where specific targeting of tumor propagating cells may be desirable. The observation that children in lengthy remission can relapse late with a novel leukemic clone (*21*), but which nonetheless appears to derive from the identical preleukemic clone that initiated the disease at presentation, suggests that the preleukemic stem cell compartment may persist even when the cells propagating the overt leukemia have been effectively eradicated.

References and Notes

1. J. C. Wang, J. E. Dick, *Trends Cell Biol.* **15**, 494 (2005).
2. M. Al-Hajj, M. S. Wicha, A. Benito-Hernandez, S. J. Morrison, M. F. Clarke, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3983 (2003).
3. D. Bonnet, J. E. Dick, *Nat. Med.* **3**, 730 (1997).
4. C. F. Kim *et al.*, *Cell* **121**, 823 (2005).
5. S. K. Singh *et al.*, *Nature* **432**, 396 (2004).
6. C. A. O'Brien, A. Pollett, S. Gallinger, J. E. Dick, *Nature* **445**, 106 (2007).
7. P. N. Kelly, A. Dakic, J. M. Adams, S. L. Nutt, A. Strasser, *Science* **317**, 337 (2007).
8. S. A. Shurtleff *et al.*, *Leukemia* **9**, 1985 (1995).
9. M. F. Greaves *et al.*, *Blood* **102**, 2321 (2003).
10. M. F. Greaves, J. Wiemels, *Nat. Rev. Cancer* **3**, 639 (2003).
11. H. Mori *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8242 (2002).
12. C. G. Mullighan *et al.*, *Nature* **446**, 758 (2007).
13. T. R. Golub *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4917 (1995).
14. C. V. Cox *et al.*, *Blood* **104**, 2919 (2004).
15. A. A. George *et al.*, *Blood* **97**, 3925 (2001).
16. M. Hotfilder *et al.*, *Blood* **100**, 640 (2002).
17. A. Castor *et al.*, *Nat. Med.* **11**, 630 (2005).
18. Although the $CD34^{+}CD38^{-}CD19^{+}$ population seen in cALL was originally described as $CD38^{-}$, we have referred to it as $CD38^{-/low}$ to reflect the observation that CD38 expression in human cells is seen as a continuum rather than existing in discrete populations.
19. Materials and methods are available as supporting material on *Science* Online.
20. M. F. Greaves *et al.*, *Blood* **61**, 628 (1983).
21. J. Zuna *et al.*, *Clin. Cancer Res.* **10**, 5355 (2004).

22. We thank I. Titley, C. Waugh, and C. Hetherington for technical assistance. This work was supported by a Specialist Programme Grant from the Leukemia Research Fund, the Medical Research Council, and EuroCSC: Targeting Cancer Stem Cells (CSCs) for Therapy.

18 September 2007; accepted 12 December 2007
10.1126/science.1150648

Effects of Molecular Memory and Bursting on Fluctuations in Gene Expression

Juan M. Pedraza[1] and Johan Paulsson[1,2]*

Many cellular components are present in such low numbers per cell that random births and deaths of individual molecules can cause substantial "noise" in concentrations. But biochemical events do not necessarily occur in single steps of individual molecules. Some processes are greatly randomized when synthesis or degradation occurs in large bursts of many molecules during a short time interval. Conversely, each birth or death of a macromolecule could involve several small steps, creating a memory between individual events. We present a generalized theory for stochastic gene expression, formulating the variance in protein abundance in terms of the randomness of the individual gene expression events. We show that common types of molecular mechanisms can produce gestation and senescence periods that reduce noise without requiring higher abundances, shorter lifetimes, or any concentration-dependent control loops. We also show that most single-cell experimental methods cannot distinguish between qualitatively different stochastic principles, although this in turn makes such methods better suited for identifying which components introduce fluctuations. Characterizing the random events that give rise to noise in concentrations instead requires dynamic measurements with single-molecule resolution.

Gene expression is a complex stochastic process, involving numerous components and reaction steps and spanning several time and concentration scales (*1*–*8*). This complexity has motivated two very different views on fluctuations in protein levels. A widespread notion in biology suggests that random variation is restrained because each individual chemical step only contributes marginally to the total, just as rolling more dice reduces relative fluctuations in the sum of the outcomes. Physics-inspired theory has instead emphasized that all the underlying processes could propagate rather than average out fluctuations, as when one die roll is used to determine how many dice to roll next. Both scenarios are plausible: The mapping from details on finer scales into effective events on coarser scales—coarse graining—depends on molecular mechanisms that support a wide range of features, including precise "gestation" periods between birth events, gradual aging of individual molecules, or sudden random bursts of synthesis. This raises two central questions: How do single-cell fluctuations in abundances depend on the coarse graining of the biochemical hardware, and how can the effective coarse graining be inferred from measurements of fluctuations in single cells?

Most experimental noise studies have measured how the variation in single-cell protein levels depends on transcription and translation rates and typically compare the results with stochastic models based on specific assumptions about the underlying molecular mechanisms. Gene activation and transcription require numerous chemical events: from repressors falling off DNA to RNA polymerase elongating nascent transcripts. For synthetically engineered gene circuits in *Escherichia coli*, these processes can produce exponential waiting times between transcription events (*9*–*11*) despite the many microscopic substeps involved. Similar Poisson statistics have been observed in a wide range of physical systems, starting with Bortkewitsch's classic study on the number of Prussian cavalry officers kicked to death by horses (*12*). However, most genes have a more complex control, involving several repressors, transcription factors, and mediators, as well as chromatin remodeling or changes in supercoiling. Such systems generate nonexponential time intervals between transcription windows, unless a single elementary reaction step is rate limiting. In particular, promoters that gradually mature through a series of inactive states (with several hidden Poisson steps) before activating can create narrowly distributed gestation periods between transcription windows (Fig. 1). The statistical uncertainty in the waiting times can also be reduced by programmed cell cycle activation, replication-activated transcription, or circadian clocks (*13*). On the other hand, fluctuations can

[1]Department of Systems Biology, Harvard University, Boston, MA 02115, USA. [2]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA.

*To whom correspondence should be addressed. E-mail: johan_paulsson@hms.harvard.edu

be increased when many copies of a transcript are made in short-lived transcription windows (Fig. 1), creating random bursts of synthesis. A great variety of scenarios is possible, but the stochastic properties of individual gene expression events have only been measured in a few simple systems (*10*, *11*), and the molecular mechanisms are not sufficiently characterized to predict specific burst or waiting-time statistics. Because gestation and bursting can decrease and increase fluctuations in gene expression, respectively, their combined effect is also hard to intuit. However, it is possible to collectively understand such complex mechanisms by mathematically analyzing families of processes. Here, we consider a cell with m molecules of an mRNA and p molecules of the protein and make two generalizations with respect to previous analyses: We allow transcripts to effectively be made both in arbitrary independent bursts of b molecules ($m \rightarrow m + b$) and at arbitrary independent time intervals T (Fig. 1), where T and b vary randomly. To understand how different types of transcription statistics affect single-cell protein fluctuations, we evaluate these assumptions in the context of a standard model (*14*, *15*), where translation ($p \rightarrow p + 1$) occurs with a constant probability per second per transcript, and where both molecules decay ($m \rightarrow m-1$, $p \rightarrow p - 1$) exponentially with average lifetimes τ_m and τ_p. Similar models have been suggested for gene activation (*6*, *16*, *17*) and many other processes, but as with the earlier gene expression model, these have not considered generalized burst or waiting-time distributions. Basic tools from probability theory (*18*) can then be used to show that the stationary variance in protein abundance—the most commonly reported noise measure in the experimental literature—is insensitive to the shapes of the distributions for b and T (Fig. 2) and approximately follows

$$\frac{\sigma_p^2}{\langle p\rangle^2} \approx \overbrace{\frac{1}{\langle p\rangle}}^{\text{low-copy noise}} + \overbrace{\underbrace{\frac{\langle b\rangle(\sigma_T^2/\langle T\rangle^2 + \sigma_b^2/\langle b\rangle^2) + 1}{2}}_{\text{coarse graining}} \times \frac{1}{\langle m\rangle}}^{\text{mRNA noise}} \times \underbrace{\frac{\tau_m}{\tau_m + \tau_p}}_{} \quad (1)$$

where $\langle \ldots \rangle$ and σ denote averages and SDs, respectively. The equation is exact for exponential time intervals and is an excellent approximation for many types of strongly nonexponential times but breaks down when large and narrowly distributed bursts occur at precise time intervals [(*19*) and (Fig. 2)]. At low average protein abundances $\langle p\rangle$, relative protein levels spontaneously fluctuate because each random birth and death of a protein then has a larger relative effect on the total. The simplicity of this low-copy noise term reflects the assumption of individual protein molecules being produced and degraded at exponential time intervals. Because mRNAs determine the rate of protein synthesis, proteins also inherit noise from mRNAs. The mRNA noise, in turn, depends on the average mRNA abundance $\langle m\rangle$ and the burst and waiting-time statistics. Because the protein level cannot immediately adjust to changes in the protein synthesis rate, proteins effectively take a time average of a series of mRNA fluctuations and $0 < \tau_m/(\tau_m + \tau_p) < 1$.

Equation 1 is related to a previous analysis in which we similarly considered a model for stochastic gene expression and used it to reinterpret experiments (*19*). That model used fluctuation-dissipation relations to generalize the concentration-dependent tendencies to return to an average, while assuming simple exponential waiting times between events. Here, we took the opposite approach and generalized the randomness of the individual events, while assuming simple average dynamics.

The different parts of Eq. 1 represent qualitatively different aspects of random processes. The overall topology of the reaction network determines which components produce fluctuations, which are captured by $\langle p\rangle$ or $\langle m\rangle$. The connections between the chemical species in the network, in turn, determine the average amplification or suppression of fluctuations: effects that can be understood from deterministic analyses in which one component adjusts to another (*17*), which is captured by the time-averaging factor. Finally, the dynamics of how fluctuations are generated—the most central aspect of stochasticity—are captured by the bursts and waiting times in the coarse-graining factor. Most studies overlook

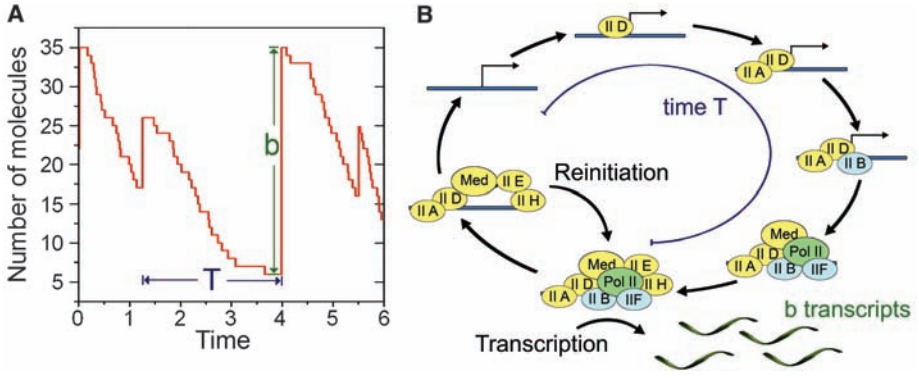

Fig. 1. (**A**) Generalized birth process where molecules are made in random bursts of b molecules at random time intervals T, where each b and T is independent, as for the mRNAs in Eq. 1. (**B**) Simplified sketch of polymerase II (Pol II)–mediated transcription in *S. cerevisiae*. Transcription factor TFIID binds DNA and is stabilized by TFIIA, which is followed by TFIIB. This complex then recruits Pol II, a Mediator, and TFIIF, after which TFIIE and TFIIH bind. Upon transcription initiation, Pol II, TFIIB, and TFIIF are released, but the rest of the complex remains, facilitating rapid reinitiation. Depending on parameters, this process can repeat and produce an effective burst of b transcripts, until the entire complex falls off the DNA and has to be reassembled, requiring a time T. The cartoon illustrates a roughly irreversible progression, which requires energy consumption, but the mathematical analysis allows for any distribution of waiting times T between transcription events.

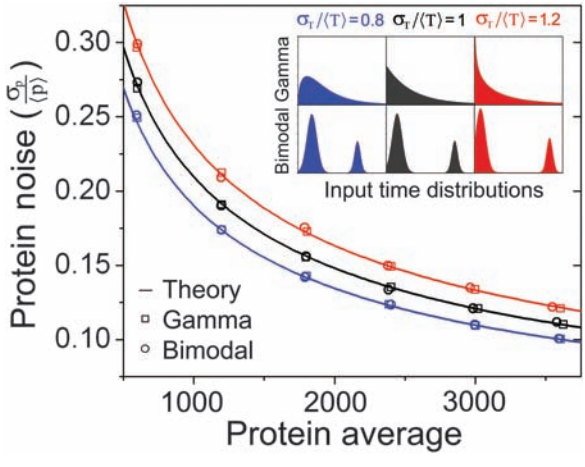

Fig. 2. Random intervals T between transcription events and their effect on protein fluctuations. Noise versus average in protein abundance for varying rates of transcription is shown. Curves are from Eq. 1 and symbols are from exact simulations that sample T from $f(T)$ in the inset. For a given $\sigma_T/\langle T\rangle$, the choice of $f(T)$ has no effect on $\sigma_p/\langle p\rangle$ despite the exotic distributions used. The parameters $\langle p\rangle/\langle m\rangle = 300$ and $\tau_p/\tau_m = 6$ (*19*) are representative for many genes, and near-perfect matches are also observed over many orders of magnitude in all parameters. (**Inset**) Hypothetical probability densities $f(T)$, with gamma distributions (exponential for $\sigma_T/\langle T\rangle = 1$) in the top row and examples of bimodal distributions (truncated sum of two Gaussians) in the bottom row, and with the same averages and variances in each column.

nontrivial coarse graining for simplicity (*20*), though some consider geometrically distributed bursts of mRNA or protein synthesis (*14*, *15*) at exponential time intervals, as observed in certain simple systems (*9*–*11*). In this case, the coarse-graining factor (for the mRNA or protein) can be expressed as $1 + \langle b \rangle$, which does not explain the effect of variation in the bursts and waiting times. By contrast, Eq. 1 shows how cells could exploit narrowly distributed gestation periods to greatly reduce variation in protein abundance from cell to cell, with the largest relative effect when synthesis, on average, occurs in large bursts. However, because the effect is determined by the sum of the normalized variances in bursts and waiting times, it also shows that narrowly distributed gestation periods only marginally reduce noise when burst sizes are widely distributed.

Many macromolecules approach their deaths gradually, passing through a series of states before finally degrading. For example, eukaryotic transcripts have their polyadenylate [poly(A)] tails sequentially chewed up (*21*) before degrading the protein-coding part of the message (Fig. 3). Measurements in *Saccharomyces cerevisiae* (by researchers using, for example, *PGK1* or *GAL10*) indeed demonstrate (*22*) strongly nonexponential mRNA decay curves, with long refractory periods where the poly(A) tail is shortened before the mRNA is degraded. Such "senescence" before death is in some sense similar to gestation before birth, reducing the uncertainty in the lifetimes of individual molecules. Here, we consider how senescence affects fluctuations in abundances across cells in the population. To keep the mathematical analysis as simple as possible, we assume that transcripts are made one by one at exponential time intervals and use the same assumptions regarding translation and proteolysis as previously indicated. The stationary mRNA distribution is then Poissonian, with $\sigma_m^2 = \langle m \rangle$, regardless of the distribution of mRNA lifetimes as long as the deaths of individual transcripts are independent of each other. However, even though the mRNA distribution is unaffected, mRNA fluc-

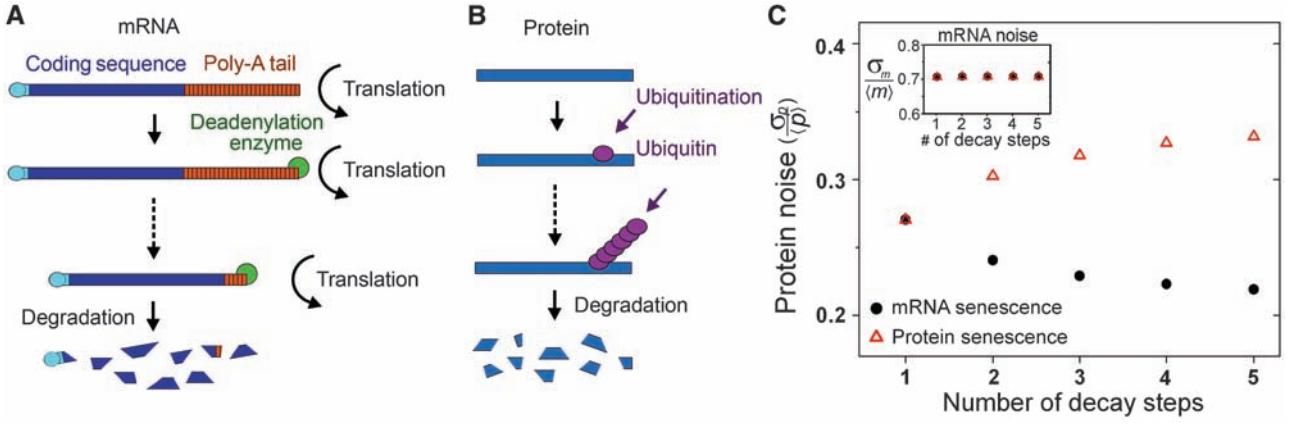

Fig. 3. Molecular senescence and its effects on noise in mRNA and protein abundance. (**A**) Cartoon of mRNA senescence when the poly(A) tail is progressively shortened before degradation. (**B**) Cartoon of protein senescence when the protein is ubiquitinated several times before degrading. (**C**) Noise in the abundance of proteins (main graph) and mRNA (inset) as a function of the number of Poisson steps N in the degradation pathway, as compared for the same average lifetimes. The graphs come from exact analytical calculations (*19*). For the mRNA senescence curve, proteins are assumed to decay exponentially, and for the protein senescence curve, mRNAs are assumed to decay exponentially. Parameters are the same as in Fig. 2 for the exponential $f(T)$. The mRNA senescence (black circles) has no effect on the mRNA variance but reduces protein noise because time averaging of the mRNA fluctuations becomes more efficient. Conversely, protein noise increases with more pronounced protein senescence (red triangles) because time averaging of mRNA fluctuations is impaired.

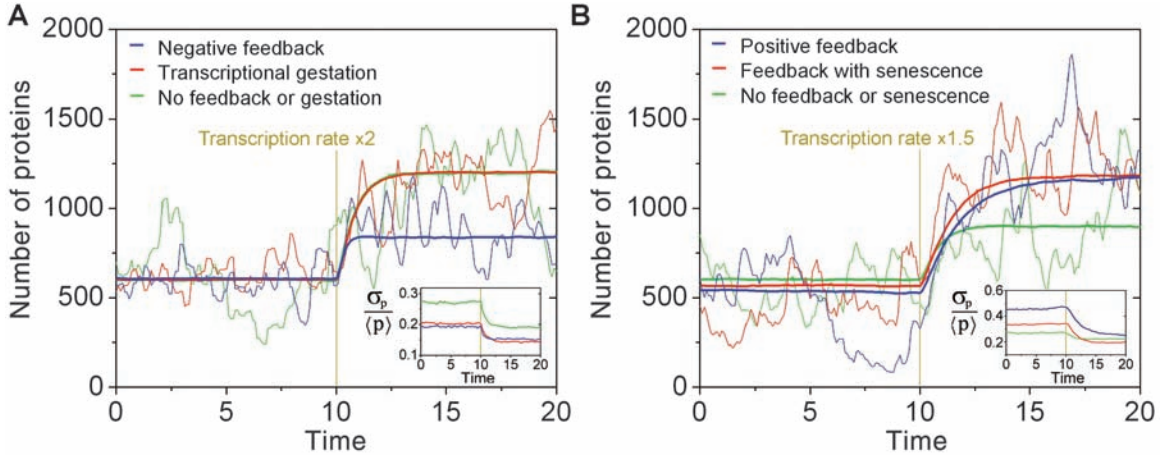

Fig. 4. Simulated time courses for averages (smooth curves) and sample paths (jagged curves) illustrating the difference between feedback and different coarse-graining mechanisms. At time t = 10, the rate of transcription is increased by a factor of 2 (A) and 1.5 (B), respectively. (**A**) Comparison of an idealized negative feedback loop (blue curve), nonexponential gestation between transcription events (red curve), and a reference gene without feedback or gestation (green curve). For the idealized negative feedback, the transcription intensity follows $k_m 2/(1 + (p/600)^2)$, where k_m is the transcription rate of the reference gene. For gestation, the time intervals between creation events follow a gamma distribution of order N = 8 and average $1/k_m$. Both negative feedback and gestation produce less noise than the reference gene (inset), but the average response is greatly reduced for negative feedback. (**B**) Comparison of an idealized positive feedback loop (blue curve), with the combination of positive feedback and mRNA senescence (five exponential time steps; red curve), and the same reference gene as in (A) (green curve). For the idealized positive feedback, the transcription intensity follows $k_m(0.01 + 0.99\ p/(600 + p))$. In both cases, the output level is roughly doubled, but when positive feedback is combined with mRNA senescence, the noise increase is smaller relative to that of the reference gene (inset).

tuations still contribute less to the width of the protein distribution than they would without senescence. If transcripts are degraded after N Poisson steps, producing gamma-distributed lifetimes, the stationary protein variance exactly follows

$$\frac{\sigma_p^2}{\langle p\rangle^2} = \frac{1}{\langle p\rangle} + \frac{1}{\langle m\rangle} \times \underbrace{\left[1 + \frac{\tau_p}{\tau_m}\left(\left(\frac{\tau_p N}{\tau_m + \tau_p N}\right)^N - 1\right)\right]}_{\text{Coarse-grained time-averaging factor}} \quad (2)$$

The time-averaging factor now depends on the coarse graining and decreases with increasing number of decay steps N [(*19*) and (Fig. 3)]. The intuitive explanation for these somewhat unexpected effects lies in the distinction between the dynamics of the noise and the overall distribution: For any given average mRNA lifetime, senescence accelerates changes in the mRNA levels in individual cells without affecting the probability of occurrence in the population. If each molecule lived for exactly τ_m time units, mRNA levels at time t would be completely uncorrelated with the levels before time $t-\tau_m$. This in turn makes it easier for the protein to "time-average out" mRNA fluctuations, because the efficiency of time averaging depends on the rate of change in the protein relative to the rate of change in the mRNA [see (*19*) for autocorrelation analyses]. Simple and well-documented molecular mechanisms of degradation (*22*) thus allow cells to exploit the internal transitions of independent molecules to reduce fluctuations in total concentrations. Protein senescence [for example, when multiple ubiquitination events are required before degradation (Fig. 3)] would instead speed up the protein response at any given average lifetime and thus prevent time averaging. Reducing the variability in the lifetimes then causes increased variability in protein concentrations (Fig. 3).

The bursting, gestation, and senescence mechanisms above modify the spontaneous noise in a system without affecting the average susceptibility to changes in parameters: A change in the rates of synthesis or degradation, which in turn may depend on upstream signaling, still has a proportional effect on average abundances (Fig. 4) in all cases. Positive or negative feedback control, by contrast, amplifies (*23*) or dampens (*24*) noise, respectively, while at the same time amplifying or dampening external signals (Fig. 4). Furthermore, negative feedback control only substantially suppresses noise when operating at high gain, but high-gain mechanisms are instead more susceptible to time lags or noisy intermediates, which destabilize feedback control and increase noise levels. Gestation and senescence, by contrast, operate without closing a potentially unstable loop and may therefore reduce spontaneous noise more efficiently. The noise reduction can indeed be extremely efficient when gestation and senescence are combined, because they prevent fluctuations from arising rather than correcting existing fluctuations.

The convenience of fluorescent reporters is creating a shift of focus in quantitative cell biology from bulk averages to individual cells, producing a wealth of data on nongenetic heterogeneity from microbes (*25*, *26*) to humans (*27*). The extent of heterogeneity is interesting in itself, but single-cell data can also be used to extract more information about the underlying processes: Individual responses to individual signaling events say much more than average responses to average signals. In particular, many quantitative studies have used the properties of the noise to infer microscopic kinetic mechanisms, testing stochastic models by analyzing how the variance in protein abundance responds to changes in the rates of transcription and translation (*3*–*5*, *7*, *25*). Analyses in *Bacillus subtilis* (*3*), *E. coli* (*4*, *28*), and *S. cerevisiae* (*5*, *6*, *25*) studied a range of genes and measured the distributions of protein abundance for different parameter values by changing the expression rates genetically, by changing growth conditions, or by adding inducers or inhibitors: in some cases with the use of dual fluorescent reporters to first separate the "intrinsic" randomness of the chemical events in gene expression from the "extrinsic" intracellular variation in the expression rates (*4*, *6*, *28*). Some studies also measured noise correlations between different proteins in genetic activation cascades (*7*) or between different time points in the same cell (*13*, *28*, *29*). The conclusions varied in the specific details but have formed a broad consensus that the intrinsic noise in protein abundance reflects low numbers of transcripts per cell, possibly with burstlike transcription resulting from brief random periods of gene activation.

The qualitative agreement between different studies and the excellent quantitative fits to the accompanying stochastic models seem to support the underlying models, which typically assume memory-less single-step transitions between births and deaths of genes, mRNAs, and proteins. However, Eqs. 1 and 2 show that very different stochastic processes can produce exactly the same response in such experiments. Changing the translation efficiency or the frequency of transcription bursts only affects $\langle p\rangle$ and $\langle m\rangle$, and the response to changes then follows $\sigma_p^2/\langle p\rangle^2 = 1/\langle p\rangle + C/\langle m\rangle$ in all cases, where C is a proportionality constant. Such experiments indicate which components in a network produce fluctuations but unexpectedly say nothing about how those fluctuations arise (*30*). For example, if protein noise was reduced by increasing the transcription rate but not by increasing the translation rate, as shown for some genes in *B. subtilis* (*3*), all versions of the mRNA-protein models in Eqs. 1 and 2 suggest that the noise comes from mRNA fluctuations (*3*, *19*). However, exactly the same fit is obtained (i) whether transcription is burstlike or perfectly regular or (ii) whether individual transcripts decay exponentially or gradually senesce (Fig. 4). Even in the ideal case where simple models without tunable parameters provide predictions that are later tested experimentally (*3*, *15*), perfect fits are equally consistent with mechanisms that are very random, very regular, or anything in between. Occam's razor can still eliminate obscuring details, but when several simple yet very different explanations work equally well, choosing one specific model can also brush interesting phenomena under Occam's "rug."

What additional tests could be used to discriminate between different types of underlying stochastic processes? One approach is to directly measure the relevant kinetic parameters. If fluc-

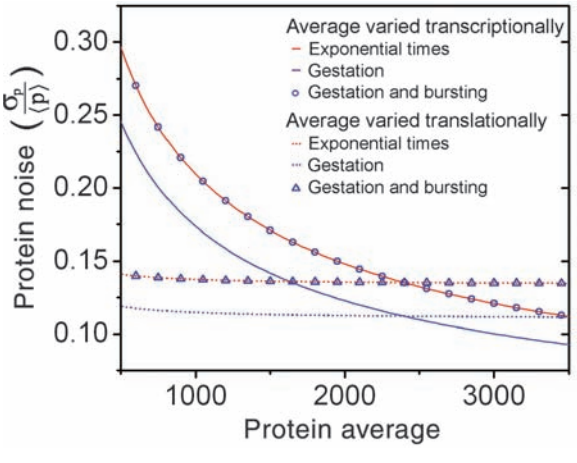

Fig. 5. Noise versus average in protein abundance from Eq. 1 for varying rates of transcription (solid lines) and translation (dotted lines). Red and blue lines correspond to exponential ($\sigma_T/\langle T\rangle = 1$) or nonexponential ($\sigma_T/\langle T\rangle = 0.6$) waiting times between transcription events, respectively, without mRNA bursting. Symbols correspond to nonexponential ($\sigma_T/\langle T\rangle = 0.6$) waiting times, with the use of mRNA bursting with $\langle b\rangle = 1$ and $\sigma_b/\langle b\rangle = 0.8$. The noise responds more sharply to changes in the transcription rate than to changes in the translation rate, reflecting the fact that, for the parameters used (*19*), most of the protein noise comes from random births and deaths of the mRNAs. However, if proteins are measured in arbitrary fluorescence units, the horizontal axis can be scaled and the transcriptional curves are indistinguishable. Similarly, if mRNA numbers are not measured directly, it is impossible to distinguish the effect of low mRNA numbers from that of high coarse-graining factors, because either effect shifts the curves vertically. Finally, the coarse-graining factor itself depends on factors not usually measured. Each line can then correspond to many combinations of timing and bursting.

tuations arise as a result of low abundances, the average levels should be measured in absolute numbers, not in arbitrary units of fluorescence (as is typically the case). This is particularly important for mRNAs: The low-copy components that introduce fluctuations must be counted (instead of the high-copy components that merely respond to underlying randomness). Measuring $\langle p \rangle$ and $\langle m \rangle$ would provide an estimate of C but would still not separate coarse graining from time averaging or other deterministic features (Eqs. 1 and 2). Time averaging could be estimated by measuring the average lifetimes of the components, but the type of time averaging should also be confirmed by manipulation experiments where the degradation rates of mRNAs and proteins are varied. If all these parameters are determined with high accuracy, the coarse-graining factor in Eq. 1 could be estimated. However, even in this ideal case, the relative contributions of gestation and bursting would still be unknown (Eqs. 1 and 2 and Fig. 5). The effects of precise gestation and random bursting could even cancel out and make it appear in these experiments as if molecules were born at exponential time intervals without bursts. More information can also be gained by considering the full distributions rather than just variances, but the exact shapes of distributions are more sensitive to experimental artifacts and have similar problems with experimental discrimination. Conclusive experimental analyses of coarse graining instead require accurate time series where the burst and dwell time statistics can be directly observed. A few pioneering studies have quantitatively monitored transcription (*11*) and translation (*9*, *10*) in *E. coli* with single-molecule resolution. For simple or synthetically engineered genes, they demonstrated exponential time intervals between geometric bursts of transcripts and proteins, respectively. These methods are now being used to study regulated genes and gene expression in eukaryotes, where gestation, senescence, and complicated bursting patterns are expected (*31*).

The fact that standard methods cannot discriminate between different types of coarse graining (Fig. 5) may explain why most experimental results have so closely matched the first models that were used. However, it also means that certain aspects of the conclusions are more robust to flaws in some of the most central model assumptions: By not distinguishing between how fluctuations arise (i.e., fitting any type of coarse graining), the methods are more suitable for identifying which components produce the fluctuations, which is both nontrivial and important.

Our findings mathematically connect noise in the single-cell protein abundance to bursting, gestation, and senescence in gene expression, describing how 10 molecules in some sense can statistically behave as if they were 5 or 20 molecules without control loops. We also show that standard single-cell measurements cannot detect or exclude these features: They only suggest which components contribute fluctuations, not how they contribute. Similar phenomena have been observed for bunching and antibunching in photon emissions (*32*), as well as for many molecular-scale cellular processes. Microtubules switch between growth and decay phases (*33*) where a burst of subunits is added or removed, and opening an ion channel can let a burst of molecules through. Nonexponential gestation periods have, in turn, been demonstrated in the rotational switching of *E. coli* flagellar motors (*34*) and in the replication control of bacterial plasmids (*14*), whereas the senescence of active rhodopsin molecules was recently shown to explain the reproducibility of the retinal signaling response to single photons (*35*). Advances in single-molecule live-cell imaging could now enable similar breakthroughs for mRNAs and proteins and finally reveal the effective coarse graining of gene expression.

References and Notes

1. O. G. Berg, *J. Theor. Biol.* **71**, 587 (1978).
2. H. H. McAdams, A. Arkin, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 814 (1997).
3. E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, A. van Oudenaarden, *Nat. Genet.* **31**, 69 (2002).
4. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, *Science* **297**, 1183 (2002).
5. W. J. Blake, M. Kaern, C. R. Cantor, J. J. Collins, *Nature* **422**, 633 (2003).
6. J. M. Raser, E. K. O'Shea, *Science* **304**, 1811 (2004).
7. J. M. Pedraza, A. van Oudenaarden, *Science* **307**, 1965 (2005).
8. G. M. Suel, R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, M. B. Elowitz, *Science* **315**, 1716 (2007).
9. L. Cai, N. Friedman, X. S. Xie, *Nature* **440**, 358 (2006).
10. J. Yu, J. Xiao, X. Ren, K. Lao, X. S. Xie, *Science* **311**, 1600 (2006).
11. I. Golding, J. Paulsson, S. M. Zawilski, E. C. Cox, *Cell* **123**, 1025 (2005).
12. L. v. Bortkewitsch, B. G. Teubner, *Monatsh. Mathematik* **9**, A39 (1898).
13. A. Amir, O. Kobiler, A. Rokney, A. B. Oppenheim, J. Stavans, *Mol. Syst. Biol.* **3**, 71 (2007).
14. J. Paulsson, M. Ehrenberg, *Q. Rev. Biophys.* **34**, 1 (2001).
15. M. Thattai, A. Van Oudenaarden, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8614 (2001).
16. P. S. Swain, M. B. Elowitz, E. D. Siggia, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12795 (2002).
17. J. Paulsson, *Phys. Life Rev.* **2**, 157 (2005).
18. Materials and methods are available as supporting material on *Science* Online.
19. J. Paulsson, *Nature* **427**, 415 (2004).
20. Nontrivial coarse graining is typically overlooked in the field of biology because the mathematical frameworks come from chemistry, where such effects are less important: The chemical master equation can account for any statistical feature, but this comes at the expense of an extended state space, which greatly complicates derivations and interpretations.
21. C. J. Wilusz, M. Wormington, S. W. Peltz, *Nat. Rev. Mol. Cell Biol.* **2**, 237 (2001).
22. C. J. Decker, R. Parker, *Genes Dev.* **7**, 1632 (1993).
23. A. Becskei, B. Seraphin, L. Serrano, *EMBO J.* **20**, 2528 (2001).
24. A. Becskei, L. Serrano, *Nature* **405**, 590 (2000).
25. A. Bar-Even *et al.*, *Nat. Genet.* **38**, 636 (2006).
26. J. R. Newman *et al.*, *Nature* **441**, 840 (2006).
27. A. Sigal *et al.*, *Nature* **444**, 643 (2006).
28. N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, M. B. Elowitz, *Science* **307**, 1962 (2005).
29. N. Friedman, S. Vardi, M. Ronen, U. Alon, J. Stavans, *PLoS Biol.* **3**, e238 (2005).
30. If the sizes of the bursts are varied, which they have been in some careful analyses (*6*), it is possible to get closer to estimating the coarse-graining statistics directly from the population variances. Such analyses say more about the underlying process but do not discriminate between different types of coarse graining of the underlying processes that create the bursting.
31. J. R. Chubb, T. Trcek, S. M. Shenoy, R. H. Singer, *Curr. Biol.* **16**, 1018 (2006).
32. C. G. Hubner *et al.*, *Phys. Rev. Lett.* **91**, 093903 (2003).
33. T. Mitchison, M. Kirschner, *Nature* **312**, 237 (1984).
34. E. A. Korobkova, T. Emonet, H. Park, P. Cluzel, *Phys. Rev. Lett.* **96**, 058105 (2006).
35. T. Doan, A. Mendez, P. B. Detwiler, J. Chen, F. Rieke, *Science* **313**, 530 (2006).

36. We thank M. Ehrenberg and O. G. Berg for discussions and A. Hilfinger for comments on the manuscript. This work was supported by Microsoft, the Human Frontier Science Program, and NSF under grant DMS-0720056.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5861/339/DC1
SOM Text
Fig. S1
Table S1
References

26 April 2007; accepted 12 December 2007
10.1126/science.1144331