

Work in progress

Renaud Dessalles

April 2016

Motivations

Since the beginning of the 2000s, quantitative fluorescent microscopy experiments permit to quantitatively measure cell by cell gene expression. In particular, the article [Taniguchi et al., 2010] presents a comprehensive study of messengers and proteins production in E. Coli. Both distributions of messengers and proteins of about 1000 genes have been measured. It depicts the behaviour of protein production, not only in terms of average production but also in terms of variance or noise. It appears that the variability in gene expression can be important (see Figure 1). For low expressed proteins (mean protein number < 10), the noise of proteins is considered dominated by the “intrinsic noise” regime (the noise scales inversely with the average protein production). For genes with higher protein production, the noise becomes independent of the protein production level, the plateau is around 10^{-1} ; this regime of gene expression is denoted as dominated by the “extrinsic noise”.

These data can be confronted with stochastic models of production of proteins which exists since the 70’s [Berg, 1978, Rigney and Schieve, 1977] (see the review of [Paulsson, 2005]). The basic principles of these models is that each event (creation and degradation of mRNA, creation of degradation of proteins, etc.), occurs at exponential times whose rate depends on the current state of the model. We have considered the two models commonly used in literature (see Figure 2): the constitutive gene and gene regulation models. Both these classical models only consider the production of one particular type of protein: only one gene is considered and it produces only one type of protein. We compute the parameters of these models in order to fit the mean and the variance of mRNA and protein production for each gene described in [Taniguchi et al., 2010]. *It appears that both constitutive gene and gene regulation models fail to represent properly the data obtained in [Taniguchi et al., 2010], especially for highly expressed proteins: with these models, there is no biologically relevant parameters that reproduce the “extrinsic regime” observed.*

Hence we proposed several other realistic stochastic models in order to understand the nature of the additional noise observed in [Taniguchi et al., 2010] dataset.

Models presentation and first results

Gene centred models with cell cycles For the classical models of Figure 2, one of aspects that are not considered is the cell cycles: what is modelled is a constant volume containing only one copy of the gene, where each compounds (mRNA and protein) has a exponential lifetime. In particular, there is no notion of growth of the cell, gene replication or division.

In order to better represent these aspects, we propose a gene centred model (each gene is independent from each other), which represents a gene in a growing volume that represents the interior of the

cell, and periodic division at time τ_D , at which each compounds (mRNAs and proteins) either go in one daughter cell or the other (see [Figure 3](#)). The model can be analytically analysed; in particular, the mean and variance of the concentration of mRNAs and proteins can be determined, compared with the values obtained in [[Taniguchi et al., 2010](#)]. It appears that the contribution of the cell cycle in the noise in protein expression is limited.

We are currently working on an extension of this model where the gene replication is also considered: in each cell cycle, at a constant time τ_I , the gene replicates, doubling therefore the production of mRNA. We expect to see, for some genes, a concentration of proteins that potentially highly depends on the time on the cell cycle, increasing therefore the noise of the protein expression other the cell cycle.

Multi-genes models with gene interaction One other neglected aspect of the classical models is the interactions between the different types of production units: when producing a mRNA, the gene sequesters a RNA-polymerase during a certain time, and so does a mRNA when producing a protein by sequestering a ribosome. The fluctuation in the number of free polymerases and ribosomes can have an impact on the protein production variance.

In order to quantify this effect, we propose a model where all the protein types are considered altogether. This multi-genes model consider a limited amount of RNA-polymerases and ribosomes that are shared among the genes to produced proteins (see [Figure 4](#)). Each elongation sequesters a polymerase or a ribosome during an exponential times before it is released. It is important to note here that, unlike the previous model that we introduced, there is no notion of cell-cycle here.

We are able to determine the invariant distribution of the number of free polymerases and asymptotic results gives us results on the distributions of mRNAs. We are currently working to have similar results for the proteins.

Simulation of multi-genes model with cell cycles We finally propose to merge the two previous aspects : the cell cycles and the interaction between proteins. Here we consider a growing cell, where the mRNA production rate is sensitive to the concentration of free polymerases and the rate of production of protein is sensitive to the concentration of ribosomes. Each elongation sequesters either a polymerase or a ribosome, hence decreasing the concentration of free ribosomes or polymerase. Each gene replicates at a certain point in the cell cycle, and when the cell reach a certain volume, it divides, thus distributing each of its compounds either in one daughter cell or the other.

Since analytical results are hard to obtain here, we propose to interpret this model by fitting the parameters in order to correspond with [[Taniguchi et al., 2010](#)] measures. The simulator is written and has been tested, and we are currently making exhaustive simulations; first interesting results can be found in cells whose protein concentration highly depends on the time of the cell cycle.

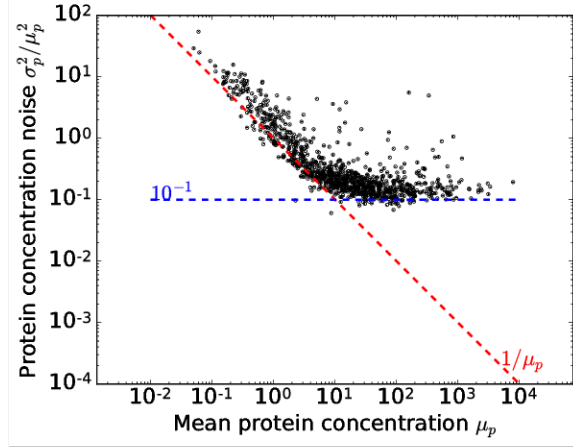
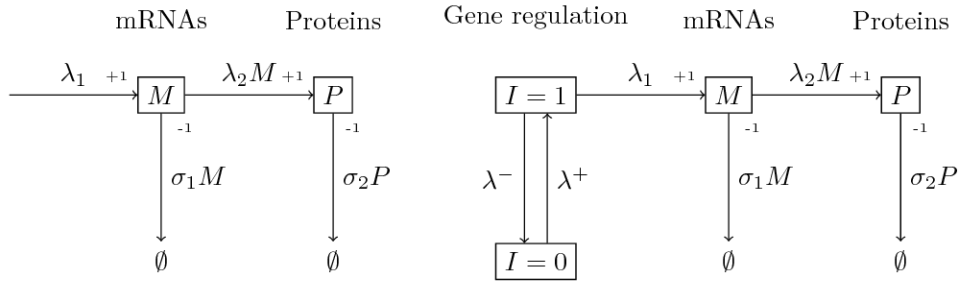


Figure 1: Protein production noise (σ_p^2/μ_p^2) function of the average protein copy number for every gene in the article [Taniguchi et al., 2010] (μ_p and σ_p denote respectively the mean and the standard deviation of protein production). For low expressed proteins (mean protein number < 10), the noise is inversely proportional to the average protein production, this part is considered lowered by an “intrinsic noise limit” (red dash). For genes with higher protein expression, the noise becomes independent of the protein expression level, protein expression is here denoted as dominated by the “extrinsic noise”.



(a) Classical model with constitutive gene. M is the number of mRNA, P is the number of proteins. Each mRNA and protein has a disappearance rate of resp. σ_1 and σ_2 ; the mRNA production rate is λ_1 , and the protein production rate per mRNA is λ_2 .

(b) Classical model with gene regulation. I represent if the gene is activated, M is the number of mRNA, P is the number of proteins. The gene activates at rate λ^+ and deactivates at rate λ^- ; when activated, the gene can produce a mRNA at rate λ_2 ; the rest is similar to the first model.

Figure 2: Classical models. Each event occurs at an exponential time whose rate depends on the current state: for instance, in both models, the rate of mRNA degradation is $M\sigma_1$ (M being the number of mRNA).

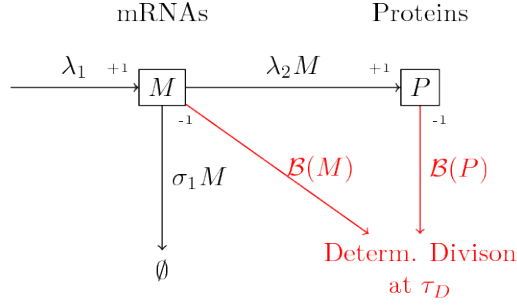


Figure 3: Model with cell cycle. The model is based on the Constitutive gene model (cf Figure 2a); here, there is no protein disappearance rate, it is replaced by a binomial segregation at division. The division occurs at deterministic time τ_D .

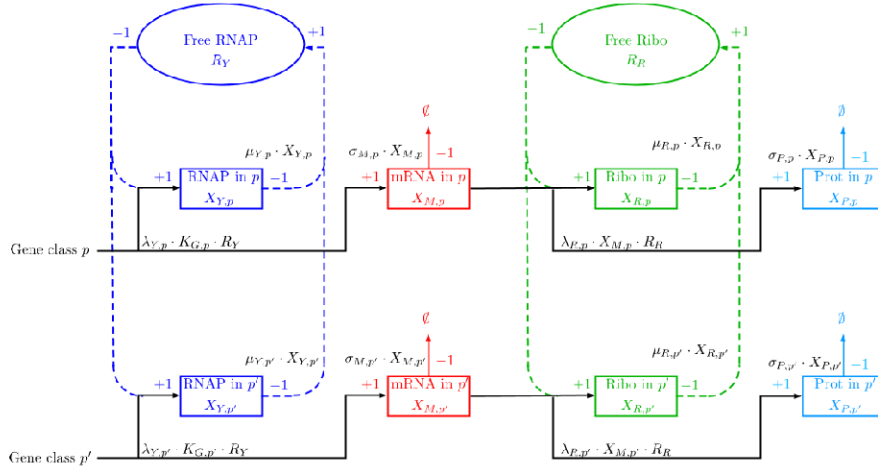


Figure 4: Multi-genes model

References

- [Berg, 1978] Berg, O. G. (1978). A model for the statistical fluctuations of protein numbers in a microbial population. *Journal of theoretical biology*, 71(4):587–603.
- [Paulsson, 2005] Paulsson, J. (2005). Models of stochastic gene expression. *Physics of Life Reviews*, 2(2):157–175.
- [Rigney and Schieve, 1977] Rigney, D. R. and Schieve, W. C. (1977). Stochastic model of linear, continuous protein synthesis in bacterial populations. *Journal of Theoretical Biology*, 69(4):761–766.

[Taniguchi et al., 2010] 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.