Review

Models of stochastic gene expression

Johan Paulsson

Department of Applied Mathematics and Theoretical Physics, Center for Mathematical Sciences, University of Cambridge, Wilberforce Rd, CB3 0WA, Cambridge, UK

Abstract

Gene expression is an inherently stochastic process: Genes are activated and inactivated by random association and dissociation events, transcription is typically rare, and many proteins are present in low numbers per cell. The last few years have seen an explosion in the stochastic modeling of these processes, predicting protein fluctuations in terms of the frequencies of the probabilistic events. Here I discuss commonalities between theoretical descriptions, focusing on a gene-mRNA-protein model that includes most published studies as special cases. I also show how expression bursts can be explained as simplistic time-averaging, and how generic approximations can allow for concrete interpretations without requiring concrete assumptions. Measures and nomenclature are discussed to some extent and the modeling literature is briefly reviewed.
© 2005 Elsevier B.V. All rights reserved.

Contents

E-mail address: j.paulsson@damtp.cam.ac.uk (J. Paulsson).

1. Introduction

All cellular events directly or indirectly depend on probabilistic collisions between molecules. If large numbers of identical events occurred in the same cell, and they were statistically independent, relative fluctuations could be ignored and deterministic rate equations would suffice to describe dynamics. But the numbers are not large and the events are not independent. Active genes are often present in zero to one copy, mRNAs can be equally rare, and most proteins are present in less than 100 molecules per bacterial cell. Substrates, enzymes and regulatory molecules can also fluctuate and further randomize expression rates.

Cells have many mechanisms for reducing or suppressing harmful fluctuations [1–5]. If these are so efficient that fluctuations are negligible, stochastic models may seem redundant. However, if we want to understand rather than just mimic a process, an absence of randomness must also be explained probabilistically. Stochastic models are equally relevant whether genes are expressed randomly or regularly. Here I review some common denominators shared by many of the stochastic models and briefly discuss possible extensions.

2. The standard model

2.1. Molecular assumptions

Many models claim to account for the critical steps in gene expression, yet there is little agreement on what those steps are, i.e., what sets of transitions that can be condensed into effective reaction steps and what concentrations that can be absorbed into rate constants. Experiments in turn show that similar genes can produce very different fluctuations depending on the exact conditions [6–9]. This makes it difficult to invest a high degree of belief in any particular model and suggests that all we can hope for is

an understanding of first principles. Whether or not we address the right first principles is always difficult to tell, but in the initial phase we can at least make sure to thoroughly understand the models.

Most theoretical descriptions [10–22] have independently focused on the same basic phenomenon. Because mRNA levels determine the rate of protein synthesis, the number of proteins per cell n_3 chases the number of mRNAs n_2, which in turn may chase the number of active genes n_1. This principle could be extended indefinitely: Any process that indirectly affects the rates of gene expression can potentially randomize protein concentrations and would then have to be included in realistic stochastic models. However, most models focus on genes, RNAs and proteins, and implicitly include all other processes in effective rate constants.

Gene activation can have many different molecular causes, including dissociation of repressors, association of activators, or chromatin remodeling. The details vary from gene to gene and organism to organism and can include transitions between numerous different states. As a first approximation, in both pro- and eukaryotes, the overall kinetic dynamics can be described by a random telegraph process [23]:

$$off \underset{\lambda_1^-}{\overset{\lambda_1^+}{\rightleftharpoons}} on \tag{1}$$

where each gene spontaneously switches on and off with rates λ_1^+ and λ_1^- respectively. Depending on growth conditions, bacteria can have several copies of partially replicated chromosomes. If the gene under study is close to the origin of replication, some cells may then have as many as eight copies of the same gene. Here I assume that a constant $n_1^{\max}$ copies independently switch on and off as in Eq. (1), ignoring cell growth and gene replication. The stationary distribution for the number of active genes is then Binomial as for the tossing of $n_1^{\max}$ unfair coins, where the probability of being on is $P_{\text{on}} = \lambda_1^+/(\lambda_1^+ + \lambda_1^-)$.

Transcription and translation are typically assumed to follow Poisson processes where the production probabilities per time unit $\lambda_2 n_1$ and $\lambda_3 n_2$ are proportional to the number of active genes and mRNAs respectively. These may or may not be good approximations. The binding of RNA polymerase may change the structure of the gene, either blocking or facilitating further transcription. Replication can change the chromatin structure or kick off activators, and repressors and thereby cause abrupt changes in expression rates. The transport of mRNAs out of the eukaryotic nucleus, or the release of finished transcripts from prokaryotic genes, may also affect dynamics.

Finally, mRNAs and proteins are often described as having exponentially distributed lifetimes, assuming that each degradation event is independent and memory-lacking. However, the events may depend on each other if the molecules compete for RNases or proteases. If these degradation enzymes operate close to saturation, the degradation rate per substrate molecule is lower in cells that by chance have a higher substrate concentration, thereby correcting perturbations less efficiently [24]. The degradation pathway may also include many rate-limiting steps. Even if each molecule is a statistically independent unit, internal molecular memory may then still allow molecules to grow old before they die, possibly reducing the variation in individual life-times.

Other likely contributors to the observed variability include fluctuations in the many enzymes and substrates involved, cell cycle effects, and random partitioning of copies at cell division. Ignoring these and other complications, the standard model only accounts for six exponential events: Constant rates of switching on and off individual genes, constant transcription per active gene, constant translation per transcript, and exponential decay of both transcripts and proteins. This is illustrated in Fig. 1 and
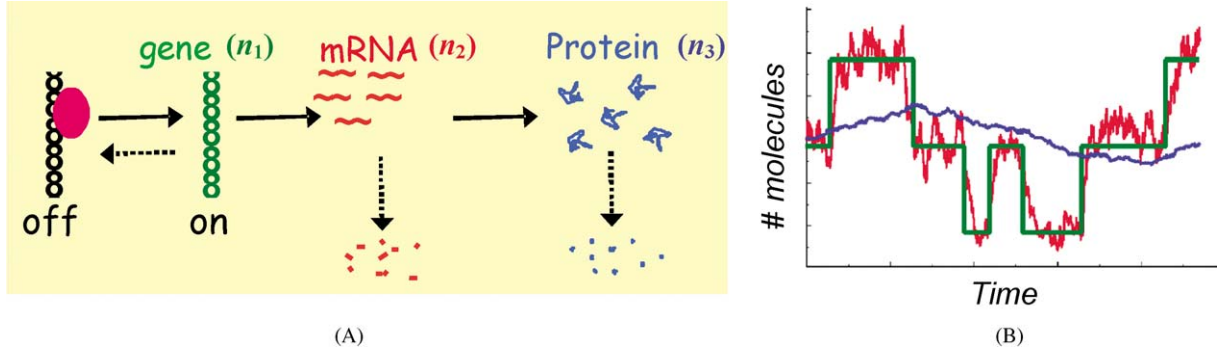
Fig. 1. (A) Cartoon of gene activation, transcription, translation, mRNA decay and proteolysis. The purple molecule represents a repressor. (B) Gillespie simulations [36] of the events described in Eq. (2), where the coloring matches the cartoon in (A). For the parameters used, most fluctuations come from genes, the mRNAs adjust quickly to changes in gene activity, and the protein adjusts slowly to changes in mRNA level. This may be common biologically, but all parameters could vary several orders of magnitude between different genes.

summarized by the following reaction diagram and corresponding dynamics for the averages $\langle x_i \rangle$:

$$
\begin{array}{lll}
\text{Gene activation:} & n_1 \xrightarrow{\lambda_1^+(n_1^{\max}-n_1)} n_1 + 1, & \dfrac{d\langle n_1 \rangle}{dt} = \lambda_1^+\left(n_1^{\max} - \langle n_1 \rangle\right) - \lambda_1^- \langle n_1 \rangle \\[2mm]
\text{Gene inactivation:} & n_1 \xrightarrow{\lambda_1^- n_1} n_1 - 1, & \qquad = \lambda_1^+ n_1^{\max} - \langle n_1 \rangle/\tau_1, \\[2mm]
\text{Transcription:} & n_2 \xrightarrow{\lambda_2 n_1} n_2 + 1, & \dfrac{d\langle n_2 \rangle}{dt} = \lambda_2 \langle n_1 \rangle - \langle n_2 \rangle/\tau_2, \\[2mm]
\text{mRNA degradation:} & n_2 \xrightarrow{n_2/\tau_2} n_2 - 1, & \\[2mm]
\text{Translation:} & n_3 \xrightarrow{\lambda_3 n_2} n_3 + 1, & \dfrac{d\langle n_3 \rangle}{dt} = \lambda_3 \langle n_2 \rangle - \langle n_3 \rangle/\tau_3. \\[2mm]
\text{Proteolysis:} & n_3 \xrightarrow{n_3/\tau_3} n_3 - 1, &
\end{array}
\tag{2}
$$

The parameter $\tau_1 = (\lambda_1^+ + \lambda_1^-)^{-1}$ is a characteristic time-scale for changes in gene activity, and τ_2 and τ_3 are the average lifetimes of mRNAs and proteins respectively. The average dynamics can be postulated immediately from the reaction scheme, but other statistical aspects must be explicitly calculated from the three-variable Markov process $dP(n_1, n_2, n_3)/dt$ that is defined by the individual events:

$$
\begin{aligned}
\frac{dP(n_1, n_2, n_3)}{dt} =\ & \lambda_1^+\left(n_1^{\max} - n_1 + 1\right)P(n_1 - 1, n_2, n_3) - \lambda_1^+\left(n_1^{\max} - n_1\right)P(n_1, n_2, n_3) \\
& + \lambda_1^-(n_1 + 1)P(n_1 + 1, n_2, n_3) - \lambda_1^- n_1 P(n_1, n_2, n_3) \\
& + \lambda_2 n_1 P(n_1, n_2 - 1, n_3) - \lambda_2 n_1 P(n_1, n_2, n_3) \\
& + (n_2 + 1)/\tau_2 P(n_1, n_2 + 1, n_3) - n_2/\tau_2 P(n_1, n_2, n_3) \\
& + \lambda_3 n_2 P(n_1, n_2, n_3 - 1) - \lambda_3 n_2 P(n_1, n_2, n_3) \\
& + (n_3 + 1)/\tau_3 P(n_1, n_2, n_3 + 1) - n_3/\tau_3 P(n_1, n_2, n_3).
\end{aligned}
\tag{3}
$$

Because the reaction rates are linear in terms of the state variables n_i, the time-dependent moments can be calculated exactly using generating functions. This procedure has been extensively described in

previous studies [10–14,19,20] of stochastic gene expression and was applied in exactly the same way to Eq. (3).

Unless stated otherwise, all equations will be evaluated at the stationary state where $dP(n_1, n_2, n_3)/dt = 0$. To reduce notational complexity this will not be indicated explicitly.

2.2. Protein fluctuations

The normalized stationary variance in the number of protein molecules per cell follows:

$$\underbrace{\frac{\sigma_3^2}{\langle n_3 \rangle^2}}_{\substack{\text{Total} \\ \text{protein noise}}} = \underbrace{\frac{1}{\underbrace{\langle n_3 \rangle}_{\text{Poisson}}}}_{\substack{\text{From individual} \\ \text{births and deaths}}} + \underbrace{\frac{1}{\underbrace{\langle n_2 \rangle}_{\text{Poisson}}} \underbrace{\frac{\tau_2}{\tau_3 + \tau_2}}_{\substack{\text{One-step} \\ \text{time-averaging}}}}_{\substack{\text{From spontaneous} \\ \text{mRNA noise}}} + \underbrace{\underbrace{\frac{1 - P_{\text{on}}}{\langle n_1 \rangle}}_{\text{Binomal}} \underbrace{\frac{\tau_2}{\tau_2 + \tau_3} \frac{\tau_1}{\tau_1 + \tau_3} \frac{\tau_1 + \tau_3 + \tau_1 \tau_3 / \tau_2}{\tau_1 + \tau_2}}_{\substack{\text{Two-step} \\ \text{time-averaging}}}}_{\substack{\text{From forced mRNA noise, originating} \\ \text{in gene activation-inactivation}}}. \quad (4)$$

The first noise term on the right-hand-side shows that if all other cellular factors are constant, proteins still display small-number Poisson fluctuations because individual birth and death events are probabilistic. This noise does not necessarily have to be exactly Poissonian if the assumptions above are not fulfilled, but some sort of small-number noise is very difficult to avoid because individual chemical events rely on random collisions between diffusing molecules.

The second and third terms both reflect random changes in the rate for protein synthesis. These are caused by mRNA fluctuations, that in this model follow:

$$\frac{\sigma_2^2}{\langle n_2 \rangle^2} = \frac{1}{\langle n_2 \rangle} + \frac{1 - P_{\text{on}}}{\langle n_1 \rangle} \frac{\tau_1}{\tau_2 + \tau_1}. \quad (5)$$

The first term of Eq. (5) again reflects small-number noise, now originating in the inherently probabilistic births and deaths of individual transcripts. The second term comes from random changes in gene activity where the first factor is a measure of stationary small-number gene fluctuations:

$$\frac{\sigma_1^2}{\langle n_1 \rangle^2} = \frac{1}{n_1^{\text{max}}} \frac{\lambda_1^-}{\lambda_1^+} = \frac{1 - P_{\text{on}}}{\langle n_1 \rangle}. \quad (6)$$

The $1 - P_{\text{on}}$ factor comes from the fact the number of active genes is Binomially rather than Poisson distributed. At any given average, Binomial variables have smaller relative fluctuations than Poissonian variables, which makes it possible to have a low number of genes without necessarily having large fluctuations as long as genes are mostly active, $P_{\text{on}} \approx 1$.

The second factor in Eq. (5) reflects time-averaging and always ranges between zero and one. Solving the second linear equation in Eq. (2) for fixed $\langle n_1 \rangle$ gives:

$$\underbrace{\langle n_2 \rangle_{t_1+t_2} - \langle n_2 \rangle_\infty}_{\substack{\text{Deviation from stationary} \\ \text{average at time } t=t_1+t_2}} = \underbrace{\left(\langle n_2 \rangle_{t_1} - \langle n_2 \rangle_\infty \right)}_{\substack{\text{Deviation from stationary} \\ \text{average at time } t=t_1}} e^{-t_2/\tau_2}. \quad (7)$$

The mRNA concentration n_2 thus exponentially forgets initial conditions with rate $1/\tau_2$. The parameter τ_2 therefore determines the effective kinetic memory in the sense that cellular events that occur further back than τ_2 time units ago have little influence on the current concentration. Rate $1/\tau_1$ similarly

determines how rapidly the gene activity changes. The ratio τ_2/τ_1 is thus a measure of how much gene activation changes up and down within the effective memory of the mRNA. If τ_2/τ_1 is large, the current mRNA level depends on a relatively long history of random ups and downs, which reduces mRNA fluctuations just like throwing many dice reduces relative fluctuations in the total outcome.

The exact same principles apply to protein fluctuations in Eq. (4). The second term represents time-averaged small-number mRNA fluctuations, and the third term represents small-number gene fluctuations, that are first time-averaged by mRNAs and then by proteins (Fig. 1B). Both time-averaging factors range between zero and one, and the compounded two-step time-averaging is always more efficient than either of the individual steps separately.

2.3. The simplest example of dynamic disorder

If genes fluctuate rapidly, or if stationary gene fluctuations are insignificant, the third term in Eq. (4) disappears. If mRNAs fluctuate rapidly, the second term instead disappears and the two-step time-averaging factor can be greatly simplified. In both cases, protein fluctuations can be written as:

$$\frac{\sigma_3^2}{\langle n_3 \rangle^2} = \frac{1}{\langle n_3 \rangle} + \frac{\sigma_E^2}{\langle n_E \rangle^2} \frac{\tau_E}{\tau_E + \tau_3} \tag{8}$$

where E represents the external environment of either mRNAs or active genes. The two types of noise represent the same basic principle, broadly defined as *dynamic disorder* [25,26]. 'Disorder' because the parameters for protein synthesis vary randomly between cells in a population, just like chemical reactions in an imperfect medium, and 'dynamic' because the parameters change in time. The Markov process has thus been extended to include both system and environment. In statistical physics, disorder often has different connotations because most studies consider how global properties emerge in terms of the spatial distributions of the underlying disorder. Here we only study local effects of the disorder itself, but percolation-type phenomena could certainly arise in tissues where cells express genes randomly and then affect adjacent cells.

3. Expression bursts

3.1. Translation bursts

Many studies have interpreted Eq. (8) in terms of bursts—brief periods of high expression intensity followed by long periods of low intensity. Most focus on the fact that each mRNA is translated several times, and often start with the mRNA-protein part of the model above assuming that the genes are constantly on, $P_{\mathrm{on}} = 1$. Further assuming that proteins decay slowly relative to their transcripts, $\tau_3 \gg \tau_2$, and measuring fluctuations by the Fano factor (variance over average) then leads to [19]:

$$\frac{\sigma_3^2}{\langle n_3 \rangle} = 1 + \frac{\langle n_3 \rangle}{\langle n_2 \rangle} \frac{\tau_2}{\tau_2 + \tau_3} \approx 1 + \frac{\langle n_3 \rangle}{\langle n_2 \rangle} \frac{\tau_2}{\tau_3} = 1 + \frac{n_1^{\mathrm{max}} \lambda_2 \tau_2 \lambda_3 \tau_3}{n_1^{\mathrm{max}} \lambda_2 \tau_2} \frac{\tau_2}{\tau_3} = 1 + \lambda_3 \tau_2 = 1 + \langle b \rangle \tag{9}$$

where b is the number of translations per transcript, typically on the order of 100 for an average *E. coli* gene. This formulation makes intuitive sense. If proteins were made in statistically independent events and decayed exponentially, stationary fluctuations would be Poissonian, with $\sigma_3^2/\langle n_3 \rangle = 1$. If they instead

were made in bursts of random size, fluctuations should be larger than Poissonian, just like a random walker who takes large and random leaps accumulate randomness more quickly. The burst interpretation can certainly be sound, but a few moderations are due.

First, to truly have brief periods of high intensity it is not enough that $\tau_3 \gg \tau_2$. Some mRNAs are present in hundreds of copies per cell at any given time, but are still so unstable that the approximation in Eq. (9) is close to exact. Protein synthesis is then not burst-like at all, regardless of the value of $\langle b \rangle$. That the equation can be written in terms of bursts does not mean that there actually are any bursts: The conditions for a mathematically correct expression are in this case more relaxed than the conditions for a physically sound interpretation. To have true bursts in the model above, it is instead necessary that most cells have zero transcripts so that the total synthesis rate switches randomly from a low to a high value. This is not uncommon in real cells where some mRNAs are exceptionally rare. However, it is not nearly as common as mRNAs having shorter life-times than proteins, which is almost universally true.

Second, the randomness of translation is represented by the first rather than the second term in Eq. (9), i.e., by the constant '1' rather than by $\langle b \rangle$. This is not merely a matter of perspective. Imagine that translation and protein decay were made entirely deterministic in the model above, using a macroscopic rate equation for n_3 where the synthesis rate $\lambda_3 n_2$ depended on the state of the Markov process for n_2. The only effect on the stationary protein variance in the equations above is then that the first term in Eq. (9) disappears. The 'translation burst noise' is thus unaffected even if we assume that translation is deterministic. If we instead assume that the mRNA is deterministic but that protein synthesis and decay are probabilistic as before, we instead keep the first term in Eq. (9) while the second one disappears. This illustrates the importance of separating fluctuation terms according to their origin in probabilistic events rather than responses to changes in parameters.

Third, if there are other sources of disorder, those noise terms can look quite bizarre when multiplied by $\langle n_3 \rangle$ and interpreted in terms of translation. A response in $\sigma_3^2/\langle n_3 \rangle$ to changes in the translation rate does not in any way reflect translational noise. The approximation in Eq. (9) in fact assumes a separation of time-scales where transcription, translation and mRNA decay are compounded into one chemical event. However, this is not a problem of the burst perspective itself, but of the measure used (see Section 5). It can be solved by simply writing Eq. (9) as $\sigma_3^2/\langle n_3 \rangle^2 = (1 + \langle b \rangle)/\langle n_3 \rangle$.

3.2. Bursting as simple time-averaging

Despite the strong intuitive notion of bursts, the moderations above show that the concept is quite restricted and that bursting should not be expected for most genes. However, it does simplify the equations by reducing the number of parameters. Would other interpretations allow for the same simplifications but at the same time allow for a physical explanation that works more broadly? Again assuming $\tau_3 \gg \tau_2$, normalizing the variance in Eq. (9) leads to:

$$\frac{\sigma_3^2}{\langle n_3 \rangle^2} \approx \frac{1}{\langle n_3 \rangle} + \frac{1}{n_1^{\max} \lambda_2 \tau_3}. \tag{10}$$

Here it is the number of transcripts made during the lifetime of the protein ($n_1^{\max} \lambda_2 \tau_3$) that determines the deviation from Poisson statistics, not the number of proteins made during the lifetime of a transcript ($\lambda_3 \tau_2$), as in the mathematically equivalent Eq. (9). As expected from Eq. (8), this is a simple version of time-averaging: $n_1^{\max} \lambda_2 \tau_3$ is the average number of independent transcription events that effectively

contribute to the current value of the protein. High numbers are more or less the same in different cells or at different times, thus reducing relative fluctuations.

3.3. Other causes of burst-like expression

There are other possible causes of burst-like expression apart from making many proteins per transcript. For example, if genes are mostly inactive ($P_{on} \ll 1$) but transcribe a large number of mRNAs while in the active state, transcription could occur in bursts. The concentration of free ribosomes, amino acids, or charged tRNAs may also fluctuate wildly [27], possibly with sharp spikes. Such changes in the underlying gene expression machinery could cause bursts, but are subject to the same moderations as above: The equations can be formulated in terms of bursts even for parameter values where there are no physical bursts. The only way to demonstrate bursts is by directly observing spikes of expression in the time-series.

4. Generic approaches

4.1. The Fluctuation–Dissipation Theorem

Rather than deriving results from particular assumptions, it is also possible to examine broader classes of random processes collectively and then afterwards suggest how the general principles are instantiated molecularly. Such approaches may seem doomed to hopeless abstraction, but can in fact be formulated in terms of physical principles that are at least as concrete as in the specific examples. A recent attempt [22] was based on a reinterpretation of the Fluctuation–Dissipation Theorem [22,28] (FDT), which in one formulation states that:

$$\frac{d\boldsymbol{\sigma}}{dt} = \mathbf{A}\boldsymbol{\sigma} + \boldsymbol{\sigma}\mathbf{A}^{\mathrm{T}} + \mathbf{B}. \tag{11}$$

Here $\boldsymbol{\sigma}$ is the matrix of covariances ($\sigma_{ii} = \sigma_i^2$), $\mathbf{A}$ is the Jacobian matrix for the dynamics of the averages, and $\mathbf{B}$ is a diffusion matrix that depends on the size of the random events. This equation has different names in different disciplines, and is sometimes called the Lyapunov Equation, the Linear Noise Approximation or the result of a 1st order van Kampen's Ω-expansion. Here we call it the FDT to emphasize physical rather than mathematical principles. It is also known as the FDT in the textbooks that most emphasize its use in chemistry [28] where both the perturbations and dynamics come from the reaction mechanisms.

A key approximation that makes the FDT practical to use is that matrices $\mathbf{A}$ and $\mathbf{B}$ are interpreted assuming that fluctuations can be ignored. The average rates are then approximated by the rates at the average concentration. For example, if $R = \lambda n^2$ then $\langle R \rangle = \lambda \langle n^2 \rangle = \lambda \langle n \rangle^2 + \lambda \sigma^2$, but when calculating $\mathbf{A}$ and $\mathbf{B}$ above, the second term would be ignored, using only $\langle R \rangle \approx \lambda \langle n \rangle^2$. This does not affect analyses of linear systems like the gene-mRNA-protein model above, but makes it possible to approximate nonlinear systems despite the fact that averages explicitly depend on fluctuations. Practically it means that we can calculate $\mathbf{A}$ from the deterministic rate equations. Here I keep the averaging brackets $\langle \ldots \rangle$ explicit, but they could also be removed if n_i were interpreted as the deterministic concentration. Matrix $\mathbf{A}$ is thus

defined by:

$$A_{ij} = \frac{\partial}{\partial \langle n_j \rangle} \frac{\partial \langle n_i \rangle}{\partial t} = \frac{\partial \langle J_i^+ \rangle}{\partial \langle n_j \rangle} - \frac{\partial \langle J_i^- \rangle}{\partial \langle n_j \rangle} \tag{12}$$

where J_i^+ and J_i^- are the total fluxes of production and elimination of species i. The diffusion matrix $\mathbf{B}$ is in turn defined by:

$$B_{ij} = \sum_k v_{jk} v_{ik} R_k \tag{13}$$

where reaction k occurs with rate R_k and produces v_{ik} molecules of species i. The flux is thus $J_{ik} = |v_{ik} R_k|$. Stationary variances can be calculated by setting Eq. (11) to zero:

$$-\mathbf{A}\sigma - \sigma \mathbf{A}^{\mathrm{T}} = \mathbf{B}. \tag{14}$$

Using this equation to obtain expressions like Eq. (4) involves three steps: Calculating $\mathbf{A}$ and $\mathbf{B}$ at steady state, solving the linear system of equations for σ, and rearranging the solution in terms of clear physical principles. The first two steps can be tedious but are always trivial, and the real conceptual challenge lies in the interpretations.

4.2. Illustrating the FDT for the gene-mRNA-protein model

For the gene-mRNA-protein system above, the Jacobian matrix can be directly calculated from Eq. (2):

$$\mathbf{A} = \begin{bmatrix} -1/\tau_1 & 0 & 0 \\ \lambda_2 & -1/\tau_2 & 0 \\ 0 & \lambda_3 & -1/\tau_3 \end{bmatrix}. \tag{15}$$

The diffusion matrix $\mathbf{B}$ is equally simple. There are six reactions and three species ($k = 1, \ldots, 6$, and $[i, j] = 1, 2, 3$), but each species is only affected by two reactions. Because no reaction events change two different species simultaneously, all off-diagonal elements in $\mathbf{B}$ are zero. Finally, each reaction only adds or removes a single molecule. This means that the product $v_{ik} v_{jk}$ only takes values zero or one, and:

$$\begin{aligned} \mathbf{B} &= \begin{bmatrix} \lambda_1^+(n_1^{\max} - \langle n_1 \rangle) + \lambda_1^- \langle n_1 \rangle & 0 & 0 \\ 0 & \lambda_2 \langle n_1 \rangle + \langle n_2 \rangle/\tau_2 & 0 \\ 0 & 0 & \lambda_3 \langle n_2 \rangle + \langle n_3 \rangle/\tau_3 \end{bmatrix} \\ &= \begin{bmatrix} 2\lambda_1^- \langle n_1 \rangle & 0 & 0 \\ 0 & 2\langle n_2 \rangle/\tau_2 & 0 \\ 0 & 0 & 2\langle n_3 \rangle/\tau_3 \end{bmatrix}. \end{aligned} \tag{16}$$

The second equality uses the fact that, at steady state, the total average synthesis flux equals the total average degradation flux. Element B_{11} can be further rewritten using:

$$\lambda_1^- = \left(1 - \frac{\lambda_1^+}{\lambda_1^+ + \lambda_1^-}\right)(\lambda_1^+ + \lambda_1^-) = \frac{1 - P_{\mathrm{on}}}{\tau_1}. \tag{17}$$

Parameter τ_1 in the denominator of B_{11} will then cancel out against the denominator of A_{11}, leaving factor $1 - P_{\mathrm{on}}$ which is a signature of the Binomial distribution.

Solving the equation system above exactly reproduces Eqs. (4)–(6), which in turn were derived using exact moment equations. The exactness is perhaps surprising given that the derivation of the FDT in

Eq. (11) typically assumes small Gaussian fluctuations while the gene-mRNA-protein model can generate very broad and skewed distributions with averages arbitrarily close to zero. However, the only essential approximation behind the FDT is that fluctuations are so small that the responses can be approximated as weakly nonlinear. Because the Markov process above is exactly linear, the variances thus follow exactly from the Fluctuation–Dissipation Theorem even though the full distributions are very non-Gaussian— something that can be shown using moment generating functions or other methods [29].

4.3. Normalizing the FDT

The algebra required to obtain expressions like Eq. (4) can be greatly reduced by normalizing the equations before rather than after they are solved. In the stationary state we then get:

$$\mathbf{M}\boldsymbol{\eta} + \boldsymbol{\eta}\mathbf{M}^{\mathrm{T}} + \mathbf{D} = 0 \tag{18}$$

where:

$$\eta_{ij} = \frac{\sigma_{ij}}{\langle n_i \rangle \langle n_j \rangle}, \quad M_{ij} = \frac{\langle n_j \rangle}{\langle n_i \rangle} A_{ij} \quad \text{and} \quad D_{ij} = \frac{B_{ij}}{\langle n_i \rangle \langle n_j \rangle}. \tag{19}$$

It is useful to rewrite the dynamic matrix $\mathbf{M}$ in more intuitive terms. The rules for differentiation of logarithms state that:

$$\frac{\partial \ln f}{\partial \ln x} = \frac{x}{f} \frac{\partial f}{\partial x} \quad \text{and} \quad \frac{\partial \ln(f/g)}{\partial \ln x} = \frac{\partial \ln f}{\partial \ln x} - \frac{\partial \ln g}{\partial \ln x}. \tag{20}$$

Applying these rules backwards to Eq. (12) gives:

$$A_{ij} = \frac{\partial \langle J_i^+ \rangle}{\partial \langle n_j \rangle} - \frac{\partial \langle J_i^- \rangle}{\partial \langle n_j \rangle} = \frac{\langle J_i \rangle}{\langle n_j \rangle} \left(\frac{\partial \ln \langle J_i^+ \rangle}{\partial \ln \langle n_j \rangle} - \frac{\partial \ln \langle J_i^- \rangle}{\partial \ln \langle n_j \rangle} \right) = -\frac{\langle J_i \rangle}{\langle n_j \rangle} \frac{\partial \ln \langle J_i^- \rangle / \langle J_i^+ \rangle}{\partial \ln \langle n_j \rangle} \tag{21}$$

where the first equality also uses the fact that, at steady state, $\langle J_i^+ \rangle = \langle J_i^- \rangle \equiv \langle J_i \rangle$. The dynamic matrix $\mathbf{M}$ is then:

$$M_{ij} = -\frac{\langle J_i \rangle}{\langle n_i \rangle} \frac{\partial \ln \langle J_i^- \rangle / \langle J_i^+ \rangle}{\partial \ln \langle n_j \rangle}. \tag{22}$$

The first factor can be further simplified by noting that, at steady state, the average turnover rate per molecule approximately equals the inverse average lifetime τ_i:

$$\frac{\langle J_i^- \rangle}{\langle n_i \rangle} = \frac{\langle J_i^+ \rangle}{\langle n_i \rangle} = \frac{\langle J_i \rangle}{\langle n_i \rangle} \approx \frac{1}{\tau_i}. \tag{23}$$

This is exact for exponential first-order degradation and approximate for nonlinear mechanisms. However, because the matrices in the stationary FDT are interpreted at the macroscopic steady state, $\mathbf{M}$ is evaluated in the hypothetical limit where there are no fluctuations and where the averages remain constant. In this idealization, each molecule is surrounded by a constant environment of other molecules, so that each species perfectly mimics first-order exponential decay regardless of any nonlinear responses to deviations from steady state. This means that though Eq. (23) is an approximation, it is not an additional approximation but already implicit in the stationary FDT approach above. Matrix $\mathbf{M}$ can thus be written

as:

$$M_{ij} = -\frac{H_{ij}}{\tau_i} \quad \text{where } H_{ij} = \frac{\partial \ln\langle J_i^-\rangle/\langle J_i^+\rangle}{\partial \ln\langle n_j\rangle}. \tag{24}$$

The second factor is a logarithmic gain or *elasticity* [1,22,30,31] and measures how the death-to-birth ratio changes with the concentrations: If $H_{ij} = 3$, then a 1% increase in component n_j will cause component n_i to decrease by increasing its death-to-birth ratio by approximately 3%. The elasticities are thus normalized measures of the strengths of the kinetic nonlinearities. Comparing the process to a random walk in a multidimensional landscape, elasticities measure the steepness of mountains and valleys.

Matrix **D** can be similarly reinterpreted. A full treatment will be published elsewhere, including arbitrary chemical events and other extensions. Here I restrict the analysis to nonlinear versions of simple cases like the gene-mRNA-protein model above, where each chemical event adds or removes a single molecule of a single species. This would not cover metabolic systems where one component turns into another component, but still applies to a large set of networks where one component only affects the rates of another. Mathematically, this means that:

$$D_{ii} = \frac{\langle J_i^+\rangle + \langle J_i^-\rangle}{\langle n_i\rangle\langle n_i\rangle} = 2\frac{\langle J_i\rangle}{\langle n_i\rangle\langle n_i\rangle} \quad \text{and} \quad D_{ij} = 0 \quad \text{for } i \neq j. \tag{25}$$

No we can use the trick from Eq. (23) and rewrite D_{ii} as:

$$D_{ii} = \frac{2}{\langle n_i\rangle}\frac{1}{\tau_i}. \tag{26}$$

This shows that within the approximation of the FDT, and assuming that one molecule is made at a time, the randomness introduced by probabilistic births and deaths of a certain species is inversely proportional to its average number of molecules. However, that does not mean that the resulting fluctuations have to be inversely proportional to the number of molecules, as that additionally depends on the dynamics and connections between the chemical species.

4.4. Examples of elasticities

A simple example will illustrate the principle behind elasticities:

$$\frac{d\langle n\rangle}{dt} = \lambda^+\langle n\rangle^\alpha - \lambda^-\langle n\rangle^\beta \quad \Rightarrow \quad H = \frac{\partial \ln(\lambda^-\langle n\rangle^\beta/\lambda^+\langle n\rangle^\beta)}{\partial \ln\langle n\rangle} = \beta - \alpha. \tag{27}$$

The elasticity thus corresponds to the difference in kinetic order between degradation and synthesis. For an unbiased and unbounded random walk, $H = 0$ and there is no stationary state. Changing from first order synthesis ($\beta = 1$) to irreversible dimerization ($\beta = 2$) can have a similar effect as changing from constitutive synthesis ($\alpha = 0$) to well-working 'hyperbolic' negative feedback control ($\alpha = -1$). Both mechanisms have been implicated in homeostatic systems [32].

Many multivariate systems are equally tractable, for example:

$$\begin{aligned} \frac{d\langle n_1\rangle}{dt} &= \lambda_1^+ - \langle n_1\rangle/\tau_1 \\ \frac{d\langle n_2\rangle}{dt} &= \lambda_2^+\langle n_1\rangle - \langle n_2\rangle/\tau_2 \end{aligned} \quad \Rightarrow \quad \mathbf{H} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \tag{28}$$

as can be seen in much the same way as in Eq. (27). If n_2 instead negatively affected its own synthesis and n_1 was an (unsaturated) enzyme that degraded n_2, one could have:

$$\frac{d\langle n_1\rangle}{dt} = \lambda_1^+ - \langle n_1\rangle/\tau_1 \qquad \Rightarrow \quad \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}. \tag{29}$$
$$\frac{d\langle n_2\rangle}{dt} = \lambda_2^+/\langle n_2\rangle - \lambda_2^-\langle n_1\rangle\langle n_2\rangle$$

The nonlinearities are only approximations. For example, if $R_2^- = \lambda_2^- n_1 n_2$ then $\langle n_1 n_2\rangle \approx \langle n_1\rangle\langle n_2\rangle$ in Eq. (29) implicitly assumes that fluctuations are negligible (see discussion below Eq. (11)).

The examples above only have one synthesis and degradation term per species. Each term is also a perfect power-law. For more complicated schemes the elasticities must be calculated explicitly using the definitions in Eqs. (21)–(24), but it is still often easy to eye-ball lower and upper bounds.

4.5. A generalized equation for dynamically disordered expression

Because the parameters above—average lifetimes, numbers of molecules and elasticities—are so physically direct and intuitive, the FDT can be solved for structural classes of random processes collectively without losing interpretability. The simplest possible generic model for disordered gene expression includes four exponential events, each adding or removing a single molecule:

$$
\begin{aligned}
n_1 &\xrightarrow{R_1^+(n_1)} n_1 + 1 \\
n_1 &\xrightarrow{R_1^+(n_1)} n_1 - 1 \\
n_2 &\xrightarrow{R_1^+(n_1,n_2)} n_2 + 1 \\
n_2 &\xrightarrow{R_1^-(n_1,n_2)} n_2 - 1
\end{aligned}
\quad \text{so that} \quad
\begin{aligned}
\frac{d\langle n_1\rangle}{dt} &= \langle J_1^+(n_1)\rangle - \langle J_1^-(n_1)\rangle \\
\frac{d\langle n_2\rangle}{dt} &= \langle J_2^+(n_1,n_2)\rangle - \langle J_2^-(n_1,n_2)\rangle
\end{aligned}
\quad \text{where } J_i^\pm = R_i^\pm. \tag{30}
$$

This is a pseudo-bivariate stochastic system, where the rates for production and degradation of the second species depend on the random state of the first, but not vice versa ($H_{12} = 0$). Using the FDT approach above, we can solve for all stationary processes that display fluctuations around a stable fixed point. Matrices $\mathbf{A}$ and $\mathbf{D}$ follow:

$$\mathbf{M} = -\begin{bmatrix} H_{11}/\tau_1 & 0 \\ H_{21}/\tau_2 & H_{22}/\tau_2 \end{bmatrix} \quad \text{and} \quad \mathbf{D} = \begin{bmatrix} 2/(\langle n_1\rangle\tau_1) & 0 \\ 0 & 2/(\langle n_2\rangle\tau_2) \end{bmatrix}. \tag{31}$$

Solving the FDT for the normalized variances $\boldsymbol{\eta}$ and rearranging the solution gives:

$$\eta_{22} = \frac{\sigma_2^2}{\langle n_2\rangle^2} \approx \overbrace{\underbrace{\frac{1}{\langle n_2\rangle}}_{\substack{\text{low-copy}\\\text{fluctuations}}} \times \underbrace{\frac{1}{H_{22}}}_{\substack{\text{effective}\\\text{stability}}}}^{\text{spontaneous or intrinsic } x_2 \text{ noise}} + \overbrace{\underbrace{\frac{\sigma_1^2}{\langle n_1\rangle^2}}_{\substack{\text{environmental}\\\text{fluctuations}}} \times \underbrace{\frac{H_{21}^2}{H_{22}^2}}_{\substack{\text{static}\\\text{susceptibility}}} \times \underbrace{\frac{H_{22}/\tau_2}{H_{22}/\tau_2 + H_{11}/\tau_1}}_{\substack{\text{one-step}\\\text{time-averaging}}}}^{\text{forced or extrinsic } x_2 \text{ noise}} \tag{32}$$

where $\eta_{11} = \sigma_1^2\langle n_1\rangle^{-2} \approx \langle n_1\rangle^{-1} H_{11}^{-1}$.

For the spontaneous or 'intrinsic' noise term, the first factor $1/\langle n_2\rangle$ corresponds to the average size of the random jumps (one molecule made at a time) relative to the average size of the population. The second factor can be interpreted in several ways. The normalized Jacobian matrix $\mathbf{M}$ in Eq. (31) has

eigenvalues $-H_{11}/\tau_2$ and $-H_{22}/\tau_2$. Because the matrix is triangular, where the environment n_1 affects the system n_2 but not vice versa, the eigenvalues are the adjustment rate constants for the two components: A perturbation in n_1 is corrected with rate constant H_{11}/τ_2 (see Eq. (7)), and a perturbation in n_2 is corrected with rate constant H_{22}/τ_2. Decreasing the average lifetimes τ increases the adjustment rates, but also increases the rate of spontaneous randomisation by the same factor—measuring time in a different unit cannot affect stationary fluctuations. Parameter H_{22} can thus be seen as the rate of adjusting to steady state normalised by the rate of deviating from steady state.

For another perspective on the same principle, consider the probability $P(birth)$ that the next event is a birth rather than a death:

$$P(birth) = \frac{J_2^+}{J_2^+ + J_2^-} = \frac{1}{1 + J_2^-/J_2^+}. \tag{33}$$

To have small fluctuations around the average $\langle n_2 \rangle$, $P(birth)$ should be low above the average and high below the average. This means that the ratio between death rates and birth rates should increase sharply with n_2. The sharpness is measured by the elasticity H_{22}: If n_2 changes by 1%, the ratio between the rates change approximately H_{22} per cent. The value of H_{22} thus measures the preference for returning to steady state over deviating further.

The forced or 'extrinsic' noise in Eq. (32) follows entirely different principles. The first factor, $\sigma_1^2/\langle n_1 \rangle^2$, simply measures the relative size of environmental fluctuations. The second factor shows how a permanent change in n_1 would eventually affect $\langle n_2 \rangle$. For example consider the following equation:

$$\frac{\partial \langle n_2 \rangle}{\partial t} = \lambda_2 n_1 - \beta_2 \langle n_2 \rangle^2. \tag{34}$$

Here we see how the average $\langle n_2 \rangle$ follows changes in n_1. If concentration n_1 changed from one fixed value to another, $\langle n_2 \rangle$ eventually adjusts to new steady states given by:

$$\langle n_2 \rangle = \sqrt{\lambda_2/\beta_2 n_1}. \tag{35}$$

A fourfold increase in n_1 then only produces a two-fold change in the stationary $\langle n_2 \rangle$, and $H_{21}/H_{22} = -1/2$. In general, we can formulate this as:

$$\frac{\partial \ln \langle n_2 \rangle}{\partial \ln n_1} = -\frac{H_{21}}{H_{22}} \tag{36}$$

where $\langle n_2 \rangle$ is evaluated at the steady state towards which the systems adjusts after n_1 changes value [20]. The two elasticities reflect two different principles. Parameter H_{21} measures how the level of n_1 affects the birth and death balance of n_2, while H_{22} measures how sensitive the eventual steady state of $\langle n_2 \rangle$ is to changes in its own birth and death rates.

When changes in n_1 are not permanent it is also necessary to account for the temporal response in n_2. As noted above, H_{11}/τ_1 and H_{22}/τ_2 are the rate constants for adjustments to steady state in components n_1 and n_2 respectively. The ratio:

$$\frac{H_{22}/\tau_2}{H_{11}/\tau_1} = \underbrace{\frac{1}{\tau_2/H_{22}}}_{\substack{\text{effective memory} \\ \text{of system } n_2}} \underbrace{\frac{1}{H_{11}/\tau_1}}_{\substack{\text{rate of change of} \\ \text{environment } n_1}} \tag{37}$$

is thus a measure for how closely n_2 follow changes in n_1. If n_2 adjusted swiftly, $H_{22}/\tau_2 \gg H_{11}/\tau_1$, the time-averaging factor would be close to one, and if n_2 adjusted slowly, the time-averaging factor would be close to zero. The adjustment rate is thus inversely related to the effective memory window: Fast systems are only influenced by the recent history and therefore more affected by environmental changes.

5. Measures and nomenclature

5.1. Measures of noise

Noise can be quantified in many ways. Autocorrelations conveniently summarize both the magnitude and frequency of fluctuations and are fairly tractable analytically. However, most models so far have focused on stationary averages and variances.

The results above were formulated in terms of the variance over squared average, which allows for a clean separation of different noise sources as long as the models are weakly nonlinear. Another common measure is the Fano factor—the variance over average—which equals one for Poisson distributions. But the comparison with the Poissonian only works well for univariate discrete random processes, where the variance often is proportional to the average with a proportionality constant that reflects the overall nature of the process. For multivariate random processes, the Poisson distribution holds no special position and using the Fano factor can be misleading. To illustrate this with a more extreme example than the bursts in Eq. (9), assume that protein fluctuations partly came from fluctuations in RNase concentrations. The variance over squared average would then contain a noise term that was more or less independent of transcription and translation rates. Multiplying by the average to obtain the Fano factor would thus force the measure to depend on anything that affected the average.

Eq. (32) suggests that no measure works well for all types of fluctuations—spontaneous fluctuations depend on the number of molecules while forced fluctuations do not. However, the variance over squared average is certainly a more suitable basis for experimental interpretations. First, in most experimental studies so far the average number of proteins per cell is too high to contribute substantial spontaneous fluctuations. Second, by plotting the variance over squared average as a function of the inverse average, any univariate scaling behavior is easily identified without introducing scaling problems for any extrinsic noise. Third, the relevance of a fluctuation typically depends on the normal size of the system. Because the variance is a second order moment it must then be normalized by the squared average—the Fano factor for the numbers of molecules per individual is more than 100,000 times higher for elephants than for mice.

5.2. Intrinsic vs extrinsic

Previous studies have used different classifications to highlight different aspects of noise. Here I focus on the distinction between 'intrinsic' and 'extrinsic' noise, and consider four aspects: The statistical nature of the fluctuations, how central a process is to gene expression, correlations between different proteins in the cell, and experimental strategies for measuring noise:

(1) From a physical viewpoint, the terms 'intrinsic' and 'extrinsic' have no specific meaning other than 'inside' and 'outside', and thus always depend on the definition of system versus environment. With

proteins as system, the first small-number term of Eq. (4) is intrinsic and the mRNA and gene terms are both extrinsic [18,22]. Other studies [7,20] define all three terms as intrinsic to distinguish them from the extrinsic fluctuations in the overall state of the cell. However, spontaneous small-number protein noise is principally different from the noise that comes from enslavement by genes or mR-NAs, but the latter two noise sources are *not* principally different from, e.g., ribosome-mediated noise that also enslaves proteins. Considering only the statistical nature of the fluctuations, there is thus no reason to label all three noise terms in Eq. (4) as intrinsic.

(2) From a biological viewpoint, noise sources could be classified according to how central the corresponding component is to gene expression. However, classifying fluctuations in gene activity as intrinsic and fluctuations in ribosomes as extrinsic would not separate the central parts of gene expression from more peripheral cell processes. The opposite would be more appropriate: Ribosomes are inherent to gene expression while spontaneous changes in gene activity can indirectly reflect regulation.

(3) Protein noise can also be classified according to correlations between different types of proteins. Some sources of randomness are shared broadly by many genes in a cell, while others are exclusive to a particular gene or a small set of genes. The most specific noise comes from having low protein numbers, originating in the probabilistic births and deaths of individual molecules. Spontaneous mRNA fluctuations are also quite specific, though some transcripts encode several different proteins. Operator fluctuations are less specific yet and typically affect all genes in an operon. Many DNA and RNA binding proteins act as repressors and activators, some of which are specific to a particular gene and others which regulate large classes of genes. This can form complicated correlation structures. For example, two genes may be targeted by the same repressor, yet their mRNAs may be degraded by different RNases. A few central factors are shared more universally, including ribosomes, core polymerases, tRNAs, amino acids etc. However, different proteins are still differently affected by such fluctuations because they have different sequences and lifetimes.

From this perspective, the gene-mRNA-protein terms in Eq. (4) could all be intrinsic, taking intrinsic to mean specific. However, it should then be emphasized that such distinctions can be made in many different ways. The mRNA term is in some sense more intrinsic than the gene term, but the gene term is not necessarily more intrinsic than fluctuations in repressor concentrations.

(4) Finally, the classifications can be tailored to the available experimental methods. A particularly clever strategy for separating noise sources is based on correlations between the expression of two physically separate but identically regulated fluorescent reporter genes [7,9]. If the two proteins are kinetically independent, i.e., if the mechanisms are linear, the normalized covariance between them then equals the sum of all shared noise terms [20]. Fluctuations in ribosomes, polymerases, RNases etc thus all end up in the 'extrinsic' covariance category, while the gene-mRNA-protein terms in Eq. (4) are 'intrinsic' because each *gfp* has its own operators and transcripts. The appeal of this approach is not only that some noise sources are separated from others, but also that the separation to some extent relates to specificity (see (3) above). The only risk is that the terms are over-interpreted, construing biological or physical meaning where none is intended: The 'intrinsic noise' only partially relates to specificity, as demonstrated by experiments where most extrinsic noise came from a repressor that was specific to that particular gene [7]. Realizing that the distinction largely is a side-effect of the experimental set-up in turn opens the doors for other applications. If the two reporters were placed under the same operator, operator fluctuations would become shared between the two genes and thus move from the intrinsic to the extrinsic category. If the reporters were encoded on

the same transcript, the mRNA term would also become extrinsic. If they were regulated by different repressors, repressor fluctuations would move from the extrinsic to the intrinsic category. And so on. The dual reporter strategy does not separate fluctuations based on a priori physical or biological principles. It is much more general and useful than that.

6. Analytical models of protein noise, 1977–2004

Detailed models of stochastic gene expression have been published for almost 30 years. Here I only discuss some analytical studies of constitutive expression (feedback analyses excluded), possibly ignoring others that deserve equal mention (if so please contact me so that the mistake can be corrected).

In the 1970s, David Rigney [10,12–14], Otto Berg [11] and co-workers pioneered the field in a series of papers that predicted mRNA and protein fluctuations in growing cells. One of their take-home messages was that mRNAs produce geometrically distributed translation bursts if they are either translated or degraded with constant probabilities, and that this in turn widens the protein distributions. Technically, the models include the mRNA and protein part of Eq. (4), but also account for deterministic cell growth, chromosome replication at a fixed time point in the cell cycle, and partitioning of protein copies at cell division. These insightful studies were carefully motivated and argued, and though some of the analytical expressions were rather complicated, the papers are as rewarding to read as any of the more recent analyses. They only failed to initiate a continuous tradition because they were 20 years before their time.

In 1987, Tapaswi et al. [15] considered a stochastic model for a biological oscillator, assuming that a transcript effectively inhibited its own gene by kicking out a gene activator. Disregarding the feedback term, their model was identical to the gene-mRNA model above and they used probability generating functions to calculate the marginal distribution of genes and mRNAs, but did not emphasize conditions when fluctuations are large. In 1995, Peccoud and Ycart [16] used the same mathematical model to describe genes and proteins, assuming that mRNAs are degraded rapidly. They used moment equations to derive exact results for how averages and variances depend on kinetic parameters, predicting that genes that slowly and randomly switch on and off produce large relative protein fluctuations. In 2001, Kepler and Elston [17] revisited the results for the gene-protein part of the model, arriving at the same results as Peccoud and Ycart, but also analyzing regulated systems in terms of bifurcations, escape times etc.

In 2000–2001, Paulsson et al. [33] slightly extended the mRNA-protein model by calculating the stationary distribution for geometrically distributed translation bursts. They also formulated the variance over squared average in terms of internal and time-averaged external fluctuations [18], as in Eq. (8). At the same time, the same model was independently analyzed by Thattai and van Oudenaarden [19] who calculated time-dependent moments and emphasized translation bursts, but also treated autoregulated systems as well as noise propagation in longer expression cascades [34]. In 2002, Swain et al. [20] presented a model that accounted for several steps in transcriptional initiation, as well as replication, cell growth and division. This was more complete than previous studies, but beneath the details the model was close to identical to Eq. (4). The apparent discrepancy comes from the fact that the gene noise was presented as a part of the small-number mRNA noise. The authors also motivated the dual reporter correlation strategy mathematically.

In 2003, Sasai and Wolynes [21] mapped stochastic gene expression onto the quantum many-body problem, presenting a different way of deriving the gene-protein model above as well as analyzing systems with multiple attractors. In 2004, Paulsson interpreted the Fluctuation–Dissipation Theorem in

terms of average numbers of molecules, lifetimes, and elasticities, and derived Eq. (32) to reinterpret some experimental findings. Raser and O'Shea [9] published results that are mathematically identical to Eq. (4), but viewed from a different perspective. More papers are coming out by the minute so this review will most likely be out of date before it is printed.

Stochastic gene expression has also been frequently modeled using Monte Carlo simulations, perhaps most notably by McAdams and Arkin [35] who in 1997 introduced the basic notions to a broader audience. Their conclusions were similar to those of earlier analytical works, for example emphasizing that protein synthesis may occur in geometrically distributed bursts. But by emphasizing biological implications they provided a much needed rallying point for subsequent studies.

7. Future directions and closing remarks

The analysis above only shows that most models have focused on the same phenomenon, not that this phenomenon really captures gene expression in living cells. Most experimental results are certainly consistent with the standard model, sometimes requiring slight modifications, but qualitatively different models can be equally consistent with the same data. A few idealizations above are particularly notable:

First, there are no strong indications that genes, RNAs and proteins are the critical molecules that contribute small-number fluctuations. For example, steps in the elongation phases of transcription and translation may also fluctuate greatly. It has been suggested that the anabolic nature of protein synthesis, joining together smaller subunits, generates enormous fluctuations in the levels of charged tRNAs [27]. How would this affect fluctuations in protein concentrations? The focus on genes, RNAs and proteins may very well turn out to be a purely sociological choice, reflecting how cartoons are drawn rather than actual physical properties.

Second, many enzyme and substrate concentrations are statistically correlated. Such correlations may even have evolved to suppress total protein fluctuations. For example assume that a gene for a certain protein is transcribed by a certain sigma factor. Further assume that the corresponding mRNA is degraded by an RNase that is transcribed by the same sigma-factor. A random increase in the concentration of the sigma factor then increases both the synthesis and degradation rates of the mRNA, and the two effects could partially cancel out. Other correlations could have the opposite effect and increase fluctuations instead.

Third, the discrete probabilistic events are assumed to be exponential, yet we know that gene activation, transcription and translation consist of numerous small steps. Does this qualitatively change the nature of the fluctuations in concentrations?

All these issues will undoubtedly be addressed in the next few years, both mathematically and experimentally. However, in addition to solving the different individual problems, the community as a whole also faces the collective challenge of converging towards a coherent discipline. Detailed studies are definitely useful to ensure that the supposed mechanisms of a certain system are consistent with the idealizations we use to understand them. But without idealized models and generic theory to begin with, details are unintelligible. Analytical theory is not automatically better than simulations though. Simply following mathematical rules and deriving complicated algebraic expressions could even prevent progress, making problems appear understood when they are not. However, these are the teething troubles of any new discipline. Given the exploding experimental progress there is little doubt that stochastic gene expression is establishing itself as one of the most central and exciting problems in molecular biology.

Acknowledgement

References

[1] Savageau MA. Biochemical systems analysis. Reading: Addison–Wesley; 1976.

[2] Savageau MA. Comparison of classical and autogenous systems of regulation in inducible operons. Nature 1974;252:546–9.

[3] Løbner-Olesen A. Distribution of minichromosomes in individual Escherichia coli cells: implications for replication control. EMBO J 1999;18:1712–21.

[4] Paulsson J, Ehrenberg M. Noise in a minimal regulatory network: plasmid copy number control. Quart Rev Biophys 2001;34:1.

[5] Becskei A, Serrano L. Engineering stability in gene networks by autoregulation. Nature 2000;405:590–3.

[6] Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. Regulation of noise in the expression of a single gene. Nature Genet 2002;31:69–73.

[7] Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science 2002;297:183–6.

[8] Blake WJ, Kærn M, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. Nature 2003;422:633–7.

[9] Raser JM, O'Shea EK. Control of stochasticity in eukaryotic gene expression. Science 2004;304:1811.

[10] Rigney DR, Schieve WC. Stochastic model of linear, continuous protein-synthesis in bacterial populations. J Theor Biol 1977;69:761–6.

[11] Berg OG. A model for statistical fluctuations of protein numbers in a microbial-population. J Theor Biol 1978;73:307.

[12] Rigney DR. Stochastic model of constitutive protein levels in growing and dividing bacterial cells. J Theor Biol 1979;76:453–80.

[13] Rigney DR. Note on the kinetics and Stochasticity of induced protein synthesis as influenced by various models for messenger RNA degradation. J Theor Biol 1979;79:247–57.

[14] Rigney DR. Stochastic models of cellular variability. Kinetic logic—a Boolean approach to the analysis of complex regulatory systems. In: Lecture Notes in Biomath, vol. 29. Berlin: Springer-Verlag; 1979 [Chapter XIII].

[15] Tapaswi PK, Roychoudhury RK, Prasad T. A stochastic model of gene activation and RNA synthesis during embryogenesis. Sankhya Ind J Stat 1987;49:51–67.

[16] Peccoud J, Ycart B. Markovian modelling of gene-product synthesis. Theor Popul Biol 1995;48:222.

[17] Kepler TB, Elston TC. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophys J 2001;81:3116–36.

[18] Paulsson J, Ehrenberg M. Noise in a minimal regulatory network: Plasmid copy number control. Quart Rev Biophys 2001;34:1.

[19] Thattai M, van Oudenaarden A. Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci USA 2001;98:8614.

[20] Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc Natl Acad Sci USA 2002;99:12795.

[21] Sasai M, Wolynes PG. Stochastic gene expression as a many-body problem. Proc Natl Acad Sci USA 2003;100:2374–9.

[22] Paulsson J. Summing up the noise in gene networks. Nature 2004;427:415–9.

[23] Gardiner CW. Handbook of stochastic methods for physics, chemistry and the natural sciences. 3rd ed. Berlin/New York: Springer-Verlag; 2004.

[24] Elf J, Paulsson J, Berg OG, Ehrenberg M. Near-critical phenomena in intracellular metabolite pools. Biophys J 2003;84:154–70.

[25] Zwanzig R. Nonequilibrium statistical mechanics. New York: Oxford Univ. Press; 2001.

[26] Xie XS. Single-molecule approach to dispersed kinetics and dynamic disorder: Probing conformational fluctuation and enzymatic dynamics. J Chem Phys 2002;117:11024–32.

[27] Elf J, Nilsson D, Tenson T, Ehrenberg M. Selective charging of tRNA isoacceptors explains patterns of codon usage. Science 2003;300:1718–22.

[28] Keizer J. Statistical thermodynamics of nonequilibrium processes. Berlin: Springer-Verlag; 1987.

[29] Lax M. Fluctuations from nonequilibrium steady state. Rev Modern Phys 1960;32:25–64.

[30] Savageau MA. Parameter sensitivity as a criterion for evaluating and comparing the performance of biochemical systems. Nature 1971;229:542–4.

[31] Kacser H, Burns JA. The control of flux. Symp Soc Exp Biol 1973;27:65–104.

[32] Das N, Valjavec-Gratian M, Basuray AN, Fekete RA, Papp PP, Paulsson J, Chattoraj D. Multiple homeostatic mechanisms in the control of P1 plasmid replication. Proc Natl Acad Sci USA. In press.

[33] Paulsson J, Berg OG, Ehrenberg M. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. Proc Natl Acad Sci USA 2000;97:7148–53.

[34] Thattai M, van Oudenaarden A. Attenuation of noise in ultrasensitive signaling cascades. Biophys J 2002;82:2943–50.

[35] McAdams HH, Arkin A. Stochastic mechanisms in gene expression. Proc Natl Acad Sci USA 1997;94:814–9.

[36] Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem 1977;81:2340.