Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Noise propagation in bacterial cells optimized for growth

Laurens H. J. Krah and Rutger Hermsen *

* *University of Utrecht, Department of Biology, Theoretical Biology Group, the Netherlands (e-mail: l.h.j.krah@uu.nl).*

Keywords: Stochasticity, Gene Expression, Metabolic Control Analysis, Cellular Growth

Single-cell microscopy experiments have shown that the instantaneous cellular growth rate of bacterial cells fluctuates strongly (Taheri-Araghi et al. (2015)). At the same time, reporter studies have demonstrated that, within each cell, also the copy numbers of individual protein species fluctuate significantly (Taniguchi et al. (2010)), affecting the metabolic flux catalyzed by these proteins, and eventually the growth rate (Kiviet et al. (2014)). This raises the question to what extend the fluctuations in a protein's concentration contribute to the observed noise in the cellular growth rate.

A common idea is that proteins with low copy numbers are most relevant for the noisy behavior of cells, since their relative fluctuations are largest (Elowitz et al. (2002), Pedraza and Oudenaarden (2005)). In sharp contrast, we here argue that actually the proteins with a high copy number are likely to contribute most to noise in the growth rate, despite their lower noise levels.

To show this, we must consider at least two factors: (i) the strength of the protein's noise, commonly quantified as its coefficient of variation (CV, defined as standard deviation over the mean) and (ii) the sensitivity of the growth rate to fluctuations in that particular protein.

The first of these two factors has been studied extensively, both experimentally and theoretically. When expression noise is dominated by the intrinsic stochasticity of transcription and translation, the distribution of the copy number of protein species i, X_i, commonly fits a gamma distribution, where CV^2 decreases with the mean abundance, $\mathbb{E}[X_i]$ (Taniguchi et al. (2010)). For proteins with a higher mean expression, noise levels reach a plateau where noise is dominated by other, extrinsic, noise sources.

To gain insight into the second factor, we recently introduced Growth Control Coefficients (GCCs, denoted as C_i^μ): first order coefficients that quantify for each protein species i its control on the instantaneous cellular growth rate μ (Kleijn et al. (2018)). Assuming that the growth rate is fully determined given all the stochastic protein copy numbers, the GCCs were defined as:

$$C_i^\mu := \left(\frac{X_i}{\mu(\boldsymbol{X})}\frac{\partial \mu}{\partial X_i}\right)\Big|_{\mathbb{E}(\boldsymbol{X})}. \tag{1}$$

Using these GCCs, we were able to link the well-known Flux Control Coefficients (FCCs) (Kacser et al. (1995)) to noise propagation inside single cells (Kleijn et al. (2018)). We moreover derived a sum rule for the GCCs (Kleijn et al. (2018)), analogous to the sum rule for FCCs:

$$\sum_i C_i^\mu = 0. \tag{2}$$

This sum rule originates from the observation that the growth rate is to good approximation an *intensive* system variable (Kiviet et al. (2014), Taheri-Araghi et al. (2015)), as opposed to metabolic flux which is generally assumed to be *extensive*.

The GCCs appear naturally after linearization of the growth rate around a particular point, say $\widetilde{\mathbf{X}}$, assuming that expression noise is small:

$$\mu(\mathbf{X}) \approx \mu(\widetilde{\mathbf{X}})\left(1+\sum_i \left(\frac{X_i}{\mu}\frac{\partial \mu}{\partial X_i}\right)\Big|_{\widetilde{\mathbf{X}}}\left(\frac{X_i-\widetilde{X_i}}{\widetilde{X_i}}\right)\right) \tag{3}$$

When we choose $\widetilde{\mathbf{X}}$ to be $\mathbb{E}[\mathbf{X}]$, the vector with mean abundances, we read:

$$\mu(\mathbf{X}) \approx \mu(\mathbb{E}[\mathbf{X}])\left(1+\sum_i C_i^\mu\left(\frac{X_i-\mathbb{E}[X_i]}{\mathbb{E}[X_i]}\right)\right) \tag{4}$$

If protein abundances are independent, equation (4) allows the variance in the growth rate to be decomposed uniquely intro contributions of each protein. Using standard properties of variance (*i.e.*, for any scalars a, b and independent stochastic variables Y_1 and Y_2, $\mathrm{Var}\{a+bY_1+Y_2\} = b^2\mathrm{Var}\{Y_1\}+\mathrm{Var}\{Y_2\}$), we find for the coefficient of varation of the growth rate:

$$\mathrm{CV}_\mu^2 \approx \sum_i (C_i^\mu)^2 \frac{\mathrm{Var}(X_i)}{\mathbb{E}(X_i)^2} = \sum_i (C_i^\mu)^2\,\mathrm{CV}_i^2. \tag{5}$$

In this decomposition the importance of the two factors described above is directly apparent.

We note that in real cells abundances are likely to be correlated. In that case, the contribution of each protein to the noise in the growth rate cannot be defined uniquely, although alternative methods exist to analyze the noise of such systems (*e.g.*, Bowsher and Swain (2012), Thomas et al. (2018)). For the purpose of this paper, however, we choose to keep the formulation simple and intuitive and assume all protein abundances to be independent.

To now further quantify which proteins are important for growth noise, we need to gain more insight into how control is distributed between proteins. Below, we present two reasons why the GCCs are not randomly distributed across the protein species.

First, bacterial cells are known to express a set of proteins, here called H-proteins, that perform important functions but do not contribute to cellular growth. Such proteins will have a negative GCC, for although they do not produce metabolic flux, their synthesis does take up resources. This set H includes 'house-keeping' proteins involved in, *e.g.*, stress-response, immunity, and DNA damage repair; in bioengineering, H may also contain engineered pathways. Assuming H-proteins are non-toxic and do not actively hinder growth, their GCC equals minus the mean of their mass fraction, *i.e.*, $C^{\mu}_{i\in H} = -\phi_i$ (Kleijn et al. (2018)). Here we write $\phi_i := \mathbb{E}[X_i] / \sum_j \mathbb{E}[X_j]$ for the proteome mass fractions ϕ_i of each protein species and ignore, for notational simplicity, that protein species have different masses. Experiments estimating the total proteomic size of the H-set have arrived at $\phi_H \approx 25-40\%$ (OBrien et al. (2016)). In light of the sum rule (2), the H-sector has an important consequence: because the H-proteins have negative GCCs, other proteins species must have a positive GCC.

Second, natural selection tends to favor cells that on average grow faster, shaping the (mean) expression levels of proteins to be (near)-optimal for growth (Towbin et al. (2017), Dekel and Alon (2005)). Such an evolutionary optimization can mathematically be described as a constrained optimization problem, where the mean growth rate, $\mathbb{E}[\mu]$, is optimized under two constrains. First, the total cell density is fixed ($\sum_i \mathbb{E}[X_i] \equiv \Omega$) and second, only a fixed part of the proteome ($\Omega - \mathbb{E}[X_H]$) can be allocated towards proteins promoting growth. Mathematically, this can be written as:

$$\underset{\mathbb{E}[\mathbf{X}_{\notin H}]}{\text{Max}} \left[\quad \mathbb{E}[\mu] \quad \middle| \quad \sum_{i\notin H} \mathbb{E}[X_i] = \Omega - \mathbb{E}[X_H] \quad \right], \tag{6}$$

with the corresponding Lagrange Multiplier function:

$$\mathcal{L} = \mathbb{E}[\mu] - \lambda \left(\sum_{i\notin H} \mathbb{E}[X_i] - (\Omega - \mathbb{E}[X_H]) \right). \tag{7}$$

In the vector of optimal abundances, $\nabla\mathcal{L} = 0$. Using (3) to find an expression for $\mathbb{E}[\mu]$, we can calculate the required derivatives explicitly. For $i \notin H$, this results in:

$$\lambda = \frac{\partial \mathbb{E}[\mu]}{\partial \mathbb{E}[X_i]} = \left(\frac{X_i}{\mu} \frac{\partial \mu}{\partial X_i} \right)\Big|_{\widetilde{\mathbf{X}}} / \widetilde{X_i}. \tag{8}$$

Lastly, we set $\widetilde{\mathbf{X}}$, the point of linearization, equal to $\mathbb{E}[\mathbf{X}^*]$, the vector of optimal expression levels (optimality is denoted with an astriks). Rearranging gives:

$$C^{\mu^*}_{i\notin H} = \lambda \mathbb{E}[\mathbf{X}^*_i]. \tag{9}$$

Using the sum rule for the GCCs and the constraint of cellular density, it is straightforward to calculate λ. We conclude that the evolutionary process affects the distribution of growth control between metabolic proteins in a very particular way (see also Berkhout et al. (2013), Dourado and Lercher (2020)). To first order, we find:

$$C^{\mu^*}_{i\notin H} = \left(\frac{\phi_H}{1-\phi_H} \right) \phi^*_i, \qquad C^{\mu}_{i\in H} = -\phi_i. \tag{10}$$

This expression shows that, in cells optimized for growth, enzymes with a higher expression level have a higher degree of control on the growth rate.

Now that we have an indication of the distribution of control coefficients, we can use (5) to decompose the noise in the growth rate after we set the noise level for each protein. As mentioned, experiments have shown that for proteins with a low mean expression, noise is dominated by the intrinsic stochasticity of the chemical reactions involved and CV^2 scales as $\mathbb{E}(X)^{-1}$. For higher mean expression, noise levels decrease and eventually reach a plateau, where fluctuations in gene expression are dominated by extrinsic noise (Taniguchi et al. (2010)). In order to make a conservative estimate for the noise contributed by highly expressed proteins, we here choose to underestimate their noise levels and assume a fixed scaling for all proteins, *i.e.*, $\text{CV}^2_{X_i} = F/\mathbb{E}[X_i]$. Using the single parameter, F (the Fano factor), we can then set noise levels inside the cell. For *E. coli*, F is measured to be of the order of 10^0-10^2 (Taniguchi et al. (2010)).

Using this scaling, we can calculate the relative contribution of protein species i to CV^2_μ, denoted as κ_i, by substituting (10) into (5). This results in:

$$\begin{aligned} \kappa^*_{i\notin H} &:= \frac{(C^{\mu}_i)^2 \, \text{CV}^2_i}{\text{CV}^2_\mu} \\ &= \left(\frac{\phi_H}{1-\phi_H} \right) \phi^*_i. \end{aligned} \tag{11}$$

The above result is highly counter-intuitive: highly abundant proteins are actually predicted to contribute most strongly to CV_μ. Indeed, although their noise levels are lower, their increased GCCs compensate and the product of these two factors scales with the mean abundance.

We confirm this scaling in a simplistic model of cellular growth (Figure 1A). After finding the optimal growth state using a gradient-based hill-climbing algorithm, a positive relationship emerges between a protein's mean expression and its GCC (Figure 1B) and, for small enough noise levels ($F \approx 1$, $\Omega = 10^4$), also between the mean expression and κ_i (Figure 1C, red points). Importantly, such positive relationships are not seen early in the evolutionary trajectory (Figure 1B, gray lines and 1C, gray points). In fact, for systems far from optimality the opposite is found: proteins with a low expression are likely to have both a large CV and a large GCC and hence tend to contribute strongly to cellular growth noise.

In summary, we have argued that, contrary to common intuitions, highly expressed proteins are expected to be most relevant for cellular noise properties because evolution has shaped protein expression to be near optimal. These results are strengthened by the fact that in the above calculation we even underestimate the noise in said highly expressed proteins by ignoring the noise floor resulting from extrinsic noise. Interestingly, our results do not hinge on specific reaction kinetics or a specific structure of the metabolic network, since no kinetic details are assumed during the derivation of (11). Therefore, we believe the first order approximation presented here yields important general insights. More detailed studies should, however, examine noise propagation in specific systems with higher noise levels. As mentioned before, correlations between protein abundances blur the interpretation of noise contributions. Imagine two proteins whose expression correlates strongly and that jointly affect growth.

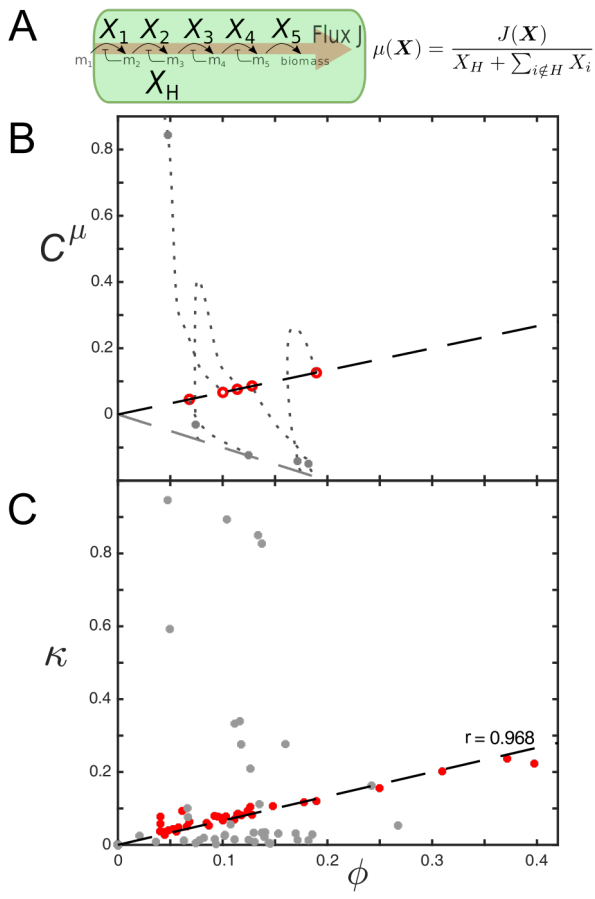

Fig. 1. **(A)** Toy model of cellular growth, where a linear reaction chain converts a fixed external metabolite (m_1) into biomass. Each protein has Michaelis-Menten-like kinetics with sampled kinetic parameters. The growth rate is defined as the steady state flux J per total expressed protein. Protein copy numbers are sampled from independent gamma distributions with evolvable means, but a fixed ratio between mean and variance ($F = 1$, $\Omega = 10^4$, $\phi_H = 0.4$). **(B)** Example trajectory of GCCs during optimisation (dotted line), ending on the predicted scaling (10) (red points on dashed line). **(C)** Predicted CV_μ^2-contributions (dashed line) compared to values measured in the simulation of 8 different 'cells' (sets of kinetic parameters) for optimal (red points), and non-optimal (gray points) mean abundances.

Attributing this noise in the growth rate to either protein is then arbitrary. Therefore, our analysis here should be interpreted in a general sense: typically a highly expressed protein is expected to contribute more growth variance than a lowly expressed protein. These observations can have important consequences for synthetic biology as well. The synthesis of byproducts often increases the set of H-proteins (Borkowski et al. (2016)), and affects all GCCs according to the sum rule (2), therewith changing cellular properties of noise propagation. Previously mentioned as one of the relevant topics in future research (Del Vecchio et al. (2018)), this work hopefully contributes to the understanding and quantification of noise propagation on a system level.

REFERENCES

Berkhout, J., Bosdriesz, E., Nikerel, E., Molenaar, D., de Ridder, D., Teusink, B., and Bruggeman, F.J. (2013). How biochemical constraints of cellular growth shape evolutionary adaptations in metabolism. *Genetics*, 194(2), 505–512. doi:10.1534/genetics.113.150631.

Borkowski, O., Ceroni, F., Stan, G.B., and Ellis, T. (2016). Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Current Opinion in Microbiology*, 33, 123–130. doi:10.1016/j.mib.2016.07.009.

Bowsher, C.G. and Swain, P.S. (2012). Identifying sources of variation and the flow of information in biochemical networks. *PNAS*, 109(20), E1320–E1328. doi: 10.1073/pnas.1119407109.

Dekel, E. and Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436(7050), 588–592. doi:10.1038/nature03842.

Del Vecchio, D., Qian, Y., Murray, R.M., and Sontag, E.D. (2018). Future systems and control research in synthetic biology. *Annual Reviews in Control*, 45, 5–17. doi:10.1016/j.arcontrol.2018.04.007.

Dourado, H. and Lercher, M.J. (2020). An analytical theory of balanced cellular growth. *Nature Communications*, 11. doi:10.1038/s41467-020-14751-w.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic Gene Expression in a Single Cell. *Science*, 297(5584), 1183–1186. doi: 10.1126/science.1070919.

Kacser, H., Burns, J.A., Kacser, H., and Fell, D.A. (1995). The control of flux. *Biochemical Society Transactions*, 23(2), 341–366. doi:10.1042/bst0230341.

Kiviet, D.J., Nghe, P., Walker, N., Boulineau, S., Sunderlikova, V., and Tans, S.J. (2014). Stochasticity of metabolism and growth at the single-cell level. *Nature*, 514(7522), 376–379. doi:10.1038/nature13582.

Kleijn, I.T., Krah, L.H.J., and Hermsen, R. (2018). Noise propagation in an integrated model of bacterial gene expression and growth. *PLOS Computational Biology*, 14(10), e1006386. doi:10.1371/journal.pcbi.1006386.

OBrien, E.J., Utrilla, J., and Palsson, B.O. (2016). Quantification and Classification of E. coli Proteome Utilization and Unused Protein Costs across Environments. *PLOS Computational Biology*, 12(6), e1004998. doi: 10.1371/journal.pcbi.1004998.

Pedraza, J.M. and Oudenaarden, A.v. (2005). Noise Propagation in Gene Networks. *Science*, 307(5717), 1965–1969. doi:10.1126/science.1109090.

Taheri-Araghi, S., Bradde, S., Sauls, J.T., Hill, N.S., Levin, P.A., Paulsson, J., Vergassola, M., and Jun, S. (2015). Cell-size control and homeostasis in bacteria. *Curr. Biol.*, 25(3), 385–391. doi:10.1016/j.cub.2014.12.009.

Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991), 533–538. doi:10.1126/science.1188308.

Thomas, P., Terradot, G., Danos, V., and Weie, A.Y. (2018). Sources, propagation and consequences of stochasticity in cellular growth. *Nat Commun*, 9(1), 1–11. doi:10.1038/s41467-018-06912-9.

Towbin, B.D., Korem, Y., Bren, A., Doron, S., Sorek, R., and Alon, U. (2017). Optimality and sub-optimality in a bacterial growth law. *Nat Commun*, 8(1), 1–8. doi: