Progress in Biophysics and Molecular Biology 100 (2009) 57–66

Contents lists available at ScienceDirect

Progress in Biophysics and Molecular Biology

journal homepage: www.elsevier.com/locate/pbiomolbio

Review

Signatures of gene expression noise in cellular systems

Julia Rausenberger [a,b,*], Christian Fleck [b,c], Jens Timmer [c,d], Markus Kollmann [e]

[a] Faculty of Biology, University of Freiburg, Schänzlestr. 1, 79104 Freiburg, Germany
[b] Center for Biological Systems Analysis (ZBSA), Habsburgerstr. 49, 79104 Freiburg, Germany
[c] Institute of Physics, University of Freiburg, Hermann-Herder-Str. 3, 79104 Freiburg, Germany
[d] Freiburg Institute for Advanced Studies (FRIAS), Albertstr. 19, 79104 Freiburg, Germany
[e] Institute for Theoretical Biology, Humboldt-University Berlin, Invalidenstr. 43, 10115 Berlin, Germany

ARTICLE INFO

Article history:
Available online 11 June 2009

Keywords:
Gene expression noise
Stochastic modeling
Transcription factor
Population distribution

ABSTRACT

Noise in gene expression, either due to inherent stochasticity or to varying inter- and intracellular environment, can generate significant cell-to-cell variability of protein levels in clonal populations. To quantify the different sources of gene expression noise, several theoretical studies have been performed using either a quasi-stationary approximation for the emerging master equation or employing a time-dependent description, when cell division is taken explicitly into account. Here, we give an overview of the different origins of gene expression noise which were found experimentally and introduce the basic stochastic modeling approaches. We extend, and apply a time-dependent description of gene expression noise to experimental data. The analysis shows that the induction level of the transcription factor can be employed to discriminate the noise profiles and their characteristic signatures. On the basis of experimentally measured cell distributions, our simulations suggest that transcription factor binding and promoter activation can be modeled independently of each other with sufficient accuracy.

Contents

Abbreviations: G/C/YFP, green/cyan/yellow fluorescent protein; TF, transcription factor; URN, uniform random number.

* Corresponding author at: Center for Biological Systems Analysis, Habsburgerstr. 49, 79104 Freiburg, Germany. Tel.: +49 (0) 761/203 97199.
E-mail address: juhe@fdm.uni-freiburg.de (J. Rausenberger).

0079-6107/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pbiomolbio.2009.06.003

1. Introduction

Stochasticity is ubiquitous in biological systems. In genetic networks random fluctuations are inevitable as each step leading towards gene expression relies upon random encounters between molecules. Moreover, the molecules participating in each of these probabilistic chemical reactions often are present in very low numbers which contribute to the final state of the system. Therefore, within a genetically identical population, individual cells show significant phenotypic heterogeneity (Avery, 2006; Raser and O'Shea, 2005; Spudich and Koshland, 1976). This variability directly affects the cell's ability to respond to environmental factors like changes in ligand concentration. Due to the progress in experimental technologies to use expression reporters in single cells, it is possible to examine the expression variability in living cells (Ko et al., 1990). A lot of effort has been undertaken to quantify the origins of gene expression noise experimentally (Bar-Even et al., 2006; Becskei et al., 2005; Blake et al., 2003; Cai et al., 2006; Colman-Lerner et al., 2005; Elowitz et al., 2002; Golding et al., 2005; Mettetal et al., 2006; Newman et al., 2006; Ozbudak et al., 2002; Pedraza and van Oudenaarden, 2005; Raj et al., 2006; Raser and O'Shea, 2004; Rosenfeld et al., 2005; Sigal et al., 2006; Volfson et al., 2006) and theoretically (Berg, 1978; McAdams and Arkin, 1997; Paulsson, 2004, 2005; Pedraza and Paulsson, 2008; Rausenberger and Kollmann, 2008; Swain, 2004; Swain et al., 2002; Thattai and van Oudenaarden, 2001; Volfson et al., 2006).

Stochasticity or noise inherent to gene expression seems to be one of the main driving forces for the observed cell-to-cell variability in clonal populations. A general framework to describe the time evolution of such a stochastic system most accurately is given by a chemical master equation (van Kampen, 1992), which determines the probability for specific molecular population at a time. The master equation determines the entire probability distribution of the molecular system. The chemical reactions themselves are viewed as distinct and instantaneous physical events in a well-stirred system (Gillespie, 2007). There are no general methods to solve the chemical master equation, but several approximations have been suggested to infer characteristics of the behavior of the system (Gardiner, 1990; van Kampen, 1992). In the case that fluctuations are negligible, the master equation can be approximated by a macroscopic ordinary differential equation for the average value of the involved chemical species. This mean-field approximation has been successfully used in several studies, e.g., to quantify cell-cycle averaged inter-cellular variability (Paulsson, 2005; Swain et al., 2002; Thattai and van Oudenaarden, 2001) or to derive analytical expressions for the distributions of the involved populations (Friedman et al., 2006; Shahrezaei and Swain, 2008). Recently, time-resolved experiments on single cells provided insight into protein levels and noise strengths at every state of the cell cycle (Sigal et al., 2006). To extract information from these non-equilibrium protein synthesis trajectories, a cell-cycle averaged description becomes invalid and a time-dependent description using the master equation approach has been suggested (Rausenberger and Kollmann, 2008; Swain et al., 2002).

The typical model for gene expression is described by a *one-gene-system*, consisting of promoter activation/inactivation, transcription, and translation (Paulsson, 2005; Swain et al., 2002). In addition, transcription factor binding to specific DNA regions can trigger or inhibit the activation of the gene under consideration and has been investigated by several authors (Bintu et al., 2005; Buchler et al., 2003; Gerstung et al., 2009; Pulkkinen and Berg, 2008). The introduced non-linearity, either due to multiple binding sites (Gerstung et al., 2009) or to autoregulation of the TF (Pulkkinen and Berg, 2008), leads to more complex dynamics of the cellular system. In the following we assume that the transcription factor (TF) acts as an activator such that transcription can start if and only if the TF is bound and the promoter is in its active state (Fig. 1).

The cell-to-cell variability of a specific protein in a large clonal population with fixed generation time can be separated into two distinct contributions, *intrinsic* and *extrinsic* noise (Swain et al., 2002). All gene specific, stochastic events contribute to the intrinsic noise. Differences between cells, either in global cellular state or in the concentration or activity of any factor that affects gene expression are referred to as extrinsic noise (Elowitz et al., 2002).

Based on the time-dependent description of gene expression (Rausenberger and Kollmann, 2008), the present work shows that the magnitudes of the different noise contributions depend strongly on the induction level of the TF, the synthesis rates, and the molecule lifetimes associated with each individual gene providing possible explanations for diverging experimental results as demanded by Kaufmann and van Oudenaarden (2007). As an example, we focus on differences in the induction level of a TF, e.g., due to different experimental set-ups, which lead to different signatures of gene expression noise even in the same organisms. Furthermore, on the basis of experimentally measured cell distributions of wild type and over-expressed cells of *Escherichia coli*, our simulations propose that, in prokaryotes, TF binding and promoter activation are independent of each other and thus can be modeled to good approximation separately.

2. Different origins of gene expression noise

In a first experiment on the different sources of gene expression noise, Elowitz et al. (2002) introduced two copies of the same promoter into the genome of *E. coli*, driving the expression of YFP and CFP proteins. Extrinsic fluctuations were quantified when the expression of both copies was affected, whereas intrinsic fluctuations were specific to one of the copies alone. Several subsequent experiments have measured the variance in protein abundances in different cellular systems (Bar-Even et al., 2006; Becskei et al., 2005; Blake et al., 2003; Colman-Lerner et al., 2005; Golding et al., 2005; Mettetal et al., 2006; Newman et al., 2006; Pedraza and van Oudenaarden, 2005; Raj et al., 2006; Raser and O'Shea, 2004; Rosenfeld et al., 2005; Sigal et al., 2006; Volfson et al., 2006). Considerable confusion stems from diverging experimental results which have identified different origins for the main contribution to

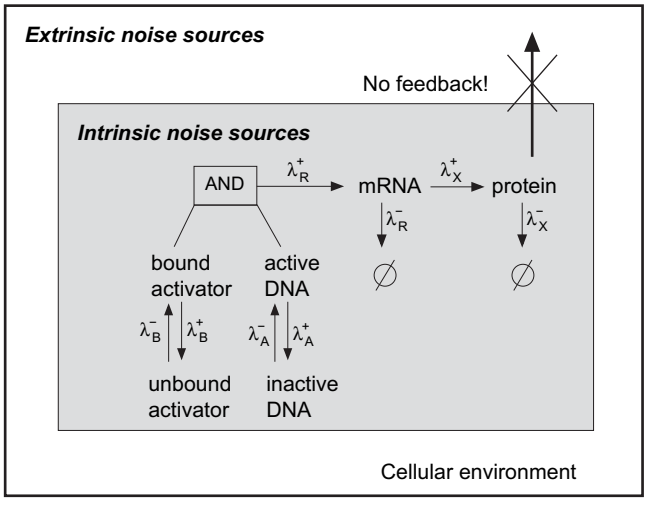

Fig. 1. Definition and reaction scheme of single-gene reporter system (shaded box) within intracellular environment (large box). Intrinsic and extrinsic noise can only be distinguished if expression level of reporter system does not influence extrinsic factors. Transition rates are defined in the text.

gene expression noise, such that a complete picture is still missing (Kaufmann and van Oudenaarden, 2007). Moreover, the cell-to-cell variations arising from gene expression noise are present in all investigated organisms ranging from microbes to humans (Raj and van Oudenaarden, 2008; Sigal et al., 2006). For prokaryotes, translational efficiency was identified as the main source of variability of expression levels consistent with a stochastic model in which proteins are produced in sharp and random bursts (Ozbudak et al., 2002). However, later experimental observations in individual living cells either by measuring mRNA levels or by real-time observations at single-molecule level indicated that promoter activation predominantly causes gene expression noise (Cai et al., 2006; Golding et al., 2005). Furthermore, extrinsic factors, like the cellular state, were also identified to give the main contribution to phenotypic variations in a clonal population (Rosenfeld et al., 2005). Similar diverging results have been found in eukaryotes, where in the budding yeast *Saccharomyces cerevisiae* a two-reporter system, expressing two fluorescent proteins from identical promoters, identified switching between active and inactive promoter states due to slow stochastic chromatin-remodeling events as the by far largest source of noise (Raser and O'Shea, 2004). In later experiments it was shown for a large set of genes at their native expression levels that the noise has a clear sign of transcriptional origin due to low-copy mRNA molecules (Bar-Even et al., 2006; Newman et al., 2006). Moreover, a direct monitoring of mRNA production from a gene at the resolution of single molecules in mammals revealed strong mRNA bursts dominating gene expression noise (Raj et al., 2006). For human cells, genes at native induction level showed significant noise contribution from long-term variations of the cellular state (Sigal et al., 2006). At first sight it seems that no general rule can be given to determine the main sources of gene expression noise. Protein levels, however, should be strongly regulated to serve for precise and reliable information processing within a cell. Any significant deviation from the optimal level would result in reduction of fitness and an evolutionary disadvantage. Thus, random fluctuations are in most cases detrimental for cellular systems and several regulatory mechanisms have evolved to minimize them. E.g., negative feedback loops have been shown both theoretically (Thattai and van Oudenaarden, 2001) and experimentally (Austin et al., 2006; Becskei and Serrano, 2000; Dublanche et al., 2006) to reduce the fluctuations around the mean, since fluctuations below and above the mean are pushed back toward the mean via the feedback. Only in rare cases noise can be used to drive phenotypic switching providing a non-genetic mechanism to population heterogeneity, as found for bacterial persistence against antibiotics (Balaban et al., 2004) and competence for DNA uptake from the environment (Suel et al., 2006).

3. Stochastic modeling approaches

In order to track down the individual contributions of the molecular mechanisms involved in protein synthesis several mathematical models have been introduced (Berg, 1978; Friedman et al., 2006; McAdams and Arkin, 1997; Paulsson, 2004, 2005; Rausenberger and Kollmann, 2008; Swain, 2004; Swain et al., 2002; Thattai and van Oudenaarden, 2001; Volfson et al., 2006). Four major steps are involved in a generic model of *intrinsic* fluctuations in gene expression in living cells: (i) TF binding (repressor unbinding), (ii) promoter (DNA) activation, (iii) transcription, and (iv) translation (see Fig. 1). TF binding and promoter activation are described as a random telegraph process, because they are assumed to switch randomly between zero and one with exponentially distributed waiting times (Golding et al., 2005; Raj et al., 2006). The state of the TF is given by the stochastic variable $B(t)$ switching between $B(t) = 1$ and $B(t) = 0$, if the TF is bound or unbound, respectively. Promoter activation $A(t)$ can be expected to occur on much slower time scales than TF binding (Alon, 2006; Cai et al., 2006; Elf et al., 2007; Golding et al., 2005) such that the time scales can be separated. The switching rate from the inactive to the active state of the promoter is denoted by λ_A^+, and λ_A^- represents the rate for the inverse process. Furthermore, the process of promoter activation does not seem to have any significant correlations with the cell cycle (Raj et al., 2006) leading to a stationary solution for the auto-correlation function $\langle A(t)A(t')\rangle = \mathrm{var}(A)e^{-(\lambda_A^+ + \lambda_A^-)|t-t'|}$ with $\mathrm{var}(A) = \lambda_A^+\lambda_A^-/(\lambda_A^+ + \lambda_A^-)$. Synthesis and degradation of mRNA and protein are typically modeled by a *birth-and-death* process (Kœrn et al., 2005), where the production probabilities λ_R^+ and λ_X^+ per time unit are proportional to the number of active genes and mRNAs. The degradation events for the mRNA (λ_R^-) and proteins (λ_X^-) are often assumed to be independent of each other, such that mRNA and proteins have exponentially distributed lifetimes. Connecting the stochastic processes in series, the time-dependent moments of the master equation can be calculated using the approach of generating functions (Swain et al., 2002; Thattai and van Oudenaarden, 2001). If the cell-cycle stage can be neglected, the levels of mRNA and proteins can be assumed to be at their steady-state values, such that cell-cycle averaged distributions for the mRNA and protein levels and the normalized stationary variances can be derived (Paulsson, 2004; Swain et al., 2002; Thattai and van Oudenaarden, 2001).

The assumption of an equilibrated system with small fluctuations around the mean becomes invalid, if cell division is included (Berg, 1978; Swain et al., 2002; Volfson et al., 2006). In accordance with recent experiments (Golding et al., 2005; Rosenfeld et al., 2005), we assume symmetric cell division and a binomial distribution of the molecules. The mathematical description of gene expression becomes more complex and the approximation via linearization fails, due to the introduction of another important time scale into the system, the generation time T_G. However, the specific intrinsic noise levels can be deduced at every stage of the cell cycle and can be compared to time-resolved experimental data (Rausenberger and Kollmann, 2008).

A generic approach to describe *extrinsic* fluctuations in the cellular environment and their effect on the time evolution of a generalized protein production rate is given by an Ohrnstein–Uhlenbeck process (Rausenberger and Kollmann, 2008; Sigal et al., 2006). The generalized protein production rate comprises all factors involved in gene expression as well as the global cellular state, and is itself expected to be subject to stochastic fluctuations. The stationary auto-correlation function of the Ohrnstein-Uhlenbeck process is, as well as the auto-correlation function for promoter activation which was described by a random telegraph process, given by an exponentially decaying function, which only depends on the time differences (Gardiner, 1990). Therefore, these two processes cannot be distinguished on the basis of stationary cell-to-cell distributions alone (Fig. 2A), exhibiting the same variance and relaxation time. If, however, time-resolved expression data is also included into the analysis, the effect of bursts and dead-times in mRNA synthesis, which contributes to the intrinsic noise, can be discriminated from continuously varying synthesis rates arising from extrinsic factors, Fig. 2B,C (Rausenberger and Kollmann, 2008).

4. Determination of noise signatures

4.1. Partial contributions to intrinsic gene expression noise

The heterogeneity in gene expression of a population can be quantified using the standard deviation σ divided by the mean μ, i.e., $\eta = \sigma/\mu$. The quantity η is commonly denoted as noise and

provides a physiologically relevant measure of gene expression variability as it quantifies relative fluctuations independent of the expression level. Two main contributions to the overall variance, σ^2_{tot}, determine the cell-to-cell variability of the amount of a protein X: the inherent stochasticity of gene expression stemming from the underlying processes (i)–(iv) is denoted as *intrinsic* variance, σ^2_I, which originates, e.g., from the different time of transcription in different cells, and is distinctive for each gene in its genomic context. The *extrinsic* variance, σ^2_E, is independent of a specific gene and acts on multiple genes in the same way. Fluctuations of upstream factors, which drive expression directly, like a given TF concentration, and the global cellular state, e.g., RNA-polymerase and ribosome concentration, are only few examples of the extrinsic variables on gene expression (Swain et al., 2002). If we do not assume any significant feedback of the expressed protein on extrinsic factors (cf. Fig. 1 and Tănase-Nicola et al., 2006), the overall variance in a large clonal population of cells with fixed generation time T_G sums up to $\sigma^2_{tot} = \sigma^2_I + \sigma^2_E$. The underlying stochastic processes of the generic model are connected in series and hence allow for assignment of the individual noise contributions to the *intrinsic noise* of protein synthesis:

$$\sigma^2_I = \left\langle \left(X(t) - \langle X(t) \rangle_{X,R,A,B} \right)^2 \right\rangle_{X,R,A,B} = \sigma^2_B + \sigma^2_A + \sigma^2_R + \sigma^2_X, \qquad (1)$$

where the average overall possible trajectories of protein copy number $X(t)$, mRNA copy number $R(t)$, promoter activation $A(t)$, and TF binding $B(t)$ is defined by $\langle \, . \, \rangle_{X,R,A,B}$. The average $\langle X \rangle_X = \sum_{X=0}^{\infty} XP(X,t|R(t))$ is conditionally dependent on the trajectory of mRNA synthesis, $R(t)$, which in turn depends on $A(t)$ and $B(t)$. Here, $P(X, t|R(t))$ denotes the probability density to observe the protein copy number X at time t given $R(t)$. The right-hand-side of Eq. (1) denotes the sum over variances corresponding to the processes of TF binding (σ^2_B), promoter activation (σ^2_A), transcription (σ^2_R), and translation (σ^2_X), respectively (Rausenberger and Kollmann, 2008). Summing up of the individual variances is only possible if there are no feedbacks from downstream to upstream processes (Tănase-Nicola et al., 2006). Otherwise there could exist a stochastic dependence of $R(t)$ on $X(t)$ making the additive form of Eq. (1) invalid (Colman-Lerner et al., 2005). We find for the noise contribution due to translation: $\eta^2_X = \sigma^2_X / \langle X \rangle^2_{X,R,A,B} = 1/\langle X \rangle_{X,R,A,B}$. Also, the binomial distribution of the proteins caused by cell division converges quite rapidly to a Gaussian distribution for an increased amount of molecules. The amount of proteins synthesized per mRNA can be estimated to be of the order 10^3 (Bar-Even et al., 2006), such that the protein copy number per cell is of the order 10^3–10^6 (Sigal et al., 2006) for the systems considered in this work. Thus, we can neglect translational noise in comparison to other noise contributions. To derive analytical expressions we assume a fixed induction level of the TF, $B(t) = B_{\mathrm{eq}}$. Due to the fast TF binding and very slow promoter activation, fluctuations from TF binding can be neglected compared to fluctuations from promoter activation and transcription such that the averaging over B cancels out in Eq. (1). Therefore, noise contribution from TF binding, σ^2_B, is not present. We introduce the average acceleration of protein synthesis in absence of mRNA and protein degradation, $\mathcal{A} = \langle A \rangle B_{\mathrm{eq}} \lambda^+_R \lambda^+_X$, where the mean $0 \leq \langle A \rangle = \lambda^+_A / (\lambda^+_A + \lambda^-_A) \leq 1$ can be interpreted as a measure of the fraction a promoter spends in its active state. If we account for cell division with generation time T_G (Rausenberger and Kollmann, 2008), the results can be directly compared to time-resolved expression data of protein levels. The

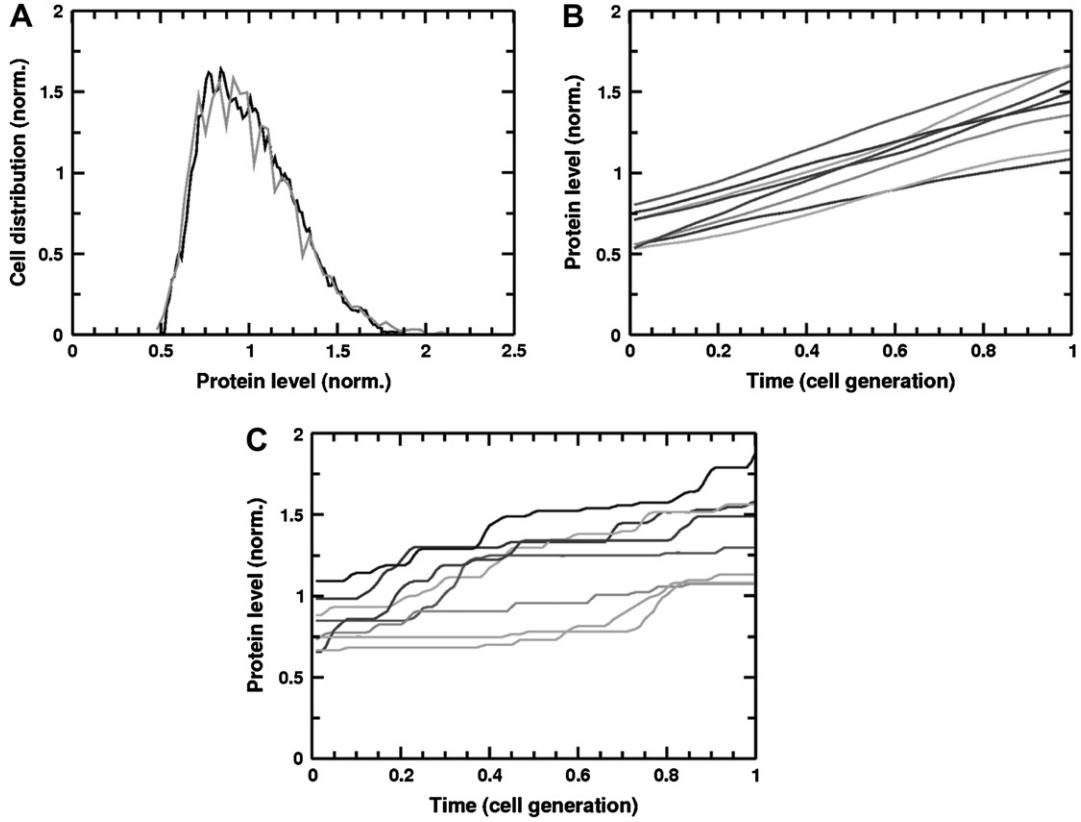

Fig. 2. A) Distribution of normalized proteins levels, sampled overall measured time points within the cell cycle of *in silico* generated trajectories assuming all noise arising from promoter activation (solid line) or from extrinsic factors (shaded line). B) Simulated trajectories using exclusively extrinsic noise contributions or C) intrinsic noise contributions stemming from promoter activation. Adapted from Rausenberger and Kollmann (2008).

derived asymptotic expressions for the mean and the variances are valid for any time t within a given cell cycle, $0 \leq t - t_0 \leq T_G$, where t_0 denotes the time point of the last cell division. The most likely physiological case consists of short mRNA lifetimes and long protein lifetime, $(\lambda_R^-)^{-1} \ll T_G \ll (\lambda_X^-)^{-1}$. If we also consider a fixed TF induction level B_{eq}, we obtain for $t - t_0, T_G \gg (\lambda_R^-)^{-1}$ the following asymptotic expressions for the average amount of protein and mRNA (Rausenberger and Kollmann, 2008):

$$\langle X(t-t_0)\rangle_{X,R,A} = \frac{\mathcal{A}}{\lambda_R^-}\left[t - t_0 + T_G\right] \tag{2}$$

$$\langle R(t-t_0)\rangle_{X,R,A} = \langle A\rangle B_{\mathrm{eq}}\,\frac{\lambda_R^+}{\lambda_R^-}. \tag{3}$$

Eq. (2) implies a linear increase of the mean amount of protein in time, whereas the mRNA level is recovered immediately after cell division and can be interpreted to be stationary (Eq. (3)). The noise contributions from transcription and promoter activation are also time-dependent and read in the limit $t - t_0, T_G \gg (\lambda_R^-)^{-1}, \gamma_A^{-1}$:

$$\sigma_R^2(t-t_0) = 2\mathcal{A}\,\frac{\lambda_X^+}{\left(\lambda_R^-\right)^2}\left[t - t_0 + \frac{1}{3}T_G\right] \tag{4}$$

$$\sigma_A^2(t-t_0) = 2\mathcal{A}^2\,\frac{\mathrm{var}(A)}{\langle A\rangle^2}\,\frac{1}{\left(\lambda_R^-\right)^2\gamma_A}\left[t - t_0 + \frac{1}{3}T_G\right]. \tag{5}$$

For the case that both protein and mRNA lifetimes are significantly shorter than the generation time, memory over generations is eliminated and both the mRNA and protein levels are recovered immediately after cell division, leading to the stationary, cell-cycle averaged description of gene expression described by the mean-field approximation. Therefore, the solutions are time-independent and agree with those found earlier by Paulsson (2004, 2005) for the mean protein level $\langle X\rangle_{X,R,A} = \mathcal{A}/(\lambda_R^-\lambda_X^-)$, mean mRNA level $\langle R\rangle_{X,R,A} = \langle A\rangle B_{\mathrm{eq}}\lambda_R^+/\lambda_R^-$, and the variances

$$\sigma_R^2 = \frac{\mathcal{A}\lambda_X^+}{\lambda_X^-\lambda_R^-\left(\lambda_R^- + \lambda_X^-\right)} \tag{6}$$

$$\sigma_A^2 = \frac{\mathcal{A}^2\mathrm{var}(A)}{\langle A\rangle^2\lambda_R^-\lambda_X^-\left(\gamma_A + \lambda_R^-\right)\left(\lambda_R^- + \lambda_X^-\right)}\left(1 + \frac{\lambda_R^-}{\gamma_A + \lambda_X^-}\right). \tag{7}$$

4.2. Noise regimes account for different experimental observations

Different origins of noise have been proposed and measured by several experimental groups (Bar-Even et al., 2006; Cai et al., 2006; Golding et al., 2005; Newman et al., 2006; Ozbudak et al., 2002; Raj et al., 2006; Raser and O'Shea, 2004; Rosenfeld et al., 2005; Sigal et al., 2006). Recently, Kaufmann and van Oudenaarden (2007) critically reviewed these experimental observations and formulated as a major goal "to identify and differentiate between the myriad possible origins of this variability". The diverging experimental results, even in the same eukaryotic organism *S. cerevisiae*, support the idea that gene expression is influenced by more than one main driving source. In the following we focus on the budding yeast *S. cerevisiae* and the experimental results found by Raser and O'Shea (2004), Bar-Even et al. (2006), and Newman et al. (2006). Raser and O'Shea (2004) measured the intrinsic noise strength of the *PHO5* and *PHO84* promoters at different rates of gene expression in promoter constructs. They distinguished between three general noise profiles depending on the relative promoter reaction rates. Therefore, noise intrinsic to gene expression seemed to be promoter-specific, e.g., noise generation at the *PHO5* promoter depended on stochastic promoter activation due to chromatin remodeling. Bar-Even et al. (2006) investigated native expression of 43 genes under 11 different conditions, and Newman et al. (2006) presented an extensive overview of protein noise for more than 2500 proteins expressed from their endogenous promoter and natural chromosomal position by the use of a combination of high-throughput flow cytometry and a library of GFP-tagged yeast strains. Both studies concluded that a random *birth-and-death* process of low-copy mRNA molecules describe the large observed variations quite well: for the great majority of proteins the noise level is inversely proportional to the mean protein abundance implying a clear signature of a Poisson process. In addition, Bar-Even et al. (2006) noted that infrequent promoter activation might also explain the observed noise trend, but biological mechanisms of promoter activation are not yet sufficiently characterized.

The following questions arise: how can one differentiate between the sources of noise? Under which conditions does promoter activation dominate the noise profile, or when do mRNA fluctuations due to low-copy number describe the observed noise trend? A possible mechanism to discriminate noise profiles is given by our stochastic model considering TF binding explicitly. For approximately constant TF concentration and high amount of protein synthesized per mRNA, the intrinsic variance σ_I^2 of Eq. (1) reduces to $\sigma_I^2 = \sigma_A^2 + \sigma_R^2$. Hence the ratio σ_A^2/σ_R^2 determines the predominant source of noise: if $\sigma_A^2/\sigma_R^2 >> 1$, promoter activation will be the dominant process while in the case of $\sigma_A^2/\sigma_R^2 << 1$ the major part of gene expression noise is due to transcription. For long protein lifetimes and short mRNA lifetimes, Eqs. (4) and (5), the ratio is given by

$$\frac{\sigma_A^2}{\sigma_R^2} = B_{\mathrm{eq}}\,\frac{\mathrm{var}(A)}{\langle A\rangle}\,\frac{\lambda_R^+}{\gamma_A} = B_{\mathrm{eq}}\,\frac{\lambda_A^-\lambda_R^+}{\gamma_A^2}. \tag{8}$$

Note that this ratio is time- and cell cycle-independent although σ_R^2, Eq. (4), and σ_A^2, Eq. (5), both depend on the cell cycle time, $t - t_0$, and generation time, T_G. It follows from Eq. (8) that for fixed rates the probability of TF binding, B_{eq}, determines the value of the ratio σ_A^2/σ_R^2, i.e., the induction level of the TF provides one possibility to distinguish between the different origins of noise and therefore selects the predominant source of noise. Thus, we expect for highly expressed genes, $B_{\mathrm{eq}} \to 1$, to show signature of noise from promoter activation provided $\lambda_A^-\lambda_R^+ > \gamma_A^2$ since in this case we find $\sigma_A^2/\sigma_R^2 > 1$. In contrast, we expect for low induced genes, $B_{\mathrm{eq}} \ll 1$, to show signature of Poissonian noise from mRNA synthesis, since in this case $\sigma_A^2/\sigma_R^2 < 1$ holds given $\lambda_A^-\lambda_R^+/\gamma_A^2$ is not too large. Therefore, the induction level of the TF, B_{eq}, provides a proper explanation for the observation of different noise contributions even in the same organism.

In Fig. 3 we present the mean protein abundance vs. noise under different induction levels of the TF. In order to select arbitrary time points t within a given cell cycle, $0 \leq t - t_0 \leq T_G$, we use the full expressions for the mean amount of protein and noise contributions (see Supporting Information of Rausenberger and Kollmann (2008)), because the approximations presented in Eqs. (2), (4) and (5) are only valid in the asymptotic limit $t - t_0 \gg (\lambda_R^-)^{-1}$. We calculate the mean amount of protein and the noise contributions for several genes at randomly selected time points for several induction levels of the TF binding B_{eq}. In Fig. 3A, we assume a low induction of the TF, where the mean induction level B_{eq} equals 0.07. The noise contribution arising from transcription (black circles) dominates the overall noise (light shaded diamonds). In Fig. 3B, noise from promoter activation (dark shaded squares) overrules noise from transcription. This can be arranged with an highly

induced TF, with mean induction level $B_{eq} = 0.7$. In fact, changes in induction level can be established by different experimental set-ups. TF induction is expected to be quite low for experiments with native genes in their chromosomal context (Newman et al., 2006; Raj et al., 2006). Bar-Even et al. (2006) investigated native genes implying a large set of low induced genes, $B_{eq} \ll 1$, such that transcription is the prevailing source of noise. In contrast, Raser and O'Shea (2004) constructed yeast strains that expressed CFP and GFP proteins from identical promoters. In promoter constructs, TF induction is likely to be rather strong $B_{eq} \to 1$, such that we expect that promoter activation noise is the dominant noise contribution. Therefore, both experimental scenarios can be qualitatively reproduced quite well with our stochastic model by varying the induction level of the TF binding.

4.3. Transcription factor binding determines population distribution

In eukaryotes, promoter activation is believed to occur due to chromatin remodeling (Raser and O'Shea, 2004) which erratically uncovers transcription-factor binding sites. TF binding, however, is assumed to be quite fast and frequent, because of the high copy number of TFs. Therefore, independence of TF binding and promoter activation seems to be a reasonable assumption in eukaryotes. For prokaryotes, the situation is less clear since a possible explanation or mechanism for the slow process of promoter activation is still lacking, although it has been measured quite accurately (Cai et al., 2006; Golding et al., 2005). In recent experiments Elf et al. (2007) measured the time scale for the binding/unbinding of TF at the single-molecule level in a living cell of *E. coli*. The experiment suggested that the binding/unbinding of highly abundant TFs is rather fast. Therefore, TF binding does not seem to be the limiting step within the process of gene expression, and now is assumed to switch randomly between 0 and 1, i.e., $B(t) \neq B_{eq} = const$. In order to gain insight into the influence of the TF binding on promoter activation in prokaryotes, we compare different theoretical scenarios with experimental data. Kollmann et al. (2005, Fig. 2a; redrawn in Fig. 4, inset), compared the mean expression of CheY in a wild type of *E. coli* and *flgM* cells, where the upstream transcription inhibitor, FlgM, was deleted. The deletion corresponds to a sevenfold over-expression of CheY. Several effects of an activator/repressor on the activation of the promoter are possible. We discuss the three most intuitive scenarios:

1. *TF binding and promoter activation are independent*: The RNA-polymerase can start transcription if and only if the TF is bound (repressor is unbound) *and* the promoter is active.
2. *TF binding enhances promoter switch-on rate*: For the experimentally observed over-expression we assume that the switch-on rate λ_A^+ of the over-expression of CheY is enhanced by the factor α compared to that of the wild type, i.e., $\lambda_{A,OE}^+ = \lambda_{A,WT}^+ \alpha$. The switch-off rate λ_A^- is not affected.
3. *TF binding decreases promoter switch-off rate*: For the experimentally observed over-expression we assume that the switch-off rate λ_A^- of the over-expression is decreased by the factor α compared to that of the wild type, i.e., $\lambda_{A,OE}^- = \lambda_{A,WT}^-/\alpha$. The switch-on rate λ_A^+ is not affected.

Of course, combinations of the mentioned scenarios are possible and likely to occur in nature. However, to keep the estimated parameters identifiable, we only focus on these three limiting scenarios.

The experimental data show that the mean protein level of the over expressing *flgM* cells is sevenfold higher than that of the wild type cells. Furthermore, the standard deviation of the population distribution for the *flgM* cells increases quite significantly compared to the wild type cells (Fig. 4 (inset) and Table 1). Fig. 4 shows the population distributions of the wild type and *flgM* cells for the different scenarios after parameter estimation (see Materials and methods). The parameters are estimated such that the wild type standard deviation and the mean fluorescence level of the *flgM* cells are represented best. The estimated parameters ($\langle B \rangle$ for the first scenario, λ_A^+, λ_A^- and the level of the over-expression factor α for second and third scenario) and the corresponding characteristics of the population distributions are summarized in Table 1.

The simulations reveal that there exits a set of parameters for the first scenario, where the TF binding does not influence the promoter activation process directly, such that the characteristic standard deviation of the wild type and the mean fluorescence level of the *flgM* cells is reproduced quite well (residual sum of squares, RSS = 0.004). Furthermore, it also mimics (without any optimization) the increased standard deviation of the *flgM* cells (cf. Table 1 and Fig. 4, light shaded crosses denotes wild type, black crosses denotes sevenfold over-expression). The mean TF binding for the *flgM* cells is 7.3 times larger than that of the wild type cells ($\langle B \rangle_{WT} = 0.13, \langle B \rangle_{OE} = 0.95$) which leads to an about sevenfold protein over-expression of the mean fluorescence level. For the second scenario, where TF binding and promoter activation are not independent of each other, but TF binding enhances the promoter

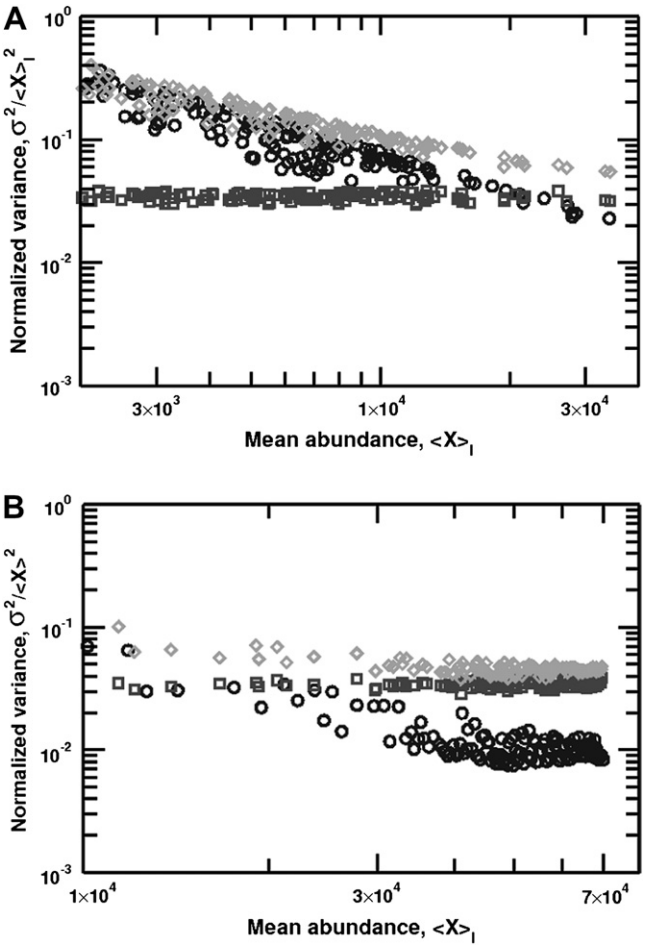

Fig. 3. Simulated mean abundance vs. noise for different genes. Transcriptional contribution (black circles), noise from promoter activation (dark shaded squares) and overall noise (light shaded diamonds). A) Low induced TF, $B_{eq} = 0.07$, leads to $\sigma_A^2/\sigma_R^2 < 1$ such that transcriptional noise dominates. B) Highly induced TF, $B_{eq} = 0.7$, leads to $\sigma_A^2/\sigma_R^2 > 1$ such that noise from promoter activation dominates.

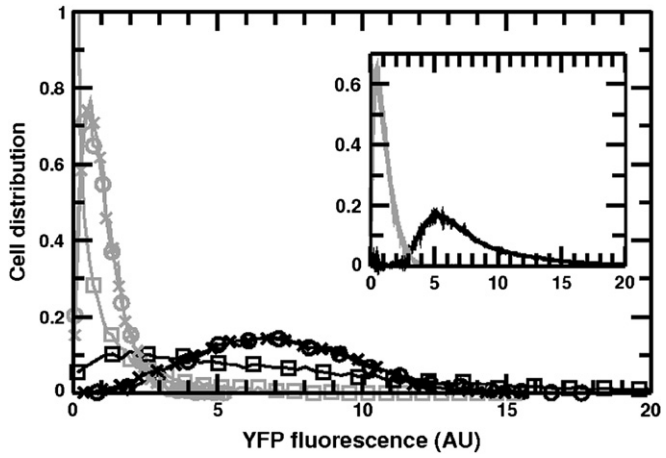

Fig. 4. Simulations of different effects of TF binding on promoter switch-on/-off rates for fixed and estimated parameters. First scenario (circles): Wild type (light shaded) and *flgM* cells (black) of assuming that promoter activation and TF binding are independent; second scenario (squares): wild type (light shaded) and *flgM* cells (black) assuming that the TF binding enhances promoter switch-on rate; third scenario (×): wild type (light shaded) and *flgM* (black) cells assuming TF binding decreases promoter switch-off rate. Means, standard deviations and estimated parameters are summarized in Table 1, fixed parameters are given in Materials and methods. The inset shows the experimental levels of CheY, expressed as YFP fusion from native chromosomal position for wild type (light shaded) and *flgM* cells (black). Redrawn from Kollmann et al. (2005).

switch-on rate, the simulations with the estimated parameters does not represent the characteristic standard deviation of the wild type and the mean fluorescence level of the *flgM* cells equivalently well (RSS = 0.67, Fig. 4 light shaded squares and black squares lines denote wild type and *flgM* cells, resp.). The standard deviation of the *flgM* cells becomes much larger than the experimental one. The estimated parameters are given by $\lambda_A^+{}_{,WT} = 0.005$ for the wild type switch-on rate and $\alpha = 413$ for the increased over-expression factor for the *flgM* cells, i.e., $\lambda_{A,OE}^+ = \lambda_{A,WT}^+ \alpha = 2.07$. This increased switch-on rate for the *flgM* cells implies that the promoter is switched-on in 95% of the time for *flgM* cells, but only in 5% of the time for wild type cells. For the third scenario, where the binding of the TF decreases the promoter switch-off rate, a set of parameters can be found such that both characteristics are reproduced (RSS = 0.04, Fig. 4 light shaded × and black × represent wild type and *flgM* cells, resp.). The standard deviation of the *flgM* cells is also increased quite well. The resulting parameters $\lambda_A^-{}_{,WT} = 0.39$ for the wild type switch-off rate and $\alpha = 924$ for the reduction factor for the *flgM* cells, i.e., $\lambda_{A,OE}^- = \lambda_{A,WT}^-/\alpha = 4.2 \times 10^{-4}$, imply that the promoter is switched-off 90% of the time for wild type and 1% of the time for *flgM* cells. The increased skewness of the *flgM* cells observed by Kollmann et al. (2005) in experiments is not reflected in any scenario of the underlying time-dependent model. The proposed time-dependent description is a reduced version of the overall system and does not take into account the complex flagella network such that an entire coincidence of the experiments and simulations is not expected. The skewness might also be influenced by external factors, like variations in ribosome or polymerase concentrations or by feedbacks of downstream to upstream processes, but none of these features are explicitly included in the present model. However, if we compare the resulting skewness of the *flgM* cells in each scenario with the experimentally measured one we find that the second scenario has the largest positive skewness, but this scenario does not fit the required characteristics quite well. For the first scenario, the characteristics are represented rather well and the skewness is also increased compared to the third scenario. Thus, TF binding and promoter activation can be considered to good approximation as independent processes in prokaryotes. Of course, this hypothesis has to be investigated in further experiments.

Table 1
Characteristic mean μ, standard deviation σ and skewness for population distributions from experiments of Kollmann et al. (2005) and simulations using the proposed stochastic model. Cells in which the upstream transcription inhibitor, the anti-sigma factor FlgM, was deleted, are denoted by *flgM* cells. Differences in the simulated scenarios, interpretation of the over-expression factor α and parameter estimation are described in the text and in Materials and methods. The residual sum of squares RSS describes the discrepancy between the data and the corresponding simulated scenario.

		μ	σ	Skewness	Estimated parameter	RSS
Experiment	Wild type	1	0.69	2.51		
	flgM cells	6.96	3.38	1.74		
Simulation	Wild type	1	0.67	1.26	$\langle B\rangle_{WT} = 0.13$	0.004
1st scenario	*flgM* cells	7.02	2.63	0.44	$\langle B\rangle_{OE} = 0.94$	
Simulation	Wild type	1	1.51	2.86	$\lambda_A^+ = 0.005$	0.67
2nd scenario	*flgM* cells	6.92	5.25	1.07	$\alpha = 413$	
Simulation	Wild type	1	0.63	1.17	$\lambda_A^- = 0.39$	0.04
3rd scenario	*flgM* cells	6.77	2.56	0.1	$\alpha = 924$	

5. Discussion

Especially during the last decade, various experimental technologies have been developed and a lot of effort has been undertaken to investigate the stochastic nature of gene expression and its implications for cell-to-cell variability. Moreover, several mathematical models emerged aiming at interpreting the experimental results as well as deriving hypotheses for the fundamental processes involved in gene expression. For the theoretical description of cell-cycle averaged experimental data, a mean-field approximation of the master equation has been successfully applied, whereas for extracting information from time-resolved experimental data, an out-of-equilibrium description seems to be appropriate. Different experiments have identified different causes for the main contribution to gene expression noise. This implies that there might be no general rule for the main source of noise or comprehensive knowledge of the overall noise architecture. The main contribution to the cell-to-cell variation within a clonal population depends strongly on the kinetic rates and the molecule lifetimes associated with the expression of each individual gene. We show that the induction level of the TF binding, B_{eq}, is one possibility to differentiate between noise stemming from promoter activation and noise originating from transcription, and thus determines the dominant source of noise. Low induced genes (Bar-Even et al., 2006) bear clear transcriptional noise signature due to low-copy number of mRNA molecules, whereas highly induced genes (Raser and O'Shea, 2004) show typical characteristics of noise stemming from promoter activation (Fig. 3). Therefore, one possible explanation for these diverging experimental results can be given by the additional incorporation of an TF binding process acting independently of promoter activation.

In eukaryotes, independence of TF binding and promoter activation is a reasonable assumption whereas in prokaryotes the situation is less clear. Based on experimentally measured cell distributions of wild type and *flgM* cells of *E. coli*, we performed parameter estimation with the time-dependent model of gene expression to discriminate between limiting cases of the effect of TF binding on promoter activation. Simulations reveal that for two out of three theoretical scenarios, a set of parameters can be found such that the characteristic standard deviation of the wild type cells and mean fluorescence level of the *flgM* cells can be reproduced. An increased standard deviation for the *flgM* cells can also be observed in both scenarios such that a qualitative distinction between both

scenarios based on simulations seems to be difficult. The biological interpretation of the estimated parameters, however, argues against one scenario: If the binding of the TF reduced the promoter switch-off rate (third scenario), the promoter would be switched-off most of the time, 90% for the wild type. However, these dead-times for the protein production, even for a repressed gene, contradict the experimental observations in *E. coli* where proteins are produced quite continuously (pers. comm. V. Sourjik). Since the protein level should be strongly optimized to allow for reliable information processing, it seems to be rather unlikely that promoters have evolved which are so strongly repressed that they are switched-off nearly all of the time. Furthermore, the over-expression factor is estimated to be rather large. This implies that an experimental depletion of an upstream inhibitor leading to a sevenfold increased mean fluorescence level corresponds to a theoretical 900 fold change of the switching rates. The cellular effort to achieve this is expected to be rather high and therefore inefficient. The independence of promoter activation and TF binding (first scenario), however, is quite likely since the cellular effort of a sevenfold increase of the probability that the TF is bound, is biologically reasonable. Thus, the independence of promoter activation and transcription factor binding might be an intrinsic property of the biological system, both for the wild type and the over-expression line. One way to regulate gene expression is to fine tune and control the mean binding of the TF which could be tested in experiments.

To move forward in this exciting field and to gain deeper insight into the underlying mechanisms contributing to the overall variance, more studies on mRNA/protein synthesis events and lifetimes are required across a variety of organisms. It would be also desirable to develop techniques bridging the gap between cell-cycle averaged and time-resolved expression data. These studies and techniques could serve to ensure or argue against the different theoretical assumptions and descriptions of gene expression with the help of idealized stochastic processes. Moreover, there are still many open questions, e.g., how cells control and regulate gene expression noise to behave most efficiently, what are the biological consequences of gene expression noise for the underlying biological system, and to what extend can the noise properties or noise regimes be generalized across different organisms. Due to the complexity of gene expression, theory alone or experimental progress on its own seem to be inefficient. Iterative studies of experiments and modeling, however, seem to be promising to establish stochastic gene expression as an on-going research area in biophysics and molecular biology.

6. Materials and methods

6.1. Simulating stochastic processes

We assume that TF binding and promoter activation can be described by a random telegraph process with transition rates λ_B^+, λ_B^- (TF binding), λ_A^+ and λ_A^- (promoter activation). The initial state of the promoter is determined by drawing a uniformly distributed random number (URN) $r \in [0,1]$ and checking whether $r < \lambda_A^+/\gamma_A$, such that the promoter is *on*. Otherwise it is *off* in its initial state. Transcription is a *birth-and-death* process with time-dependent synthesis rate $\lambda_R^+(t) = \lambda_R^+ A(t)B(t)$, i.e., mRNA can only be synthesized if the TF is bound and the promoter is in its *on*-state (Fig. 1). The original Gillespie-algorithm (Gillespie, 1977) has been refined (for review see Gillespie and Petzold, 2006), but also modified and extended to model growing cell volume via time-dependent reaction rates (Lu et al., 2004). To determine the next time τ of the reaction and the next reaction μ for time-dependent reaction rates, we follow the lines of Gillespie (1977) and Lu et al. (2004) and arrive at the cumulative distribution function

$$F(\tau) \;=\; 1 - \exp\left[-\sum_{\mu} \int_0^{\tau} a_\mu(t+\tau')\mathrm{d}\tau' \right] \;=:\; 1 - P_0(\tau). \tag{9}$$

Drawing a URN $\overline{u}_1 \in [0,1]$ we set $\overline{u}_1 = 1 - P_0(\tau)$ and obtain, since $1 - \overline{u}_1$ is also a URN, the new URN $u_1 = P_0(\tau)$. The stochastic time τ for the next reaction to occur is obtained by inverting this equation. We formulate the cumulative distribution function for transcription with a time-dependent synthesis rate $a_1(t) = \lambda_R^+(t) = \lambda_R^+ A(t)B(t)$ of mRNA. The degradation does not depend on time $\tau > t$ with transition rate $a_2 = \lambda_R^- R(t)$ where $R(t)$ represents the actual amount of mRNA at time t. We find that at time t, the next stochastic time τ has to satisfy

$$\ln(u_1) \;=\; -\lambda_R^+ \int_0^{\tau} B(t+\tau')A(t+\tau')\mathrm{d}\tau' - \lambda_R^- R(t)\tau. \tag{10}$$

Drawing a second URN $u_2 \in [0,1]$, the next reaction μ must fulfill the inequality

$$\lambda_R^+ A(t+\tau)B(t+\tau) < u_2\left(\lambda_R^+ A(t+\tau)B(t+\tau) + \lambda_R^- R(t+\tau)\right) \tag{11}$$

The modified Gillespie-algorithm with time-dependent reaction rate $a_1(t)$ determines the next stochastic time τ in Eq. (10) as the upper bound of the integral. In our case, however, the integrand has a very special form, i.e., it is 1 if and only if the TF is bound as well as the promoter is *on*. Otherwise the integrand is 0. Therefore, the integration becomes a simple summation overall *on*-states, O_τ, of the product of the TF times the promoter within the time interval $[t, t+\tau]$. We define α_τ as the ratio of the *on*-states to the time interval $[t, t+\tau]$, i.e., $0 \le \alpha_\tau = O_\tau/\tau \le 1$. The next time τ can therefore be calculated according to Eq. (10) which reduces for an exponentially distributed stochastic variable $\ln(u_1) =: -z$ to

$$\tau \;=\; \frac{z}{\lambda_R^+ \alpha_\tau + \lambda_R^- R(t)}. \tag{12}$$

Note that α_τ depends on the single realization of $A(t)B(t)$ and is thus also a stochastic variable. If $A(t)$ and $B(t)$ are time-independent, e.g., $A(t)B(t) \equiv 1$, it follows that $\alpha_\tau = 1$ and the original Gillespie-algorithm (Gillespie, 1977) is recovered. Therefore, we use the original Gillespie-algorithm to determine the next time τ and reaction μ and check afterwards whether the proposed Gillespie-step can be performed or not. If $A(t+\tau)B(t+\tau) = 1$, mRNA synthesis can be realized, but if $A(t+\tau)B(t+\tau) = 0$ and mRNA synthesis is selected as reaction μ, the step is rejected and new URNs are drawn. In general, this procedure will always select a stochastic time τ which is smaller than that of the modified algorithm of Eq. (12) since $0 \le \alpha_\tau \le 1$. However, the above procedure of taking the original Gillespie-algorithm and rejecting specific reactions is equivalent to the determination of the next time τ via Eq. (12). To obtain the same stochastic time for both procedures, the following equation should hold:

$$z \;=\; \frac{\lambda_R^+ \alpha_\tau + \lambda_R^- R(t)}{\lambda_R^+ + \lambda_R^- R(t)}\,\overline{z}; \tag{13}$$

where z and $\overline{z}$ are exponentially distributed variable stemming from a URN u via $\ln(u)$. This equation holds since an exponentially distributed variable can be described by the product of a constant (given a specific α_τ) times another exponentially distributed stochastic variable $\overline{z}$. Therefore, taking the original Gillespie-algorithm and rejecting specific reactions according to the time-dependent trajectory $A(t)B(t)$ is just another realization of the modified Gillespie-algorithm from Eq. (12) and averaging over a lot of trajectories yields the same result.

The generated mRNA trajectory $R(t)$ can directly be used to calculate the appropriate mean protein number $\langle X(t)\rangle_X$. At the end of one cell cycle, the cell divides symmetrically into two daughter cells. The mother's cell amount of protein and mRNA is divided binomially to both daughter cells.

6.2. Effect of TF binding on promoter activation: parameter estimation

A least-square fit is performed with MATLAB such that, after data normalization to mean wild type fluorescence of 1, the experimental standard deviation of the wild type cells ($\sigma = 0.69$) and mean *flgM* fluorescence ($\mu = 6.96$) of the population distributions are best represented. For each optimization step, $20 \times 2^9 = 10\,240$ realizations of the time-dependent gene expression model of Rausenberger and Kollmann (2008) are generated. The following parameters are fixed for the simulations: $\lambda_R^+ = 2$, $\lambda_R^- = 0.2$, $\lambda_X^+ = 4$ and $\lambda_X^- = 10^{-4}$. In the first scenario, the promoter switch-on/-off rates are set to $\lambda_A^+ = 0.05$ and $\lambda_A^- = 0.1$, representing realistic kinetic rates for a repressed gene (Golding et al., 2005). The mean TF binding rates of wild type, $\langle B\rangle_{WT}$, and *flgM* cells, $\langle B\rangle_{OE}$, are estimated separately. For the second and third scenario we assume a mean TF binding of $\langle B\rangle = 0.5$ for the wild type as well as for the *flgM* cells. The promotor switch-off rate is set to $\lambda_A^- = 0.1$ and the promoter switch-on $\lambda_{A,WT}^+$ for the wild type and the strength of the over-expression, α, are estimated in the second scenario. If the TF is bound, the promoter switch-on rate for the *flgM* cells, $\lambda_{A,OE}^+$, will be enhanced by the over-expression factor α, i.e., $\lambda_{A,OE}^+ = \lambda_{A,WT}^+\alpha$. In the third scenario we set the promoter switch-on rate to $\lambda_A^+ = 0.05$ and estimate the promoter switch-off rate $\lambda_{A,WT}^-$ for the wild type and the strength of the over-expression α. If the TF is bound, the promoter switch-off rate for the *flgM* cells, $\lambda_{A,OE}^-$, will be reduced by the over-expression factor α, i.e., $\lambda_{A,OE}^- = \lambda_{A,WT}^-/\alpha$. The estimated parameters are summarized in Table 1.

Funding statement

This work was financially supported by the International Graduate School "Signal Systems in Plant Model Organisms" (DFG grant no. GRK1305 for JR), the DFG Emmy Noether-Program (for MK), BMBF FRISYS 0313921 (for CF and JR), and by the Excellence Initiative of the German Federal and State Governments.

References

Alon, U., 2006. An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapman & Hall/CRC.
Austin, D.W., Allen, M.S., McCollum, J.M., Dar, R.D., Wilgus, J.R., Sayler, G.S., Samatova, N.F., Cox, C.D., Simpson, M.L., 2006. Gene network shaping of inherent noise spectra. Nature 439, 608–611.
Avery, S.V., 2006. Microbial cell individuality and the underlying sources of heterogeneity. Nat. Rev. Microbiol. 4, 577–587.
Balaban, N.Q., Merrin, J., Chait, R., Kowalik, L., Leibler, S., 2004. Bacterial persistence as a phenotypic switch. Science 305, 1622–1625.
Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y., Barkai, N., 2006. Noise in protein expression scales with natural protein abundance. Nat. Genet. 38 (6), 636–643.
Becskei, A., Kaufmann, B.B., van Oudenaarden, A., 2005. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. Nat. Genet. 37 (9), 937–944.
Becskei, A., Serrano, L., 2000. Engineering stability in gene networks by autoregulation. Nature 405, 590–593.
Berg, O.G., 1978. A model for statistical fluctuations of protein numbers in a microbial population. J. Theor. Biol. 71, 587–603.
Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., Phillips, R., 2005. Transcriptional regulation by numbers: models. Curr. Opin. Genet. Dev. 15, 116–124.
Blake, W.J., Kœrn, M., Cantor, C.R., Collins, J.J., 2003. Noise in eukaryotic gene expression. Nature 422, 633–637.
Buchler, N.E., Gerland, U., Hwa, T., 2003. On schemes of combinatorial transcription logic. Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5136–5141.
Cai, L., Friedman, N., Xie, X.S., 2006. Stochastic protein expression in individual cells at the single molecule level. Nature 440, 358–362.
Colman-Lerner, A., Gordon, A., Serra, E., Chin, T., Resnekov, O., Endy, D., Pesce, C.G., Brent, R., 2005. Regulated cell-to-cell variation in a cell-fate decision system. Nature 437, 695–706.
Dublanche, Y., Michalodimitrakis, K., Kummerer, N., Foglierini, M., Serrano, L., 2006. Noise in transcription negative feedback loops: simulation and experimental analysis. Mol. Sys. Biol. 2, 41.
Elf, J., Li, G.W., Xie, X.S., 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. Science 316, 1191–1194.
Elowitz, M.B., Levine, A.J., Siggia, E.D., Swain, P.S., 2002. Stochastic gene expression in a single cell. Science 297, 1183–1186.
Friedman, N., Cai, L., Xie, X.S., 2006. Linking stochastic dynamics to population distribution: an analytical framework to gene expression. Phys. Rev. Lett. 97, 168302.
Gardiner, C.W., 1990. Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences, second ed. Springer-Verlag.
Gerstung, M., Timmer, J., Fleck, C., 2009. Noisy signaling through promoter logic gates. Phys. Rev. E 79, 011923.
Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81 (25), 2340–2361.
Gillespie, D.T., 2007. Stochastic simulation of chemical kinetics. Annu. Rev. Phys. Chem. 58, 35–55.
Gillespie, D.T., Petzold, L.R., 2006. Numerical simulation for biochemical kinetics. In: Szallasi, Z., Stelling, J., Periwal, V. (Eds.), Systems Modelling in Cellular Biology: From Concepts to Nuts and Bolts. MIT Press, pp. 125–147.
Golding, I., Paulsson, J., Zawilski, S.M., Cox, E.C., 2005. Real-time kinetics of gene activity in individual bacteria. Cell 123, 1025–1036.
Kaufmann, B.B., van Oudenaarden, A., 2007. Stochastic gene expression: from single molecules to the proteome. Curr. Opin. Genet. Dev. 17 (2), 107–112.
Ko, M.S., Nakauchi, H., Takahashi, N., 1990. The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. EMBO J. 9, 2835–2842.
Kœrn, M., Elston, T.C., Blake, W.J., Collins, J.J., 2005. Stochasticity in gene expression: from theories to phenotypes. Nat. Rev. Genet. 6, 451–464.
Kollmann, M., Løvdok, L., Bartolome, K., Timmer, J., Sourjik, V., 2005. Design principles of a bacterial signalling network. Nature 438, 504–507.
Lu, T., Volfson, D., Tsimring, L., Hasty, J., 2004. Cellular growth and division in the Gillespie algorithm. Syst. Biol. 1 (1), 121–128.
McAdams, H.H., Arkin, A., 1997. Stochastic mechanisms in gene expression. Proc. Natl. Acad. Sci. U.S.A. 94, 814–819.
Mettetal, J.T., Muzzey, D., Pedraza, J.M., Ozbudak, E.M., van Oudenaarden, A., 2006. Predicting stochastic gene expression dynamics in single cells. Proc. Natl. Acad. Sci. U.S.A. 103 (19), 7304–7309.
Newman, J.R.S., Ghaemmaghami, S., Ihmels, J., Breslow, D., Noble, M., DeRisi, J.L., Weissmann, J.S., 2006. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. Nature 441, 840–846.
Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D., van Oudenaarden, A., 2002. Regulation of noise in the expression of a single gene. Nat. Genet. 31, 69–73.
Paulsson, J., 2004. Summing up the noise in gene networks. Nature 427, 415–418.
Paulsson, J., 2005. Models of stochastic gene expression. Phys. Life Rev. 2, 157–175.
Pedraza, J., Paulsson, J., 2008. Effects of molecular memory and bursting on fluctuations in gene expression. Science 319, 339–343.
Pedraza, J., van Oudenaarden, A., 2005. Noise propagation in gene networks. Science 307, 1965–1969.
Pulkkinen, O., Berg, J., 2008. Dynamics of gene expression under feedback. arXiv:0807.3521v1.
Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., Tyagi, S., 2006. Stochastic mRNA synthesis in mammalian cells. PLoS Biol. 4 (10), 1707–1719.
Raj, A., van Oudenaarden, A., 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell 135, 216–262.
Raser, J.M., O'Shea, E.K., 2004. Control of stochasticity in eukaryotic gene expression. Science 304, 1811–1814.
Raser, J.M., O'Shea, E.K., 2005. Noise in gene expression: origins, consequences, and control. Science 309, 2010–2013.
Rausenberger, J., Kollmann, M., 2008. Quantifying origins of cell-to-cell variations in gene expression. Biophys. J. 95, 4523–4528.
Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S., Elowitz, M.B., 2005. Gene regulation at the single-cell level. Science 307, 1962–1965.
Shahrezaei, V., Swain, P.S., 2008. Analytical distributions for stochastic gene expression. Proc. Natl. Acad. Sci. U.S.A. 105, 17256–17261.
Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y., Rosenfeld, N., Danon, T., Perzov, N., Alon, U., 2006. Variability and memory of protein levels in human cells. Nature 444, 643–646.
Spudich, J.L., Koshland, D.E.J., 1976. Non-genetic individuality: chance in the single cell. Nature 262, 467–471.
Suel, G.M., Garcia-Ojalvo, J., Liberman, L.M., Elowitz, M.B., 2006. An excitable gene regulatory circuit induces transient cellular differentiation. Nature 440, 545–550.
Swain, P.S., 2004. Efficient attenuation of stochasticity in gene expression through post-transcriptional control. J. Mol. Biol. 344, 965–976.
Swain, P.S., Elowitz, M.B., Siggia, E.D., 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc. Natl. Acad. Sci. U.S.A. 99 (20), 12795–12800.

Thattai, M., van Oudenaarden, A., 2001. Intrinsic noise in gene regulatory networks. Proc. Natl. Acad. Sci. U.S.A. 98 (15), 8614–8619.
Tănase-Nicola, S., Warren, P.B., ten Wolde, P.R., 2006. Signal detection, modularity and the correlation between extrinsic and intrinsic noise in biochemical networks. Phys. Rev. Lett. 97 (6), 068102.
van Kampen, N.G., 1992. Stochastic Processes in Physics and Chemistry, rev. and enlarged Edition. North Holland.
Volfson, D., Marciniak, J., Blake, W.J., Ostroff, N., Tsimring, L.S., Hasty, J., 2006. Origins of extrinsic variability in eukaryotic gene expression. Nature 439, 861–864.