Quantifying Origins of Cell-to-Cell Variations in Gene Expression

Julia Rausenberger* and Markus Kollmann†
*Institute of Physics and Faculty of Biology, University of Freiburg, Freiburg, Germany; and †Institute for Theoretical Biology, Humboldt-University Berlin, Berlin, Germany

ABSTRACT A general dynamic description of protein synthesis was employed to quantify different sources of gene expression noise in cellular systems. To test our approach, we use time-resolved expression data of individual human cells and, from this information, predict the stationary cell-to-cell variation in protein levels in a clonal population. For three of the four human genes investigated, the cellular variations in expression level are not due to fluctuations in promoter activity or transcript copy number, but are almost exclusively a consequence of long-term variations of gene regulatory factors or the global cellular state. Moreover, we show that a dynamic description is much more reliable to discriminate extrinsic and intrinsic sources of noise than it is on grounds of cell-cycle averaged descriptions. The excellent agreement between the theoretical predictions and the experimentally measured noise strengths shows that a quantitative description of gene expression noise is indeed possible on the basis of idealized stochastic processes.

INTRODUCTION

Individual cells within a genetically identical population show significant phenotypic heterogeneity (1,2). One of the main driving forces for cell-to-cell variability in clonal populations is stochasticity or noise inherent to gene expression. Numerous experiments have measured the variance in protein abundances in different cellular systems (1,3–12) and showed that cellular variations of protein levels can be correlated over generations (13,14). Possible molecular origins contributing to stochasticity in protein synthesis can be seen in concentration fluctuations of factors involved in transcription, uncovering transcription factor binding sites by chromatin remodeling, polymerase binding, random bursts of mRNA and protein synthesis, fluctuations in ribosome concentration, uncovering ribosome binding sites by mRNA unfolding, and more (1,2). Which of these known processes are dominating seems to be gene-, organism-, and environment-dependent (5).

As protein levels are in general under selection within a given concentration range, any significant long-lived deviation from the optimal level would result in a reduction of fitness. Despite the large amount of data quantifying cell-cycle averaged intercellular variability, the specific molecular origins of gene expression noise are still unclear. Therefore, significant effort is now undertaken to generate single-cell time-resolved expression data (13). However, only a small amount of work has been done to extract information from nonequilibrium protein synthesis trajectories, because most previous theoretical approaches lack dynamic description or enforce a priori (quasi-) stationarity (15–20).

It is commonly agreed to define all gene-specific stochastic events of protein synthesis as intrinsic noise. Differences between cells, either in the global cellular state or in the concentration or activity of any factor that affects gene expression, is referred to as extrinsic noise (3,21). Thus, in a large clonal population of cells with fixed generation time, T_G, the cell-to-cell variation in the amount of a specific protein is characterized by two main contributions to the total variance, $\sigma^2_{tot} = \sigma^2_I + \sigma^2_E$. Here, the intrinsic noise contribution, σ^2_I, is specific to each gene in its genomic context in a fixed intracellular environment. The extrinsic noise contribution, σ^2_E, is usually separated into fluctuations of upstream factors that drive expression directly, like a given activator concentration and the global cellular state that influences gene expression, e.g., via ribosome, polymerase, and protease concentrations.

THEORY

Three major steps are involved in a generic model of gene expression in living cells (19): 1), promoter activation; 2), transcription; and 3), translation. The underlying stochastic processes of these steps are connected in series and hence allow for assignment of the individual noise contributions to the intrinsic noise of protein synthesis corresponding to processes 1–3:

$$\sigma^2_I = \left\langle [X - \langle X\rangle_{\mathbf{I}}]^2 \right\rangle_{\mathbf{I}} = \sigma^2_A + \sigma^2_R + \sigma^2_X. \quad (1)$$

Here, the subscript, $\mathbf{I} = \{X, R, A\}$, denotes the average over all possible trajectories including protein copy number $X(t)$; mRNA copy number $R(t)$; and promoter activation $A(t)$ (for derivation, see Data S1in Supplementary Material). In Eq. 1, the translational contribution is given by $\sigma^2_X = \langle (X - \langle X\rangle_X)^2\rangle_{X,R,A}$, where the average, $\langle X\rangle_X = \sum_0^\infty XP(X, t|R(t))$, is conditionally dependent on the trajectory of mRNA synthesis, $R(t)$, that in turn depends on $A(t)$. Here, $P(X, t|R(t))$, is the probability density to observe the protein copy number X at time t given $R(t)$. Equivalently, the other intrinsic noise contributions read as $\sigma^2_R = \langle (\langle X\rangle_X - \langle X\rangle_{X,R})^2\rangle_{R,A}$ and $\sigma^2_A = \langle (\langle X\rangle_{X,R} - \langle X\rangle_{X,R,A})^2\rangle_A$. The summing up of partial noise contributions is only possible if no significant feedback exists of the expressed protein, $X(t)$, on gene regulation or any interference of intrinsic noise with extrinsic factors (22). Otherwise, a stochastic dependence of $R(t)$ on $X(t)$ could exist, making the additive form of Eq. 1 invalid.

Submitted December 5, 2007, and accepted for publication July 7, 2008.

Address reprint requests to Markus Kollmann, Tel.: 0-30-20-93-89-20; E-mail: m.kollmann@biologie.hu-berlin.de.

Editor: Michael D. Stern.

© 2008 by the Biophysical Society
0006-3495/08/11/4523/06 $2.00

doi: 10.1529/biophysj.107.127035

In the following, we denote by $\langle . \rangle_{\mathbf{E}}$ the average over all extrinsic factors that influence synthesis rates. In our approach we follow along the lines of previous approaches (19) and model synthesis and degradation by a birth-and-death process and exponential decay, respectively. Promoter activation is described as a random telegraph process. For the cellular systems considered in this work, the amount of protein synthesized per mRNA can be estimated to be $\sim 10^3$ (5), with protein copy numbers per cell of $\sim 10^3 - 10^6$ (13). Therefore, the normalized noise contribution due to stochastic independent translation events, $\sigma_X^2/\langle X \rangle_{\mathbf{I}}^2 = 1/\langle X \rangle_{\mathbf{I}}$, is expected to be small compared to other sources of noise. Note that standard deviation divided by the mean provides a physiologically relevant measure of gene expression noise as this value quantifies relative fluctuations independent of the expression level. The time evolution equation for the probability, $P(n, t)$, to observe $n \in \{X, R\}$ molecules in a given cell at time t is governed by the equation

$$\partial_t P(n,t) = \lambda^+(t)P(n-1,t) + (n+1)\lambda^- P(n+1,t) - (n\lambda^- + \lambda^+(t))P(n,t) \quad (2)$$

with rate $\lambda^+(t) = \lambda_X^+ R(t)$, if n is the amount of proteins or $\lambda^+(t) = \lambda_R^+ A(t)$, if n reflects the amount of mRNA. The corresponding degradation rate, $\lambda^- \in \{\lambda_R^-, \lambda_X^-\}$, is assumed to be independent of molecule number and constant in time. The trajectory $A(t)$ switches randomly between zero and one with exponentially distributed waiting times. Transcription at time t is only possible if the promoter is in the on-state, $A(t) = 1$. Experiments show that the main effect of cell division seems to be binomial partitioning of molecules (10,14) that changes the time evolution equation of the conditional probability of Eq. 2, $P(n, t|n', t')$, to

$$P(n,t|n',t') = \sum_{n_0^-=0}^{\infty} \sum_{n_0^+=0}^{n_0^-} P(n,t|n_0^+,t_0) \times \binom{n_0^-}{n_0^+} q_0^{n_0^+} (1-q_0)^{n_0^- - n_0^+} P(n_0^-,t_0|n',t'), \quad (3)$$

where n_0^- and n_0^+ represent the amount of molecules before and after the last cell division occurring at time t_0. The ratio of volumes between the daughter cell and its mother cell of the cell linage under investigation is denoted by q_0. The explicit solution for the generating function $G(s,t|n',t') = \sum_{n=0}^{\infty} s^n P(n,t|n',t')$ for Eq. 3 is given by

$$G(s,t|n',t') = \left[(s-1)e^{-d(t,t')-\lambda^-(t-t')}+1\right]^{n'} \times \exp\left[(s-1)\int_{t'}^{t} \lambda^+(t'')e^{-d(t,t'')-\lambda^-(t-t'')}dt''\right], \quad (4)$$

with $d(t,t') = -\sum_{n=0}^{\infty} \Theta(t_{-n} - t')\ln(q_{-n})$ (23). Here, $\Theta(t)$ is the Heaviside step function, q_{-n} the volume fraction, and t_{-n} the time n cell-divisions-ago in the cell lineage. In the following, we assume equally dividing cells ($q_i = 1/2$), with concerted cell divisions across the population and constant generation time, T_G. These mathematical simplifications are fulfilled for the experimental data used in this work, because generation times are indeed similar (standard error over mean ≈ 0.1) and all relevant parameters are extracted from one cell cycle only. The process of promoter activation does not seem to have any significant correlations with the cell cycle (7). Therefore, we can employ a stationary solution for the autocorrelation function $\langle A(t), A(t') \rangle = var(A)e^{-\gamma_A|t-t'|}$ with $var(A) = \lambda_A^+\lambda_A^-/\gamma_A^2, \langle A \rangle = \lambda_A^+/\gamma_A$, and $\gamma_A = \lambda_A^+ + \lambda_A^-$. The switching rate between the gene states *off*→*on* and *on*→*off* are denoted by λ_A^+ and λ_A^-, respectively.

Intrinsic noise

Explicit expressions for the different intrinsic noise contributions of Eq. 1 can be obtained from analytical solutions of Eqs. 2–4 (for derivation, see Data S1). In the following, we give asymptotic expressions for a single gene, valid for any time t within a given cell cycle, $0 \le t - t_0 \le T_G$. Multiple copies of M identical genes in different loci of the DNA would result in the simple transformation $\sigma_{\mathbf{I}}^2 \to \sigma_{\mathbf{I}}^2/M$ due to stochastic independence (4); therefore, they are not considered explicitly in this work.

If we consider the most physiological case in protein synthesis involving short mRNA lifetimes, $\lambda_R^- \gg T_G^{-1}$, and long protein lifetimes, $\lambda_X^- \ll T_G^{-1}$, we obtain for $t - t_0 \gg (\lambda_R^-)^{-1}$ the asymptotic expressions for the average protein and mRNA numbers in the system

$$\langle X \rangle_{\mathbf{I}} = \mathcal{V}[(t-t_0) + T_G], \quad (5)$$

$$\langle R \rangle_{\mathbf{I}} = \mathcal{V}/\lambda_X^+. \quad (6)$$

Here, we denote the average rate of protein synthesis by $\mathcal{V} = \langle A \rangle \lambda_R^+ \lambda_X^+ / \lambda_R^-$. Note that Eq. 5 implies a linear increase in the mean amount of proteins in time as well as a doubling of protein copy number over one cell cycle. The noise contribution arising from transcription and promoter activation reads in the limit of fast mRNA degradation and high promoter switching frequency, $\lambda_R^-, \gamma_A \gg (t-t_0)^{-1} > T_G^{-1}$,

$$\sigma_R^2 = 2\mathcal{V}\frac{\lambda_X^+}{\lambda_R^-}\left[t - t_0 + \frac{1}{3}T_G\right], \quad (7)$$

$$\sigma_A^2 = 2var(\mathcal{V}_A)\frac{1}{\gamma_A}\left[(t-t_0) + \frac{1}{3}T_G\right], \quad (8)$$

with $var(\mathcal{V}_A) = \mathcal{V}^2 var(A)/\langle A \rangle^2$ the variance of the synthesis rate due to fluctuations in $A(t)$. In this limit the cellular memory of gene expression contributes at minimum one-third to the cell-to-cell variation. In the extreme case $\gamma_A \ll T_G^{-1}$, the contributions from previous generations sum up to threefold the noise generated within one cell cycle

$$\sigma_A^2 = var(\mathcal{V}_A)\left[(t-t_0)^2 + 2T_G(t-t_0) + T_G^2\right]. \quad (9)$$

Here, the first term reflects noise in protein copy number from stochastic promoter activation events that occur entirely in the actual generation. The second term includes correlations of promoter states between past generations and the actual generation. The last term describes the accumulated noise contribution of promoter activation events that have occurred entirely in previous generations.

For the case that protein lifetimes and mRNA lifetimes are significantly shorter than the generation time, memory over generations is eliminated and the stationary protein level is recovered immediately after cell division. Consequently, the solutions are time-independent and agree with those found earlier by Paulsson (18,19) for the mean protein level $\langle X \rangle_{\mathbf{I}} = \mathcal{V}/\lambda_X^-$, mean mRNA level $\langle R \rangle_{\mathbf{I}} = \mathcal{V}/\lambda_X^+$ and the scaled variances

$$\frac{\sigma_R^2}{\langle X \rangle_{\mathbf{I}}^2} = \frac{1}{\langle R \rangle_{\mathbf{I}}}\frac{\lambda_X^-}{\lambda_R^- + \lambda_X^-}$$

$$\frac{\sigma_A^2}{\langle X \rangle_{\mathbf{I}}^2} = \frac{var(A)}{\langle A \rangle^2}\frac{\lambda_R^-}{\gamma_A + \lambda_R^-}\frac{\lambda_X^-}{\lambda_R^- + \lambda_X^-}\left(1 + \frac{\lambda_R^-}{\gamma_A + \lambda_X^-}\right). \quad (10)$$

Extrinsic noise

Fluctuations in the cellular environment can contribute substantially to gene expression noise. In the following, we concentrate on the most predominant case in cellular systems involving short mRNA lifetimes, $\lambda_R^- \gg T_G^{-1}$, and long protein lifetimes, $\lambda_X^- \ll T_G^{-1}$. The global cellular state and the factors involved in gene regulation are also expected to be subject to stochastic fluctuations that change in turn the protein production rate, $\mathcal{V} = \mathcal{V}_E(t)$. A generic approach to describe the time evolution of $\mathcal{V}_E(t)$ is given by the Ornstein-Uhlenbeck process as proposed by Sigal et al. (13)

$$\partial_t \mathcal{V}_E(t) = -\gamma_E[\mathcal{V}_E(t) - \langle \mathcal{V}_E \rangle_{\mathbf{E}}] + D^{1/2}\xi(t), \quad (11)$$

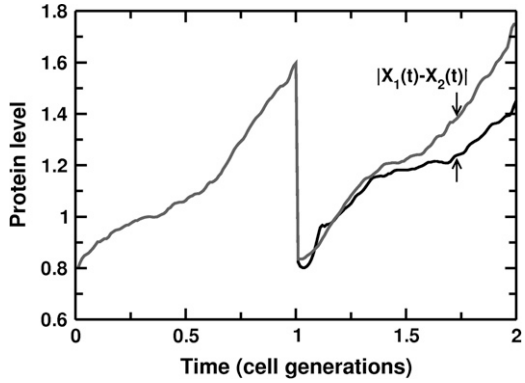

FIGURE 1 Instructive picture of the normalized protein abundance trajectories within one cell cycle for the mother cell and its two daughter cells, $X_1(t)$ and $X_2(t)$, of the human gene TOP1 taken from Sigal et al. (13). The difference in expression levels of the two daughter cells is indicated by the arrows.

where γ_E^{-1} represents the effective relaxation time of the extrinsic factors, D a diffusion constant in velocity space, and $\xi(t)$ δ-correlated white noise. The average $\langle . \rangle_{\mathbf{E}}$ corresponds in this simplified approach to averaging over the stochastic variable $\xi(t)$. The stationary autocorrelation function is given by $\langle \mathcal{V}_E(t), \mathcal{V}_E(t') \rangle_{\mathbf{E}} = var(\mathcal{V}_E)\exp[-\gamma_E|t-t'|]$, with $var(\mathcal{V}_E) = D/(2\gamma_E)$. The actual protein concentration can be inferred from

$$X(t) = \int_{t_0}^{t} \mathcal{V}_E(t')dt' + \sum_{n=1}^{\infty}\left(\frac{1}{2}\right)^n \int_{t_0^- - nT_G}^{t_0^- - (n-1)T_G} \mathcal{V}_E(t')dt', \quad (12)$$

where t_0^- denotes a time point right before the last cell division. Using the expression for the stationary autocorrelation function of Eq. 11, the extrinsic noise contributions to cell-to-cell variability can be derived from Eq. 12 (check Data S1 for details). With the definitions $I_1(t-t_0) = 2\,var(\mathcal{V}_E)\gamma_E^{-1}[t-t_0-\gamma_E^{-1}(1-\exp[-\gamma_E(t-t_0)])]$, $I_2(t-t_0) = var(\mathcal{V}_E)\gamma_E^{-2}(1-\exp[-\gamma_E(t-t_0)])$ Z_E, and $Z_E = (1-\exp[-\gamma_E T_G])(1-2^{-1}\exp[-\gamma_E T_G])^{-1}$, the variance of $X(t)$ reads as

$$\sigma_E^2 = I_1(t-t_0) + I_2(t-t_0) + \frac{I_1(T_G)+I_2(T_G)}{3}. \quad (13)$$

Note that the autocorrelation function for promoter activation $A(t)$ has the same functional form as for the Ornstein-Uhlenbeck process. Therefore, the contributions from promoter activation to gene expression noise, σ_A, result in equivalent expressions as given for extrinsic noise, Eq. 13, but with $var(\mathcal{V}_A)$ substituted for $var(\mathcal{V}_E)$ and γ_A for γ_E. Therefore, the limit $\gamma_E \ll T_G^{-1}$ for σ_E^2 results in the equivalent expression as found for σ_A^2 in the limit $\gamma_A \ll T_G^{-1}$ (see Eq. 9).

RESULTS

In the following, we concentrate on the single cell measurements for human H1299 lung carcinoma cells as reported by Sigal et al. (13). These are fast proliferating cells with generation time of ~18 h. A typical time course for the expression of a human gene (TOP1) is depicted in Fig. 1. The apparent correlation in shape of the two daughter trajectories is observed frequently and could be a signature of the underlying gene regulatory network that changes mean expression at different time points within the cell cycle. The increase of the mean protein level at the end of the cell cycle, however, is due to a readily performed DNA replication in a not-yet-divided cell. A crucial test for our analytical framework is to derive the stationary cell-to-cell variation in protein abundance from the average divergence of the trajectories between any two daughter cells, $|X_1(t) - X_2(t)|$. For our analysis we use expression data of four human genes (HMGA2, SET, TOP1, USP7) for which a sufficient amount of daughter-cell trajectories has been measured (13). The functions of the proteins range from transcriptional regulation (HMGA2), chromatin remodeling (SET), control and alteration of the topologic states of DNA during transcription (TOP1), to de-ubiquitination (USP7) (13). From the linear increase of $\langle X \rangle_{\mathbf{I,E}}$ (Fig. 2 *A*), we expect that these genes are described to good approximation by the asymptotic limit of Eqs. 5 and 6,

$$\langle X \rangle_{\mathbf{I,E}}(t) \approx \langle \mathcal{V} \rangle_{\mathbf{I,E}}[t - t_0 + T_G]. \quad (14)$$

From Eq. 14 follows that after averaging over trajectories, the mean amount of protein available after cell division equals the amount of protein synthesized on average over one cell cycle, $\langle X \rangle_{\mathbf{I,E}}(t_0 + T_G) = 2\langle X \rangle_{\mathbf{I,E}}(t_0)$. This finding is confirmed by the experimental data shown in Fig. 2 *A*, which verifies stationarity of the trajectories as the expression value at the end of the cell cycle being precisely the double of the starting

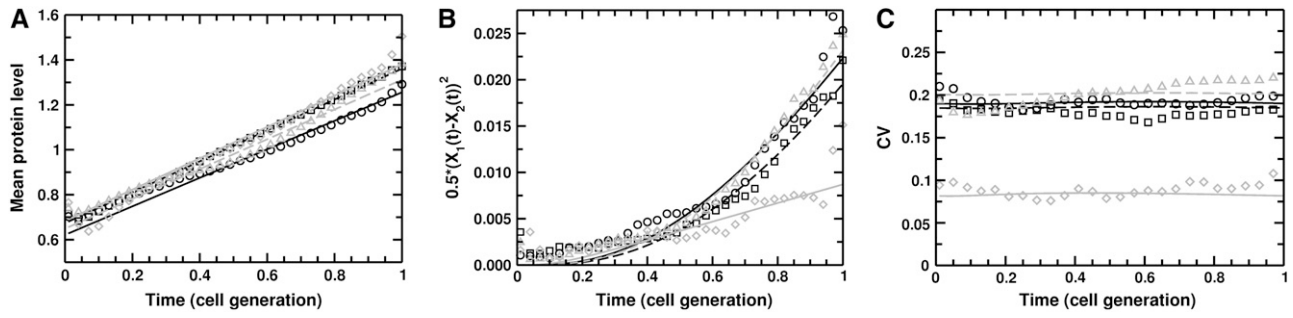

FIGURE 2 Dynamics of protein synthesis over one cell cycle for four different human genes, HMGA2 (*circles*), SET (*squares*), TOP1 (*triangles*), and USP7 (*diamonds*). Least mean-square fits to data are indicated by dark solid line for HMGA2, dark dashed line for SET, shaded dashed line for TOP1, and shaded line for USP7. (*A*) Normalized average cellular protein abundances. (*B*) Mean-squared distance of daughter cell pairs. (*C*) Coefficient of variation from the asymptotic analysis, $CV = \sigma_E/\langle X \rangle_{\mathbf{I,E}}$ in comparison with the experimentally observed cell-to-cell variation, $CV = \sigma_{exp}/\langle X \rangle_{exp}$, taken from Sigal et al. (13).

TABLE 1 Parameters estimated from a least mean-square fit to normalized daughter cell trajectories

Gene	Cell pairs	$\langle \mathcal{V}_E \rangle$	$var(\mathcal{V}_E)$	γ
HMGA2	16	0.634	0.06	2.18
SET	37	0.687	0.053	1.59
TOP1	69	0.663	0.062	1.75
USP7	22	0.697	0.061	12.19

value. Linear regression gives an estimate for the corresponding values of $\langle \mathcal{V} \rangle_E$ for the different genes, which in our case are similar because of data normalization (Fig. 2 *A*, Table 1). As the average distance of daughter cell trajectories (see Fig. 2 *B*) do not show signature of strong contributions from short-time stochastic variations of mRNA synthesis and decay—which would be reflected by a linear increase (see Eq. 7)—we expect that the main contributions to cell-to-cell variability arise either from promoter activation (with $\gamma_A \ll T_G^{-1}$) or from extrinsic noise sources or combinations of both. Using Eq. 12, the explicit expression for half the mean-squared distance between the expression levels of two daughter cells, $\sigma_{12}^2(t) = 1/2\langle (X_1(t) - X_2(t))^2 \rangle_{\mathbf{I,E}}$ (see Fig. 2 *B*), is given in the asymptotic limit $\lambda_R^- \gg T_G^{-1}$ and $\lambda_X^- \ll T_G^{-1}$ by

$$\sigma_{12}^2(t) = 2\, var(\mathcal{V}) \times \left[\frac{t - t_0}{\gamma} - \frac{1 - e^{-\gamma(t - t_0)}}{\gamma^2} - \frac{1}{2\gamma^2}\left(1 - e^{-\gamma(t - t_0)}\right)^2 \right]. \quad (15)$$

Note that the two daughter cells have undergone the same cellular history and we therefore receive no noise contribution from previous generations. This is reflected by the absence of the generation time, T_G. We also assumed that noise is dominated by a single process with relaxation rate γ. For the case of long-term extrinsic correlation, $\gamma \to 0$, Eq. 15 simplifies to $\sigma_{12}^2(t) = (2/3) var(\mathcal{V}) \gamma (t - t_0)^3$, whereas short-lived fluctuations result in $\sigma_{12}^2 = 2 var(\mathcal{V}) \gamma^{-1} (t - t_0)$ (see Eq. 8). The unknown parameters $var(\mathcal{V})$ and γ can be determined in time units of T_G from a least mean-square fit of Eq. 15 to the experimentally found time course for $\sigma_{12}^2(t)$ (Fig. 2 *B*) and are listed in Table 1. The average distance of the daughter cell trajectories increases faster than $(t - t_0)^2$ for three genes, as estimated from a parabola fit to the data (not shown). This indicates a very slow relaxation of the factors that drive synthesis rates, meaning that correlations in expression level can span several cell cycles (13). Moreover, contributions from fast fluctuating processes—that should be apparent by a linear increase (see Eqs. 7 and 8)—seem to be negligibly small for the genes HMGA2, TOP1, and SET. In contrast to latter proteins, USP7 seem to be less susceptible to slowly varying factors for an unknown reason.

A comparison of calculated gene expression noise, as given by the coefficient of variation $CV = \sigma_E / \langle X \rangle_{\mathbf{I,E}}$, with σ_E from Eq. 13, gives excellent agreement with the measured variance of cell-to-cell variability of Sigal et al. (13) (Fig. 2 *C*). To discriminate between intrinsic and extrinsic origins of gene expression noise, we simulated the trajectories of protein abundance numerically, using the estimated values for $var(\mathcal{V})$ and γ from Table 1 and compared them with the experimentally measured trajectories of the SET gene. The results, assuming all noise to be extrinsic (Fig. 3 *B*), are in good agreement with experimental time courses (Fig. 3 *A*). Assuming noise to be exclusively a consequence of fluctuations in promoter activation shows dead-times in protein synthesis that are not observed in experiments (Fig. 3 *C*). The trajectories of the genes TOP1 and HMGA2 (not shown) show almost identical dynamic behavior to the trajectory for the SET gene, indicating that, also for these two genes, extrinsic noise contributes strongly.

Although the simulated trajectories differ significantly for intrinsic and extrinsic noise sources for the SET gene (see Fig. 3, *B* and *C*), we find almost identical stationary cell-to-cell distributions (Fig. 4). The reason is that the process describing promoter activation and the process for extrinsic factors (Eq. 11) share an exponentially decaying autocorrelation function. The two parameters determining variance and relaxation time of the autocorrelation function are taken from Table 1 and are thus equal for both processes. Thus, to leading order, effect of bursts and dead-times in mRNA synthesis cannot be distinguished from the continuously varying synthesis rates arising from extrinsic factors after

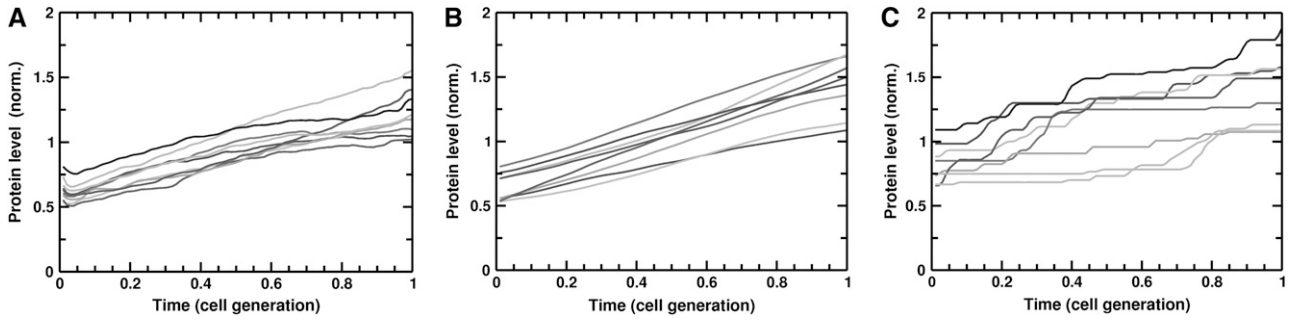

FIGURE 3 Relative amount of synthesized SET proteins using a random sample of eight data sets. (*A*) Experimental data of Sigal et al. (13). (*B*) Trajectories using exclusively extrinsic noise contributions calculated from Eq. 11 with constants γ_E and $var(\mathcal{V}_E)$ from Table 1. (*C*). Trajectories using exclusively intrinsic noise from promoter activation with γ_A and $var(\mathcal{V}_A)$ taking the same values as γ_E and $var(\mathcal{V}_E)$ in panel *B*.

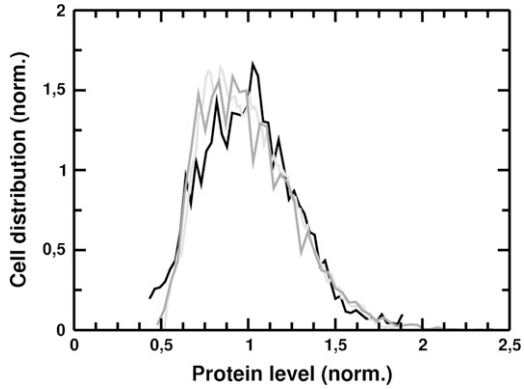

FIGURE 4 Distribution of normalized proteins levels, averaged over all measured time points within the cell cycle. Set of 38 measured trajectories of one clone for the protein SET (*solid line*) taken from Sigal et al. (13) in comparison with in silico generated trajectories assuming all noise arising from promoter activation (*dark shaded line*) or from extrinsic factors (*light shaded line*).

averaging over trajectories. This shows that experiments on cell-cycle averaged distributions of protein levels are not suitable for a clear discrimination of extrinsic and intrinsic noise contributions such that time-resolved expression data seems to be indispensable. Furthermore, a much deeper insight into the origins of cell-to-cell variations can be gained from a time-dependent theoretical description of gene expression than from a stationary description alone.

DISCUSSION

For three of the investigated four human genes, we find that, essentially, all intercellular variation in expression levels exists because of extrinsic factors that vary either the components involved in the regulation of the reporter gene or the kinetic rates associated with its expression. The fact that these extrinsic factors change on timescales of the generation time results in a significant cell-to-cell variation with standard deviation of ~20% of the mean level. Using Eq. 13, the contribution to cell-to-cell variations in protein levels can be separated into stochastic events that occur entirely in the actual generation, $CV_a^2 = I_1(t-t_0)/\langle X\rangle_{I,E}$; stochastic correlations across the last cell division, $CV_c^2 = I_2(t-t_0)/\langle X\rangle_{I,E}$; and stochastic events that have occurred before the last cell division, $CV_p^2 = [I_1(T_G)+I_2(T_G)]/(3\langle X\rangle_{I,E})$. The corresponding time-averaged partial contributions to the total cell-to-cell variation, $\langle CV^2\rangle_t$, are listed in Table 2.

TABLE 2 Fractional contributions to cell-cycle-averaged total variance, $\langle CV^2\rangle_t$

Gene	$\langle CV^2\rangle_t$	$\%\langle CV_a^2\rangle_t$	$\%\langle CV_c^2\rangle_t$	$\%\langle CV_p^2\rangle_t$
HMGA2	0.036	30%	20%	50%
SET	0.034	27%	23%	50%
TOP1	0.04	28%	22%	50%
USP7	0.007	47%	5%	48%

Here, we also applied Eq. 13 to the gene USP7, to illustrate that the $\langle CV^2\rangle_t$ of genes with short-time stochastic fluctuations do receive significant contributions from previous generations. However, correlations decline quickly in this case, leading to a small contribution of CV_c^2. We have shown in this article that trajectories of protein synthesis in single cells contain necessary information for the discrimination of intrinsic and extrinsic noise contributions that could not be otherwise deduced by cell-cycle averaged distributions of protein levels. However, this time-resolved information is by no means sufficient. To gain deeper insight into the underlying mechanisms contributing to intrinsic noise, additional information on mRNA synthesis events and mRNA lifetimes is required that, it is hoped, will be available in the future.

SUPPLEMENTARY MATERIAL

To view all of the supplemental files associated with this article, visit www.biophysj.org.

This work was financially supported by the Deutsche Forschungsgemeinschaft Emmy Noether-Program and the International Graduate School ''Signal Systems in Plant Model Organisms'' (DFG grant No. GRK1305, to J.R.).

We thank A. Cohen, A. Sigal, R. Milo, and U. Alon for helpful comments and for making their experimental data publicly available and C. Fleck for carefully reading the manuscript. M.K. acknowledges help from A. Raj and J.R. thanks J. Timmer, C. Kreutz, and F. Geier for support and useful discussions.

REFERENCES

1. Raser, J. M., and E. K. O'Shea. 2004. Control of stochasticity in eukaryotic gene expression. *Science*. 304:1811–1814.
2. Kœrn, M., T. C. Elston, W. J. Blake, and J. J. Collins. 2005. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet*. 6:451–454.
3. Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain. 2002. Stochastic gene expression in a single cell. *Science*. 297:1183–1186.
4. Volfson, D., J. Marciniak, W. J. Blake, N. Ostroff, L. S. Tsimring, and J. Hasty. 2006. Origins of extrinsic variability in eukaryotic gene expression. *Nature*. 439:861–864.
5. Bar-Even, A., J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, and N. Barkai. 2006. Noise in protein expression scales with natural protein abundance. *Nat. Genet*. 38:636–643.
6. Becskei, A., B. B. Kaufmann, and A. van Oudenaarden. 2005. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat. Genet*. 37:937–944.
7. Raj, A., C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi. 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol*. 4:e309.
8. Blake, W. J., M. K. A. Ern, C. R. Cantor, and J. J. Collins. 2003. Noise in eukaryotic gene expression. *Nature*. 422:633–637.
9. Pedraza, J. M., and A. van Oudenaarden. 2005. Noise propagation in gene networks. *Science*. 307:1965–1969.
10. Golding, I., J. Paulsson, S. M. Zawilski, and E. C. Cox. 2005. Real-time kinetics of gene activity in individual bacteria. *Cell*. 123:1025–1026.
11. Mettetal, J. T., D. Muzzey, J. M. Pedraza, E. M. Ozbudak, and A. van Oudenaarden. 2006. Predicting stochastic gene expression dynamics in single cells. *Proc. Natl. Acad. Sci. USA*. 103:7304–7309.

12. Newman, J. R., S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, and J. S. Weissman. 2006. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*. 441:840–846.

13. Sigal, A., R. Milo, A. Cohen, N. Geva-Zatorsky, Y. Klein, Y. Liron, N. Rosenfeld, T. Danon, N. Perzov, and U. Alon. 2006. Variability and memory of protein levels in human cells. *Nature*. 444: 643–646.

14. Rosenfeld, N., J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz. 2005. Gene regulation at the single-cell level. *Science*. 307: 1962–1965.

15. Thattai, M., and A. van Oudenaarden. 2001. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA*. 98:8614–8619.

16. Swain, P. S., M. B. Elowitz, and E. D. Siggia. 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA*. 99:12795–12800.

17. Swain, P. S. 2004. Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *J. Mol. Biol*. 344:965–966.

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

19. Paulsson, J. 2005. Models of stochastic gene expression. *Phys. Life Rev*. 2:157–175.

20. Friedman, N., L. Cai, and X. S. Xie. 2006. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett*. 97:168302.

21. Raser, J. M., and E. K. O'Shea. 2005. Noise in gene expression: origins, consequences, and control. *Science*. 309:2010–2013.

22. Tanase-Nicola, S., P. B. Warren, and P. R. ten Wolde. 2006. Signal detection, modularity, and the correlation between extrinsic and intrinsic noise in biochemical networks. *Phys. Rev. Lett*. 97:068102.

23. Gardiner, C. W. 2004. Handbook of Stochastic Methods. Springer, Berlin.