Noise propagation in an integrated model of bacterial gene expression and growth

Istvan T. Kleijn[*, 1, 2], Laurens H. J. Krah[*, 1], and Rutger Hermsen[†, 1]

[1]Theoretical Biology, Department of Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2]Biomathematics, Department of Mathematics, Imperial College London, Sir Ernst Chain Building, London SW7 2AZ, United Kingdom

Abstract

In bacterial cells, gene expression, metabolism, and growth are highly interdependent and tightly coordinated. As a result, stochastic fluctuations in expression levels and instantaneous growth rate show intricate cross-correlations. These correlations are shaped by feedback loops, trade-offs and constraints acting at the cellular level; therefore a quantitative understanding requires an integrated approach. To that end, we here present a mathematical model describing a cell that contains multiple enzymes that are each expressed stochastically and jointly limit the growth rate. Conversely, metabolism and growth affect protein synthesis and dilution. Thus, expression noise originating in one gene propagates to metabolism, growth, and the expression of all other genes. Nevertheless, under a small-noise approximation many statistical quantities can be calculated analytically. We identify several routes of noise propagation, illustrate their origins and scaling, and establish important connections with metabolic control analysis. We then present a many-protein model parameterized by previously measured protein abundance data and demonstrate that the predicted cross-correlations between gene expression and growth rate are in broad agreement with published measurements.

Keywords: bacterial growth, stochasticity, modeling, metabolism, control analysis.

* I.T.K. and L.H.J.K. contributed equally.

† Corresponding author. Email: r.hermsen@uu.nl

The authors declare that they have no conflict of interest.

Few processes are more fundamental to life than the growth and proliferation of cells. Bacterial cells in particular are highly adapted to grow rapidly and reliably in diverse habitats (Ingraham et al., 1983). This requires them to dynamically adapt to changing conditions. Indeed, the composition of bacterial cells depends strongly on their growth environment (Hui et al., 2015; Schmidt et al., 2016). At the same time, the composition of individual cells grown in a constant environment fluctuates vigorously, in part due to the stochastic nature of gene expression (McAdams and Arkin, 1997; Thattai and van Oudenaarden, 2001; Elowitz et al., 2002). Many experimental and theoretical studies have shed light on the origins, characteristics and consequences of this "noisy" expression (McAdams and Arkin, 1997; Thattai and van Oudenaarden, 2001; Swain et al., 2002; Elowitz et al., 2002; Pedraza and van Oudenaarden, 2005; Rosenfeld et al., 2005; Cai et al., 2006; Friedman et al., 2006; Yu et al., 2006; Levine and Hwa, 2007; Shahrezaei and Swain, 2008; Bruggeman et al., 2009; Oyarzún et al., 2015; Wolf et al., 2015). However, it remains unknown to what extent, and by what routes, noise propagates through the cell to affects its rate of growth (Tănase-Nicola and ten Wolde, 2008; Kiviet et al., 2014; Hashimoto et al., 2016).

A clear understanding of these processes is complicated by the fact that gene expression and growth are interdependent as a result of regulatory feedback and fundamental cellular constraints (Maaloe, 1969; Tan et al., 2009; Klumpp et al., 2009; You et al., 2013; Klumpp and Hwa, 2014; Hui et al., 2015; Schmidt et al., 2016). The importance of this interplay is highlighted by a series of recent studies. In particular, experiments on *E. coli* cultures in balanced exponential growth have revealed striking linear relations between their proteomic composition and their growth rate (Scott et al., 2010; You et al., 2013; Klumpp and Hwa, 2014; Hui et al., 2015; Schmidt et al., 2016). Phenomenological models have demonstrated how such "growth laws" can be understood as near-optimal solutions to constrained allocation problems (Molenaar et al., 2009; Scott et al., 2014; Bosdriesz et al., 2015; Maitra and Dill, 2015). For instance, a cell's capacity for protein synthesis is fundamentally constrained by

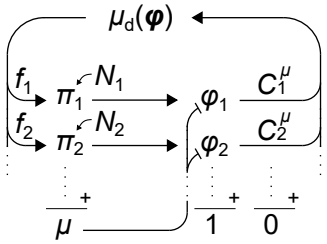

Figure 1: Integrated model of stochastic gene expression and cell growth. The cell contains many protein species, with proteome mass fractions ϕ_i that sum to 1. Mass fractions are increased by protein synthesis but diluted by growth. The synthesis rate π_i of each species i is modulated by a noise source N_i. The instantaneous growth rate μ reflects the total rate of protein synthesis. Proteins affect metabolism and thus the expected growth rate $\mu_d(\phi)$, as quantified by growth-control coefficients C_i^μ. A fraction f_i of the total metabolic flux is allotted to the synthesis of protein i. The inherent noise in the expression of each gene reverberates through the cell, affecting cell growth and the expression of every other gene.

its ribosome density (Maaloe, 1969; Vind et al., 1993). By assuming that (translational) resources are allocated optimally, one can explain the observed negative relation between the expression of carbon-catabolic enzymes and growth rate under variation of the available carbon substrate(s) (You et al., 2013). These results stress that global physiological variables strongly affect the expression of individual genes. They suggest a "holistic" perspective: the behavior of individual components cannot be understood without some knowledge of the cell's global physiological state (Berthoumieux et al., 2013; Shahrezaei and Marguerat, 2015).

Similar conclusions follow from experiments in which the instantaneous growth of individual cells is monitored in real time (Taheri-Araghi et al., 2015). Such experiments have revealed large fluctuations in the growth rate, with coefficients of variation of the order of 25%, which in part result from noise in the concentrations of metabolic enzymes (Kiviet et al., 2014). Conversely, growth-rate fluctuations affect the concentrations of individual enzymes, because cell growth dilutes the cell's constituents (Tsuru et al., 2009). Again, a full understanding of (stochastic) gene expression and growth requires an approach that integrates expression, metabolism and growth.

Here, we present a model of growing cells in which both growth rate and gene expression fluctuate due to the stochastic synthesis of many protein species that together control the rates of metabolism and growth. We offer a highly simplified description that imposes several of the cellular constraints mentioned above. As a result, noise in the expression of each gene propagates and affects the expression of every other gene as well as the growth rate—and *vice versa*.

Below, we first introduce the generic modeling framework and its assumptions, with an excursion to the theory of growth control. We then discuss how the concentration of each protein is affected by the synthesis noise in all other proteins. This exposes a hidden assumption in a standard operational definition of intrinsic and extrinsic expression noise. Next, we explain the noise modes that characterize the noise propagation between gene expression and growth in the context of a toy model with just two proteins. Lastly, we present a many-protein model that includes 1021 protein species with experimentally measured parameters. We demonstrate that the cross-correlations functions between expression and growth rate predicted by this model capture the main features of published measurements.

Results

Modeling framework

We here discuss the key assumptions of the modeling framework (Fig. 1); see SI Text 2 for details.

The mass density of *E. coli* cells is dominated by the protein content (Bremer and Dennis, 1996) and under tight homeostatic control (Kubitschek et al., 1984). We assume that this homeostasis also eliminates long-lived protein-density fluctuations in single cells. Then, the volume of a cell is proportional to its protein mass $M := \sum_i n_i$, where n_i is the abundance (copy number) of protein i. (We ignore that different proteins have different molecular weights.) The instantaneous growth rate is then defined by $\mu := \dot{M}/M$, and the proteome fraction $\phi_i := n_i/M$ of enzyme i measures its concentration. Differentiation of ϕ_i with respect to time then yields

$$\dot{\phi}_i = \pi_i - \mu\phi_i, \tag{1}$$

where π_i is the synthesis rate per protein mass. (Here we neglect active protein degradation, which on average amounts to about 2% of the dilution rate (Maurizi, 1992).) By definition, proteome fractions obey the constraint $\sum_i \phi_i = 1$. Combined with Eq. 1 this results in

$$\mu = \sum_i \pi_i. \tag{2}$$

That is, the growth rate equals the total rate of protein synthesis.

Similar assumptions connecting the increase in biomass, the cellular growth rate, protein synthesis, and growth-mediated dilution were explored in a recent review article (de Jong et al., 2017).

Another key assumption for our model is that the cellular growth rate is an *intensive* quantity. That is: given fixed mass fractions, the expected growth rate does not depend on the cell size, as suggested by the observation that individual *E. coli* cells grow exponentially within their cell cycle (Iyer-Biswas et al., 2014; Kiviet et al., 2014). This is equivalent

to assuming that the metabolic flux scales linearly with the cell size, which is a common hypothesis in metabolic control analysis (Kacser et al., 1995). Based on this, we express the synthesis rate of protein i as:

$$\pi_i = f_i \mu_\mathrm{d}(\phi) + N_i, \tag{3a}$$

$$\mu_\mathrm{d}(\phi) \coloneqq J/M. \tag{3b}$$

The first term in Eq. 3a is an intensive function; it captures the effect of the cellular composition $\phi = (\phi_1, \phi_2, \ldots)$ on the metabolic flux J that delivers the amino acids consumed in protein synthesis. The coefficients f_i specify which fraction of this flux is allocated towards the synthesis of protein species i. Gene regulation can be included by allowing the f_i to depend on intra- and extra-cellular conditions, but below we focus on the simplest case where the f_i are constant. The second term of Eq. 3a couples each synthesis rate π_i to an independent colored (Ornstein–Uhlenbeck) noise source N_i that represents the stochasticity of both transcription and translation (Dunlop et al., 2008). We neglect other noise sources, such as the unequal distribution of molecules over daughter cells during cell division.

Combining Eq. 2 and 3a reveals that

$$\mu = \mu_\mathrm{d}(\phi) + \sum_i N_i, \tag{4}$$

which identifies $\mu_\mathrm{d}(\phi)$ as the growth rate in the zero-noise limit. Given a function $\mu_\mathrm{d}(\phi)$, Eq. 1–3 fully define the dynamics of the cell.

Fig. 1 is an illustration of the model. Noise in the synthesis of a protein species induces fluctuations in its mass fraction (Eq. 1). Through their effect on metabolism, these fluctuations propagate to the expected growth rate μ_d, which affects the synthesis of all protein species (Eq. 3). In parallel, all noise sources directly impact the growth rate μ (Eq. 4) and thus the dilution of all proteins.

Linearization under a small-noise approximation

The results below rely on the assumption that $\mu_\mathrm{d}(\phi)$ may be linearized around the time-averaged composition $\phi_\mathbf{0}$. This transforms Eq. 4 to

$$\frac{\delta\mu}{\mu_0} = \sum_i C_i^\mu \frac{\delta\phi_i}{\phi_{0,i}} + \sum_i N_i, \tag{5}$$

where $\delta\phi_i$ is the deviation of ϕ_i from its time average and $\delta\mu$ the deviation of μ from $\mu_0 \coloneqq \mu_\mathrm{d}(\phi_\mathbf{0})$. The coefficients C_i^μ are defined as

$$C_i^\mu \coloneqq \left[\frac{\phi_i}{\mu_\mathrm{d}} \frac{\partial \mu_\mathrm{d}}{\partial \phi_i}\right]_{\phi_\mathbf{0}}. \tag{6}$$

In the terminology of linear noise models, the C_i^μ are transfer coefficients: they quantify to what extent fluctuations in ϕ_i transmit to μ_d. However, they have an additional biological interpretation as growth-control coefficients, as we now discuss.

Growth-control coefficients and their sum rule

In Metabolic Control Analysis, flux control coefficients (FCCs) C_i^J quantify to what extent an enzyme concentration ϕ_i limits (controls) a flux J (Kacser et al., 1995):

$$C_i^J \coloneqq \left[\frac{\phi_i}{J} \frac{\partial J}{\partial \phi_i}\right]_{\phi_\mathbf{0}}. \tag{7}$$

Analogously, Eq. 6 defines growth control coefficients (GCCs) that quantify each enzyme's control of the growth rate. Eq. 3b links FCCs and GCCs (see also (Wortel et al., 2016), SI Text 4, and Fig. S1):

$$C_i^\mu = C_i^J - \phi_i. \tag{8}$$

Based on the assumption that fluxes are extensive quantities, a famous sum rule has been derived for FCCs (Kacser et al., 1995):

$$\sum_i C_i^J = 1. \tag{9}$$

Under our assumption that the growth rate is instead intensive, the GCCs obey a markedly different sum rule:

$$\sum_i C_i^\mu = 0. \tag{10}$$

This new sum rule articulates a delicate trade-off, where the excess of one protein implies the lack of another.

In reality, many proteins do not directly contribute to metabolism; in *E. coli*, the total proteome fraction ϕ_H occupied by such "housekeeping" proteins is estimated at around 25% (O'Brien et al., 2016; Hui et al., 2015). By definition, the FCCs of such proteins vanish. The GCC of a housekeeping protein h with mass fraction ϕ_h then follows from Eq. 8:

$$C_h^\mu = -\phi_h. \tag{11}$$

Given the sum rule for GCCs, the existence of a sector of proteins with a negative GCC implies that the sum of GCCs of all other proteins must be positive and equal to $\sum_{i\notin\mathrm{H}} C_i^\mu = \sum_{h\in\mathrm{H}} \phi_h = \phi_\mathrm{H}$.

Because the transfer coefficients of the linear noise model are in fact GCCs, the above results have implications for the noise propagation. In particular, because the sum of all transfer coefficients is zero and GCCs of housekeeping proteins are negative, some transfer coefficients must be positive. That is, even if a cell's expression levels are optimized for growth, some protein concentrations are growth-limiting and their fluctuations will affect the growth rate.

In- and extrinsic noise components

Within the above framework, many statistical properties can be calculated analytically (Dunlop et al., 2008; Kiviet et al., 2014). In particular, the noise level of the concentration of

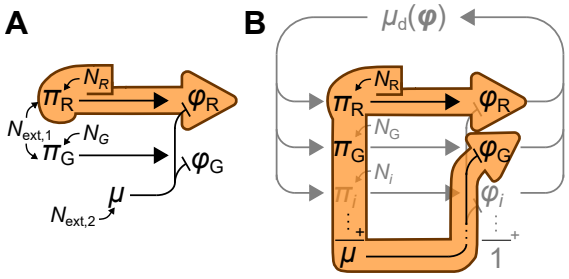

Figure 2: Limitations of the operational definition of in- and extrinsic expression noise. (*A*) Extrinsic noise is measured by the covariance between the expression levels of two identical reporter systems R and G. This presupposes that the intrinsic noise N_R of system R affects concentration ϕ_R but not ϕ_G (orange outline), so that the covariance between ϕ_R and ϕ_G quantifies the contribution of extrinsic sources $N_{ext,i}$. (*B*) But in our model, N_R affects the growth rate and thus the dilution of ϕ_G. This adds a negative term to the covariance, which no longer measures just the extrinsic noise.s

protein i, quantified by the coefficient of variation η_i, can be expressed as:

$$\eta_i^2 = (1-\phi_{0,i})^2 S_i + \sum_{j\neq i} \phi_{0,j}^2 S_j. \qquad (12)$$

Here the quantities S_i depend on the parameters of noise source i; they are always positive (see SI Text 3). Eq. 12 shows that the coefficient of variation has two components: the first term results from the noise in the synthesis of the protein itself, the second from the noise in the synthesis of all other proteins. This confirms that the inherent noise in the synthesis of one protein affects all other proteins.

A fundamental distinction is commonly made between intrinsic and extrinsic noise in gene expression (Paulsson, 2004). Intrinsic noise results from the inherently stochastic behavior of the molecular machinery involved in gene expression; extrinsic noise from fluctuations in the intra- and extracellular environment of this machinery. In this sense, the two terms in Eq. 12 can be identified as intrinsic and extrinsic contributions. Complications arise, however, if the standard operational definition of these terms is applied (Swain et al., 2002; Elowitz et al., 2002). This definition considers two identical reporter constructs R and G expressed in the same cell (Fig. 2A). Noise sources extrinsic to both reporters affect both identically, inducing positively correlated fluctuations in the concentrations of the reporter proteins. Intrinsic noise sources instead produce independent fluctuations in both concentrations. Extrinsic noise is therefore measured by the covariance between both expression levels; intrinsic noise by their expected squared difference.

This operationalization, however, implicitly assumes that intrinsic noise does not propagate between the reporters. This assumption is violated in our model because the synthesis of reporter R directly contributes to the dilution of protein G (Fig. 2B). Consequently, the covariance between the expression levels has two contributions:

$$\frac{\mathrm{Cov}(\phi_R,\phi_G)}{\phi_{0,b}^2} = \underbrace{-2\phi_{0,b}(1-\phi_{0,b})S_b}_{\text{transmission between R and G}} + \underbrace{\sum_{j\neq R,G}\phi_{0,j}^2 S_j}_{\text{other sources}}, \qquad (13)$$

where the label "b" indicates quantities that are by definition identical for both expression systems. The second term is positive and stems from noise sources that affect both reporters identically. The first term, however, is negative; it reflects the transmission of noise between reporters R and G. It would be misleading to identify Eq. 13 as the extrinsic component of the noise—it is not even guaranteed to be positive. We conclude that the operational definition is not suitable when noise propagates between arbitrary genes.

Expression–growth correlations in a two-protein toy model

The circulation of noise in the cell can be studied by measuring cross-correlations between expression and growth rate in single-cell experiments (Kiviet et al., 2014). Interpreting measured cross-correlations, however, is non-trivial. To dissect them, we now discuss a toy version of the model with just two protein species, X and Y. Despite its simplicity, it displays many features seen in more realistic models.

Within the linear noise framework, ϕ_Y–μ and π_Y–μ cross-correlations can be calculated analytically (Dunlop et al., 2008). Fig. 3AB plots them (see caption for parameters). To aid interpretations, the cross-correlations can be decomposed into the following four noise modes:

The **control mode** (Fig. 3C) reflects the control of enzyme Y on the growth rate. Noise N_Y in the synthesis of Y causes fluctuations in ϕ_Y, which transfer to the growth rate in proportion with the GCC C_Y^μ. Because the effect of ϕ_Y on μ is instantaneous, the contribution to the ϕ_Y–μ cross-correlation is *symmetric*. In contrast, the effect of π_Y on μ involves a delay; hence the contribution to the π_Y–μ cross-correlation is *asymmetric*. In both cases, the amplitude is proportional to C_Y^μ.

The **autogenic mode** (Fig. 3D) is a consequence of Eq. 2. Because the growth rate matches the total rate of protein synthesis, noise in the synthesis of Y instantly affects the growth rate, resulting in a symmetric noise mode in the π_Y–μ cross-correlation. With a delay, this noise also affects ϕ_Y, adding an asymmetric mode to the ϕ_Y–μ cross-correlation. This mode does not depend on the control of Y; instead, its amplitude is proportional to the mean concentration $\phi_{0,Y}$.

The **dilution mode** (Fig. 3E) pertains only to the ϕ_Y–μ cross-correlation. It reflects that the growth rate is also the dilution rate of protein Y (Eq. 1). With a delay, upward fluctuations in μ therefore cause downward fluctuations in ϕ_Y. A subtle complication is that noise in the synthesis rate of both

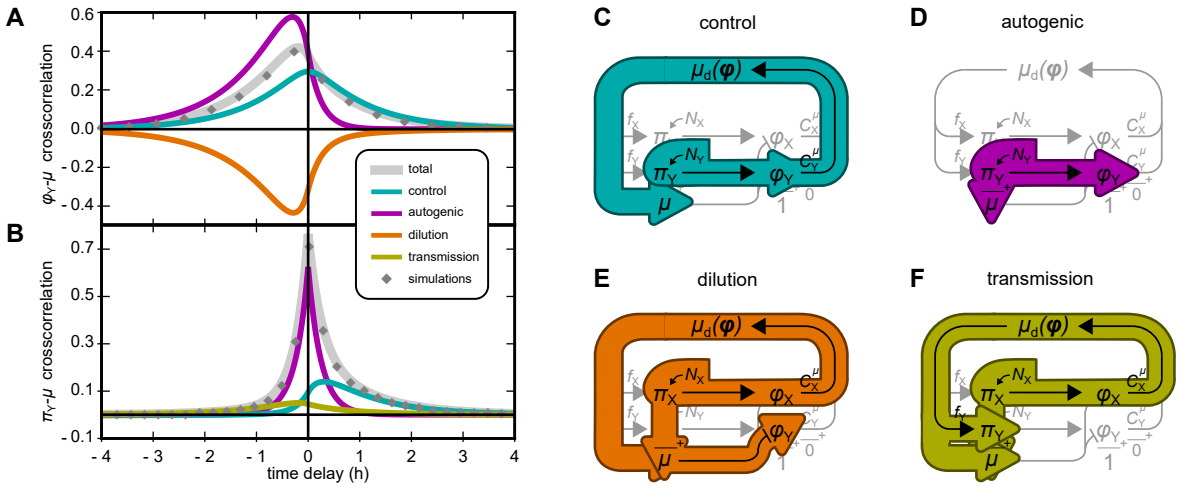

Figure 3: Noise modes in a toy model containing only two protein species, X and Y. (*A*): Analytical solution for the cross-correlation between protein Y's proteome fraction ϕ_Y and growth rate μ (gray curve), verified by simulations (gray diamonds, details in SI Text 5). The contributing noise modes are indicated (colored curves). (*B*): Same as (A), but for the synthesis rate π_Y. (*C*)–(*F*): Noise propagation routes underlying the noise modes. The control mode and the autogenic mode arise from noise source N_Y alone. Both noise sources N_X and N_Y contribute to the dilution and transmission modes, but only the contribution of N_X is illustrated in Fig. (D) and (F). Parameters for (A) and (B): $C_Y^\mu = 0.25$; $\phi_{0,Y} = 0.33$; mean growth rate $\mu_0 = 1\ \mathrm{h}^{-1}$; noise sources of N_Y and N_X have relative amplitudes $\theta_Y = 0.2$ and $\theta_X = 0.1$ and reversion rates $\beta_Y = \beta_X = 4\mu_0$.

proteins reaches μ via two routes: the immediate effect of π_Y on μ (Eq. 2), and the delayed effect of π_Y on ϕ_Y, which in turn affects μ in proportion with C_Y^μ. Together, these routes result in a mode towards which each protein contributes both a symmetric and an asymmetric function.

Lastly, the **transmission mode** (Fig. 3F) is unique to the π_Y–μ cross-correlation. It reflects that all noise sources affect the cell's composition ϕ and therefore μ_d; this in turn induces fluctuations in the synthesis rate π_Y. The noise sources also affect the growth rate via the two routes explained above, causing a symmetric and an asymmetric component to the π_Y–μ cross-correlation for each protein.

The above analysis shows that, even in a highly simplified linear model, the cross-correlations is a superposition of several non-trivial contributions. The intuitions gained from this exercise will be used below when we present the results of a more complex model.

The effects of gene regulation

Above, we assumed that the cell allocates a fixed fraction f_i of its metabolic flux towards the synthesis of protein i. Within this two-protein model all cross-correlations can still be computed if the f_i are linear(ized) functions of the concentrations ϕ (see SI Text 5 and Fig. S2). The resulting feedback regulation affects the decay of fluctuations: a negative feedback shortens the correlation time scales and reduces variance, whereas positive feedback lengthens them and increases variance (*cf.* (Thattai and van Oudenaarden, 2001; Friedman et al., 2006; Dublanche et al., 2006)).

Expression–growth correlations in a many-protein model

In single *E. coli* cells, the cross-correlations between gene expression and growth rate have been measured by Kiviet *et al.*(Kiviet et al., 2014). To test whether the above framework can reproduce their results, we constructed a model that included 1021 protein species with realistic parameters.

In the experiments, micro-colonies of cells were grown on lactulose (a chemical analog of lactose) and expression of the *lac* operon was monitored using a green fluorescent protein (GFP) reporter inserted in the operon. As fluctuations in GFP expression can affect the expression of other genes and the growth rate, we modeled this reporter construct explicitly (see Fig. 4A). Specifically, the *lac* operon O was represented as a collection of three proteins Y, Z, and G, (for LacY, LacZ, and GFP) affected by a shared noise source N_O in addition to the genes' private sources N_Y, N_Z, and N_G. The GCC of the operon as a whole is the sum of the GCCs of its genes.

By varying the mean expression of the *lac* operon with a synthetic inducer, Kiviet *et al* measured cross-correlations in three growth states with different macroscopic growth rates: "slow", "intermediate", and "fast" (Kiviet et al., 2014). Empirically, the macroscopic growth rate followed a Monod law (Monod, 1949) as a function of the mean *lac* expression. We therefore mimicked the three growth states by choosing their mean *lac* expression levels and growth rates according to three points on a Monod curve that approximated the empirical one (Fig. 4B, labels D, E, and F). Via Eq. 6, the same curve also determined the GCC of the *lac* operon in each condi-

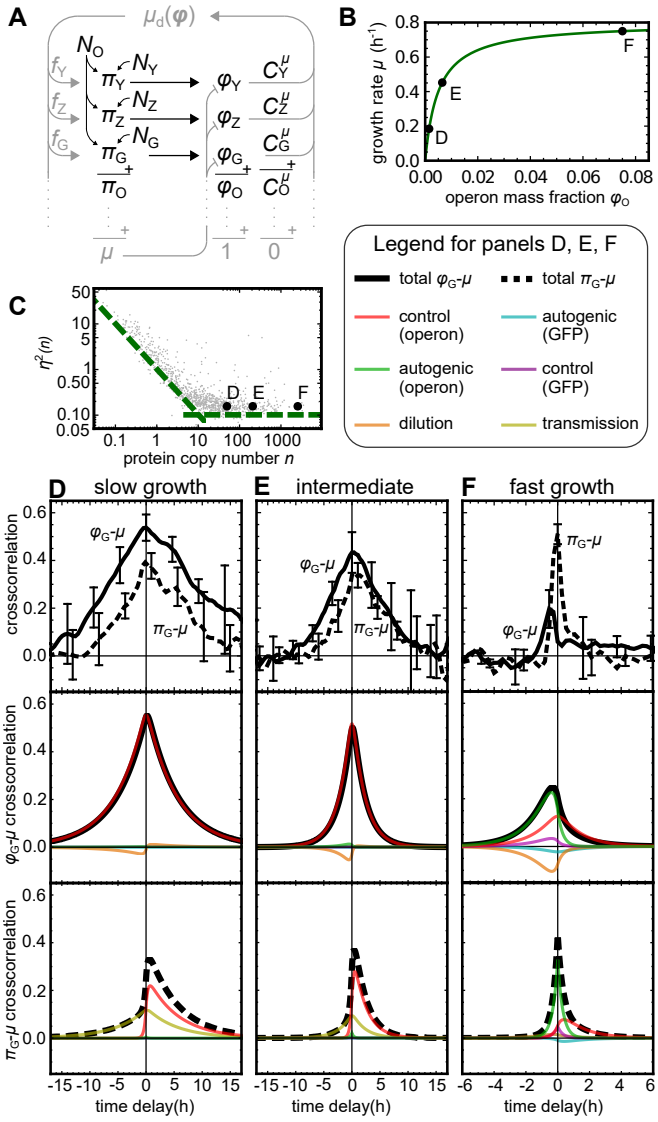

Figure 4: Expression–growth cross-correlations in the many-protein model. (*A*) Cartoon of the noise propagation network. (*B*) Monod curve describing the mean growth rate as a function of *lac* expression. Black dots indicate the operon mass fractions and growth rate used to calculate the cross-correlations in (D)-(F). (*C*) Noise distribution of the proteome (gray cloud) taken from Ref. (Taniguchi et al., 2010), and the values chosen for proteins on the *lac* operon (black dots). Green dashed lines are guides for the eye. (*D*)-(*F*) Experimental (Kiviet et al., 2014) (top panels) and theoretical (middle and bottom panels) cross-correlations for three growth conditions. Proteome fraction–growth and production–growth cross-correlations are plotted as solid and dashed black lines, respectively. As in Fig. 3AB, colored lines show the contributing noise modes.

tion. Under "slow" growth conditions, the *lac* enzymes limit growth considerably (large GCC); under "fast" conditions, *lac* activity is almost saturated (small GCC).

To choose realistic parameter values for all other proteins, we used measured means and variances of *E. coli* protein abundances (Taniguchi et al., 2010). For each of the 1018 proteins in the dataset, the model included a protein with the exact same mean and variance (see Fig. 4C). This uniquely fixed the amplitudes of all noise sources. The GCCs of all proteins were randomly sampled from a probability distribution that obeyed the sum rule of Eq. 10. (See Materials and Methods.)

Comparison with measured cross-correlations

The experimental results on the cross-correlations between GFP synthesis π_G, GFP expression ϕ_G, and growth rate μ are reproduced in Fig. 4D–F (top panels), together with the model predictions (middle and bottom panels).

The predicted cross-correlations are linear superpositions of the same noise modes as described for the two-protein model. However, the dilution and transmission modes are now driven by all 1022 noise sources, and there are two instances of the control and autogenic modes: one associated with the expression and GCC of the operon as a whole, and one with the expression and GCC of GFP separately.

At slow growth, the $\phi_G - \mu$ cross-correlation is almost symmetric (Fig. 4D, middle panel). Here the control mode of the operon dominates due to its large GCC. At higher growth rates, the autogenic modes become more prominent because their amplitudes are proportional to the expression level of the *lac* genes, whereas the amplitudes of the control modes decrease with the GCCs (Fig. 4F). As a result, the cross-correlation becomes weaker and more left-skewed.

At slow growth, the π_G–μ cross-correlation is right-skewed because the operon control mode is dominant (Fig. 4D, bottom panel). It also shows a notable transmission mode. With increasing growth rate, the autogenic modes increase in importance, which narrows the peak, increases its height, and reduces its asymmetry. The patterns seen in both cross-correlations are in good qualitative agreement with the experimental data (top panels).

Alternative dataset, similar results

In the dataset that we used to parameterize protein expression, the abundances are consistently low compared with other studies (Arike et al., 2012; Schmidt et al., 2016). However, an alternative analysis based on different abundance data (Arike et al., 2012) and sampled variances (Wolf et al., 2015) yielded similar results (SI Text 6 and Fig. S3). We conclude that the qualitative trends are insensitive to the precise dataset used.

Discussion

We have presented a model of cell growth in which the growth rate and the expression of all genes mutually affect each other. Systems in which all variables communicate to create interlocked feedback loops are generally hard to analyze. Analytical results were obtained by virtue of stark simplifying assumptions. Nevertheless, the predicted and measured cross-correlations have similar shapes and show similar trends under variation of the growth rate.

That said, a few differences are observed. Chiefly, at slow and intermediate growth rates the model consistently underestimates the decorrelation timescales (peak widths). In the model, the longest timescale is the doubling time; this timescale is exceeded in the experimental data. This suggests a positive feedback that is not included in the model, possibly as a result of gene regulation (also see Fig. S2), or else a noise source with a very long correlation time.

Alongside their measurements, Kiviet *et al.* published their own linear noise model, which fits their data well. In fact, the shapes of the noise modes of that model are mathematically identical to those presented above (Dunlop et al., 2008). Yet, the models differ strongly in their setup and interpretation. Kiviet *et al.* model a single enzyme E that is produced and diluted by growth. It features only three noise sources: one affects the production of E ("production noise"), one the growth rate μ ("growth noise"), and one affects both ("common noise"). While these ingredients are sufficient to fit the data, the interpretation and molecular origins of the common and growth noise were not specified. In our model, which includes many proteins, similar noise modes emerge without explicit growth or common noise sources. To us, each enzyme perceives fluctuations in the expression of *all* genes as noise in the growth rate; this results in a dilution mode similar to that of Kiviet *et al*. Furthermore, noise in the synthesis of each enzyme instantaneously affects the growth rate (Eq. 2) due to the assumed homeostatic control of protein density. Hence, this noise behaves as a common noise source, which explains why the autogenic mode is mathematically identical to the common-noise mode of Kiviet *et al*. We conclude that noise in the expression of many enzymes, combined with homeostatic control of protein density, can contribute to the observed but unexplained common- and growth-noise modes.

Control coefficients are routinely used in metabolic control analysis (Kacser et al., 1995; Fell, 1992; Moreno-Sánchez et al., 2008) and have also been studied in the context of evolutionary optimization (Wortel et al., 2016; Berkhout et al., 2013). In our linearized model, GCCs emerged as transfer coefficients, indicating that these quantities also affect the propagation of noise. Conversely, this suggests that GCCs could be inferred from noise-propagation measurements. For example, the Pearson correlation (cross-correlation at zero delay) between ϕ_i and μ might be used as an indication of control. However, we have seen in Fig. 3 that the ϕ–μ correlation in-

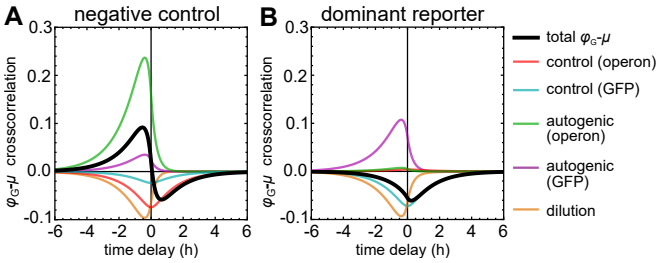

Figure 5: Deceptive concentration–growth cross-correlations. (A) Positive Pearson correlation despite a negative operon GCC, due to a dominant autogenic mode. Same parameters as Fig. 4F, but with $C^{\mu}_{\mathrm{O}} = -0.035$. (B) Negative Pearson correlation despite a positive operon GCC, due to noisy GFP expression. Same parameters as figure 4F, but with operon noise much smaller than GFP noise (see Materials and Methods).

volves several noise modes that are independent of the GCC. As a result, the signs of the Pearson correlation and the GCC do not necessarily agree (see Fig. 5A). In addition, the intrinsic noise and GCC of the reporter protein can result in a negative cross-correlation even if the operon's control is positive (Fig. 5B). Alternatively, the asymmetry of the control mode in the π–μ cross-correlation could perhaps be exploited (Kiviet et al., 2014) (Fig. S4). Unfortunately, this asymmetry is also affected by e.g. the transmission mode, which can mask the effect (Fig. S4C). We conclude that, in any case, such results have to be interpreted with great caution, ideally guided by a quantitative model.

Future theoretical work should aim to relax assumptions and remove limitations. The assumed strict control of protein density can be relaxed by allowing density fluctuations. If these are long-lived, they will likely weaken the autogenic mode and introduce new modes of their own. Also, additional noise sources can be included that do not stem directly from protein synthesis. In particular, we ignored noise originating from cell division despite its importance (Golding et al., 2005; Rosenfeld et al., 2005; Walker et al., 2016). In addition, gene regulation will affect some noise modes; this can be studied by allowing the f_i to depend on ϕ. It will also be interesting to include non-protein components of the cell, such as RNAs.

Lastly, we hope that this work will inspire new experiments that can confirm or falsify the assumptions and results presented above. In particular, single-cell measurements of mass-density of protein-density fluctuations (Grover et al., 2011; Martínez-Martín et al., 2017) could establish whether our assumption of density homeostasis is warranted. Also, additional single-cell measurements could determine whether expression noise indeed propagates between reporter proteins, adding to their covariance, and whether the amplitude of the various noise modes scales with the GCCs and mass fractions as predicted.

Methods

We here specify the parameters used for the many-protein model; also see the Supplemental Information.

Growth rates and protein abundances

The Monod curve (Fig. 4B) is given by $\mu_0 = \mu_{\max}\phi_{0,O}/(\phi_{\text{half}} + \phi_{0,O})$, with μ_0 the mean growth rate, $\phi_{0,O}$ the mass fraction of the *lac*-operon proteins, $\mu_{\max} = 0.8/\text{h}$, and $\phi_{\text{half}} = 0.005$. The three growth states correspond to three points on this curve, with $\phi_{0,O}/\phi_{\text{half}} = \{0.3, 1.3, 15\}$; this mass is shared equally among proteins Y, Z, and G. The mass fractions of the remaining proteins matched the proportions of the dataset (Taniguchi et al., 2010).

Ornstein–Uhlenbeck noise sources

The amplitudes of all noise sources were uniquely fixed by the constraints that (i) the CV of each *lac* protein was 0.15, (ii) the amplitude of N_O was 1.5× that of N_G (Elowitz et al., 2002), and (iii) all other CVs agreed with the dataset (Taniguchi et al., 2010). All noise reversion rates were $\beta = 4\mu_{\max}$.

GCCs

We first randomly assigned proteins to the "housekeeping" sector H, ($\approx$ 25% of the total mass). After the *lac* reporter construct was added, their GCC was set by Eq. 11. In each growth state, the GCC of the *lac* operon was calculated from the Monod curve, which yielded $C^\mu_O = \{0.77, 0.43, 0.063\}$. Assuming GFP is non-metabolic and the GCCs of Y and Z are equal, we set $C^\mu_G = -\phi_{0,G}$ and $C^\mu_Y = C^\mu_Z = \left(C^\mu_O - C^\mu_G\right)/2$. The GCCs of all other proteins were sampled from a probability distribution that respects Eq. 10 and assumes that proteins with a larger abundance tend to have a larger GCC (see SI Text 6).

Acknowledgements

We thank Daan Kiviet and Philippe Nghe for sharing their data.
This work was supported by the NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek) (Grant 022.005.023).

References

Arike L, Valgepea K, Peil L, Nahku R, Adamberg K, Vilu R. Comparison and applications of label-free absolute proteome quantification methods on Escherichia coli. J Proteomics, 2012;75(17):5437–5448.

Berkhout J, Bosdriesz E, Nikerel E, Molenaar D, Ridder Dd, Teusink B, et al. How biochemical constraints of cellular growth shape evolutionary adaptations in metabolism. Genetics, 2013;194(2):505–512.

Berthoumieux S, de Jong H, Baptist G, Pinel C, Ranquet C, Ropers D, et al. Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. Mol Syst Biol, 2013;9(1):634–634.

Bosdriesz E, Molenaar D, Teusink B, Bruggeman FJ. How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. FEBS J, 2015;282(10):2029–2044.

Bremer H, Dennis PP. Modulation of chemical composition and other parameters of the cell by growth rate. In: Neidhardt FC, Curtiss III R, Ingraham JL, Lin ECC, Low KB, Magasanik B, et al., editors. Escherichia coli and Salmonella, vol. 2, 2nd ed. Washington, D.C.: ASM Press, 1996.p. 1553–1569.

Bruggeman FJ, Blüthgen N, Westerhoff HV. Noise management by molecular networks. PLOS Comput Biol, 2009;5(9):e1000506.

Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. Nature, 2006;440(7082):358–362.

Dublanche Y, Michalodimitrakis K, Kümmerer N, Foglierini M, Serrano L. Noise in transcription negative feedback loops: Simulation and experimental analysis. Mol Syst Biol, 2006;2:41.

Dunlop MJ, Cox RS III, Levine JH, Murray RM, Elowitz MB. Regulatory activity revealed by dynamic correlations in gene expression noise. Nat Genet, 2008;40(12):1493–1498.

Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science, 2002;297(5584):1183–1186.

Fell DA. Metabolic control analysis: a survey of its theoretical and experimental development. Biochem J, 1992;286(Pt 2):313–330.

Friedman N, Cai L, Xie X. Linking stochastic dynamics to population distribution: An analytical framework of gene expression. Phys Rev Lett, 2006;97(16):168302.

Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. Cell, 2005;123(6):1025–1036.

Grover WH, Bryan AK, Diez-Silva M, Suresh S, Higgins JM, Manalis SR. Measuring single-cell density. Proc Natl Acad Sci USA, 2011;108(27):10992–10996.

Hashimoto M, Nozoe T, Nakaoka H, Okura R, Akiyoshi S, Kaneko K, et al. Noise-driven growth rate gain in clonal cellular populations. Proc Natl Acad Sci USA, 2016;113(12):3251–3256.

Hui S, Silverman JM, Chen SS, Erickson DW, Basan M, Wang J, et al. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. Mol Syst Biol, 2015;11(2):784.

Ingraham JL, Maaløe O, Neidhardt FC. Growth of the Bacterial Cell. Sunderland MA: Sinauer Associates, 1983.

Iyer-Biswas S, Wright CS, Henry JT, Lo K, Burov S, Lin Y, et al. Scaling laws governing stochastic growth and division of single bacterial cells. Proc Natl Acad Sci USA, 2014;111(45):15912–15917.

de Jong H, Casagranda S, Giordano N, Cinquemani E, Ropers D, Geiselmann J, et al. Mathematical modelling of microbes: metabolism, gene expression and growth. J R Soc Interface, 2017;14(136):20170502.

Kacser H, Burns JA, Fell DA. The control of flux. Biochem Soc Trans, 1995;23(2):341–366.

Kiviet DJ, Nghe P, Walker N, Boulineau S, Sunderlikova V, Tans SJ. Stochasticity of metabolism and growth at the single-cell level. Nature, 2014;514(7522):376–379.

Klumpp S, Hwa T. Bacterial growth: Global effects on gene expression, growth feedback and proteome partition. Curr Opin Biotechnol, 2014;28:96–102.

Klumpp S, Zhang Z, Hwa T. Growth rate-dependent global effects on gene expression in bacteria. Cell, 2009;139(7):1366–1375.

Kubitschek HE, Baldwin WW, Schroeter SJ, Graetzer R. Independence of buoyant cell density and growth rate in Escherichia coli. J Bacteriol, 1984;158(1):296–299.

Levine E, Hwa T. Stochastic fluctuations in metabolic pathways. Proc Natl Acad Sci USA, 2007;104(22):9224–9229.

Maaloe O. An analysis of bacterial growth. Dev Biol Suppl, 1969;3:33–58.

Maitra A, Dill KA. Bacterial growth laws reflect the evolutionary importance of energy efficiency. Proc Natl Acad Sci USA, 2015;112(2):406–411.

Martínez-Martín D, Fläschner G, Gaub B, Martin S, Newton R, Beerli C, et al. Inertial picobalance reveals fast mass fluctuations in mammalian cells. Nature, 2017;550(7677):500–505.

Maurizi MR. Proteases and protein degradation in Escherichia coli. Experientia, 1992;48(2):178–201.

McAdams HH, Arkin A. Stochastic mechanisms in gene expression. Proc Natl Acad Sci USA, 1997;94(3):814–819.

Molenaar D, van Berlo R, de Ridder D, Teusink B. Shifts in growth strategies reflect tradeoffs in cellular economics. Mol Syst Biol, 2009;5(1):323.

Monod J. The growth of bacterial cultures. Annu Rev Microbiol, 1949;3(1):371–394.

Moreno-Sánchez R, Saavedra E, Rodríguez-Enríquez S, Olín-Sandoval V. Metabolic control analysis: a tool for designing strategies to manipulate metabolic pathways. J Biomed Biotechnol, 2008;2008:597913.

Oyarzún DA, Lugagne JB, Stan GBV. Noise propagation in synthetic gene circuits for metabolic control. ACS Synth Biol, 2015;4(2):116–125.

O'Brien EJ, Utrilla J, Palsson BO. Quantification and classification of E. coli proteome utilization and unused protein costs across environments. PLoS Comput Biol, 2016;12(6):e1004998.

Paulsson J. Summing up the noise in gene networks. Nature, 2004;427(6973):415–418.

Pedraza JM, van Oudenaarden A. Noise propagation in gene networks. Science, 2005;307(5717):1965–1969.

Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB. Gene regulation at the single-cell level. Science, 2005;307(5717):1962–1965.

Schmidt A, Kochanowski K, Vedelaar S, Ahrné E, Volkmer B, Callipo L, et al. The quantitative and condition-dependent Escherichia coli proteome. Nat Biotechnol, 2016;34(1):104–110.

Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. Interdependence of cell growth and gene expression: origins and consequences. Science, 2010;330(6007):1099–1102.

Scott M, Klumpp S, Mateescu EM, Hwa T. Emergence of robust growth laws from optimal regulation of ribosome synthesis. Mol Syst Biol, 2014;10(8):747.

Shahrezaei V, Marguerat S. Connecting growth with gene expression: of noise and numbers. Curr Opin Microbiol, 2015;25:127–135.

Shahrezaei V, Swain PS. Analytical distributions for stochastic gene expression. Proc Natl Acad Sci USA, 2008;105(45):17256–17261.

Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc Natl Acad Sci USA, 2002;99(20):12795–12800.

Taheri-Araghi S, Brown SD, Sauls JT, McIntosh DB, Jun S. Single-cell physiology. Annu Rev Biophys, 2015;44(1):123–142.

Tan C, Marguet P, You L. Emergent bistability by a growth-modulating positive feedback circuit. Nat Chem Biol, 2009;5(11):842–848.

Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, et al. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science, 2010;329(5991):533–538.

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

Tsuru S, Ichinose J, Kashiwagi A, Ying BW, Kaneko K, Yomo T. Noisy cell growth rate leads to fluctuating protein concentration in bacteria. Phys Biol, 2009;6(3):036015.

Tănase-Nicola S, ten Wolde PR. Regulatory control and the costs and benefits of biochemical noise. PLoS Comput Biol, 2008;4(4):e1000125.

Vind J, Sørensen MA, Rasmussen MD, Pedersen S. Synthesis of proteins in escherichia coli is limited by the concentration of free ribosomes: Expression from reporter genes does not always reflect functional mrna levels. J Mol Biol, 1993;231(3):678–688.

Walker N, Nghe P, Tans SJ. Generation and filtering of gene expression noise by the bacterial cell cycle. BMC Biol, 2016;14:11.

Wolf L, Silander OK, van Nimwegen E. Expression noise facilitates the evolution of gene regulation. eLife, 2015;4:e05856.

Wortel MT, Bosdriesz E, Teusink B, Bruggeman FJ. Evolutionary pressures on microbial metabolic strategies in the chemostat. Sci Rep, 2016;6:29503.

You C, Okano H, Hui S, Zhang Z, Kim M, Gunderson CW, et al. Coordination of bacterial proteome with metabolism by cyclic AMP signalling. Nature, 2013;500(7462):301–306.

Yu J, Xiao J, Ren X, Lao K, Xie XS. Probing gene expression in live cells, one protein molecule at a time. Science, 2006;311(5767):1600–1603.