Division stochasticity can be transmitted to protein expression through chromosome replication

César Nieto * **César Augusto Vargas García** **
Abhyudai Singh *** **Juan Manuel Pedraza** *

* Department of physics. Universidad de los Andes. Bogotá. Colombia
** Department of mathematics and engineering. Fundación
Universitaria Konrad Lorenz. Bogotá. Colombia
(e-mail:cesara.vargasg@konradlorenz.edu.co)
*** Electrical and Computing engineering department, University of
Delaware, Newark, Delaware, USA

Abstract: Stochastic fluctuations (noise) are a fundamental characteristic of protein production. In this work, we explore how these fluctuations can originate from the stochasticity on division events. Here we consider the classical gene expression model with chromosome replication following the known Helmstetter & Cooper model. This model predicts intervals of the cell cycle where bacteria can have more than one copy of a particular gene. Considering the transcription rate as proportional to the number of chromosomes and division based on a continuous rate model, we explore how stochasticity in division or equivalently in cell size, could be transmitted to gene expression. Our simulations suggest that division can be an important source of such fluctuations only if chromosomes are replicating, otherwise, this noise is not well transmitted. This effect happens even if replication is deterministic. This work can be helpful in understanding cell cycle dynamics and their interplay with phenotypic variability.

Keywords: Stochastic hybrid systems, Control in system biology, Kinetic modeling and control of biological systems, Monte Carlo methods, Bio-signals analysis and interpretation.

1. INTRODUCTION

Stochasticity is a fundamental property of biochemical reactions (Elowitz et al. (2002); Blake et al. (2003)). In bacteria, this property can cause fluctuations in protein concentration with consequences for phenotype variability (Kaern et al. (2005)), physiology (Bressloff and Newby (2013)), sensing (Yi et al. (2000)) and self-organization (Misteli (2001)). Gene expression is particularly noisy (Elowitz et al. (2002)). It is known that part of the observed fluctuations in protein synthesis are intrinsic to the transcription and translation processes (Elowitz et al. (2002)). These intrinsic fluctuations arise mainly due to mRNA number variation since, unlike the proteins which have typically numbers in the thousands, mRNAs are present in low numbers inside the bacteria (typically of the order of ten) (Paulsson (2005)). Once protein is produced at rate proportional to the number of mRNA inside the cells, a variation in this RNA number—from ten to nine, for instance— is relatively high (almost 10% in some cases) resulting on a similar variation in protein synthesis. This effect is known as *noise transmission* because although intrinsic protein production is not so noisy (fluctuations in protein are about one over one thousand) these fluctuations come from the mRNA number fluctuations.

Other sources of noise are *extrinsic* and their nature is still under debate. This kind of noise is transmitted to all of the genetic pathways in the cell. This noise could be due to RNA polymerase number fluctuations, nucleotide availability and aminoacid concentration fluctuations, ribosome number fluctuations, among others. Some authors have pointed out division stochasticity as one of the possible causes of this noise (Amir (2014); Modi et al. (2017)) but we do not have a precise model which explicitly shows how stochasticity in division can transmit noise to gene expression.

In this work, we study how stochasticity in division can be transmitted to protein fluctuations. To do that we consider classical models of transcription and translation together with a Helmstetter & Cooper model for chromosome replication (Wallden et al. (2016); Levin and Taheri-Araghi (2019)). Division was modeled using a continuous rate model previously studied by us (Nieto et al. (2020)) consisting on division triggered by the occurrence of a fixed number of successive processes each with occurrence rate proportional to the size.

First, to study the noise transmission (noise is measured by the squared coefficient of variation C_V^2, which is the variance over the mean squared), we study how much noise in protein concentration can be due to chromosome replication and division. To do that we simulated the protein synthesis including noise selectively in each step of gene expression (division, replication, transcription and translation) for a simple genetic pathway with typical

parameters. We show that division can be an important source of noise in gene expression.

To explore specifically how noise in division can be transmitted, we use the fact that the fluctuations on division are approximately inversely proportional to number of the needed processes (steps) to divide. Then, changing this number and, therefore, changing the division noise, we explore two scenarios: one with fixed number of chromosomes and another with dynamic chromosome replication. We observe how the noise in protein concentration is affected by the tuning of noise in size. We found that the noise in division is transmitted only in the case of dynamic replicating chromosome.

2. METHODS

Gene expression is taken as a discrete stochastic process modeled by a continuous-time Markov chain. The processes involved in gene expression that are considered in this model are shown in Fig. 1
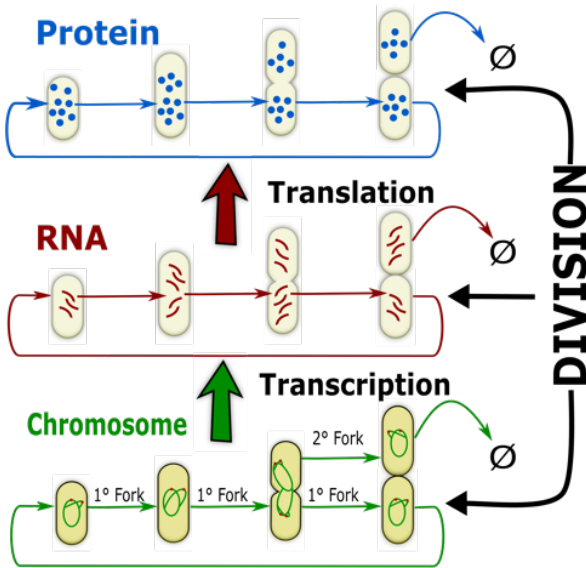

Fig. 1. Diagram of the processes involved on gene expression.

2.1 Division

As we explained in previous publications (Nieto et al. (2020); Vargas-García and Singh (2018)), division can be described by a continuous rate model. In this model, the division is triggered by the occurrence of a process of M steps. That is, once these M steps happen the cell divides. Each step can be related to the accumulation of a compound (such as FtsZ filaments to form the septum ring) and has a stochastic occurrence with continuous rate h which can be function of the size s. In the case of the *adder* strategy(Nieto et al. (2020)), the occurrence rate can be modeled simply as a constant times the size, which for later convenience we will write as:

$$h = Cs = Mk_d s \qquad (1)$$

with k_d a constant determining the division rate. Let (m) be the number of steps that have been completed at time t. If the size s grows exponentially in time with growth

rate μ and doubling time $\tau = \log(2)/\mu$, the associated deterministic equation describing the dynamics of the current number of steps completed (m) is:

$$\frac{dm}{dt} = h = Mk_d s = Mk_d s_0 e^{\mu t}, \qquad (2)$$

Once the M steps are done $(m = M)$, division occurs, the number of steps is reset form M to zero and the size is halved.

By this model, the added size Δ, defined as the difference $(\Delta = s_d - s_b)$ between size at division s_d and size at the beginning of the cell cycle (s_b), follows an Erlang distribution:

$$\rho(\Delta) = \left(\frac{Mk_d}{\mu}\right)^M \frac{\Delta^{M-1}}{(m-1)!} \exp\left(-M\frac{k_d}{\mu}\Delta\right), \qquad (3)$$

With mean added size $\langle\Delta\rangle = \frac{\mu}{k_d}$.

Thus, if the size at $t = 0$ is s_0 and n divisions happened until t, the size at that time is given by:

$$s(t; n) = \frac{s_0 \exp(\mu t)}{2^n} \qquad (4)$$

where we have considered perfectly symmetric division.

Thus, if $P_{n,m}(t)$ is the distribution defining the state with n divisions and m division steps up to time t, the master equation describing its dynamics is:

$$\frac{dP_{0,0}}{dt} = -Mk_d s_0 e^{\mu t} P_{0,0}$$

$$\frac{dP_{n,0}}{dt} = -Mk_d \frac{s_0 e^{\mu t}}{2^n} P_{n,0} + Mk_d \frac{s_0 e^{\mu t}}{2^{n-1}} P_{n,M}$$

$$\frac{dP_{n,m}}{dt} = -Mk_d \frac{s_0 e^{\mu t}}{2^n} P_{n,m} + Mk_d \frac{s_0 e^{\mu t}}{2^n} P_{n,m-1} \qquad (5)$$

As can be observed, although there must be two sibling cells after division, in our model we track only one of these cells and the other one is discarded as is shown in Fig. 1

2.2 Chromosome replication

To describe chromosome replication, we use the Helmstetter & Cooper model (H & C) (Cooper and Helmstetter (1968)). As shown in Fig. 2, a replication fork starts, on the chromosome origin site, at a fixed size per origin (s_0). It has been observed that this size is approximately constant over most of the growth conditions.

During fast growth, one replication fork starts when cells reach a size s_0, a second fork starts at size $2s_0$ and a third starts at $4s_0$ and so on. A relatively fixed time $C + D$ is needed to finish the replication. Once a time $C + D$ has passed, bacteria divide and the number of origins is halved. By this process, if bacteria grow fast multiple copies of the gene promoter can coexist inside the cell.

We model the replication as a multi-step process. The accumulation of steps $(n_R^{(i)})$ needed for replication of the ith fork is modeled by:

$$\frac{dn_R^{(i)}}{dt} = k_i \qquad (6)$$

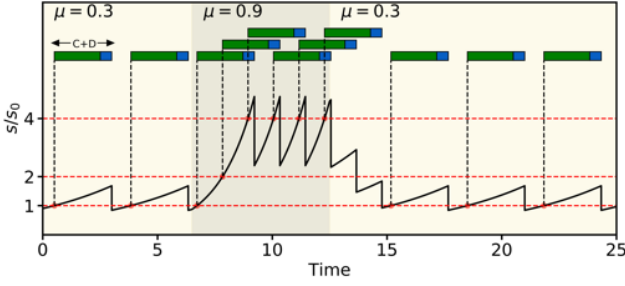where the replication rate (k_i) for the i fork is given by:

Fig. 2. Simulation of the Helmstetter & Cooper model. In slow growth ($\mu = 0.3$) only one chromosome copy is in the cell. In fast growth ($\mu = 0.9$) multiple chromosome copies can exist in the cell.

$$k_1 = N_R k_R s \theta(s - s_0)\theta(N_R - n_R)$$
$$k_2 = N_R k_R s \theta(s - 2s_0)\theta(N_R - n_R)$$
$$k_3 = N_R k_R s \theta(s - 4s_0)\theta(N_R - n_R) \tag{7}$$

with N_R the number of steps needed to finish the replication, n_R the number of these steps that the system has already done, k_R a constant defining the speed of replication rate, s the size defined by (4), and s_0 the size per origin.

The addition of the function $\theta(N_R - n_R)$ on (7) means that replication finishes, in that fork, once the current number of replication steps n_R has reached the objective number of steps N_R. The dependence of the replication rate (k_i) on the size is considered to result in an *adder* strategy for replication. This is, adding, on average, a fixed size independently of the size at birth.

During a division event, the second fork becomes the first fork, the third fork becomes second fork and so on. The number of copies of a sample gene is chosen, arbitrarily, to duplicate every time each fork reaches half of steps needed to finish each replication.

For instance, if 20 steps are needed to finish the chromosome, the first fork has done 17 steps, the second fork has done 11 steps (one greater than the half of needed steps to finish the replication) and the third fork has done 3 steps, the number of copies of the promoter n_c is: one from the original chromosome, another one because the first fork has duplicated the promoter, another two due to the second fork replication and zero due to the third fork. This is, $n_c = 4$.

Although H & C considers that division happens after a period $(C + D)$ after the replication initiation, there is evidence that division occurs in a way independent of the replication. Thus, in this work, unlike H & C, we consider the division and replication to be independent of each other.

2.3 Transcription

The classical modelPaulsson (2005) describes the concentration of RNA as a first order reaction. If the concentration of the transcription factor is fixed and its binding-unbinding rates are fast enough, the translation rate can be considered constant k_r; if degradation happens at a rate γ_r and dilution by growth happens at a rate μr, the RNA concentration r follows the differential equation:

$$\frac{dr}{dt} = k_r - \gamma_r r - \mu r. \tag{8}$$

If the process is considered stochastic, usually (8) is transformed into an *Ornstein-Uhlenbeck* process with r described as a continuous variable. However, since we are dealing mostly with discrete variables, we will base our models on discrete master equation theory.

To do that, we define the number of mRNA n_r molecules inside the cell such that the concentration is given by $r = \frac{n_r}{s}$ with s, the cell size. Equation (8) can be obtained if the dynamics of n_r is given by:

$$\frac{dn_r}{dt} = k_r s - \gamma_r n_r. \tag{9}$$

This equation implies (8) if exponential growth $\frac{ds}{dt} = \mu s$ is considered and taking advantage of the properties of the derivatives:

$$\frac{dr}{dt} = \frac{d}{dt}\left(\frac{n_r}{s}\right) = \frac{1}{s}\frac{dn_r}{dt} - \frac{n_r}{s}\frac{1}{s}\frac{ds}{dt}. \tag{10}$$

To obtain the mRNA concentration (r), one must divide this number of mRNAs (n_r) by the current size (s) which is given by (4).

During division, which happens once the number of division steps M is reached, mRNA is divided following a binomial distribution with parameter 0.5

In our general model, where the number of chromosomes may vary along the cell cycle, the transcription rate can be taken proportional to the number of copies of the gene promoter inside the cell, i.e., $k_r \rightarrow k_r n_c$.

2.4 Translation

Translation is, mathematically, a process similar to transcription. The dynamics of protein concentration (p) are given by the classical expression:

$$\frac{dp}{dt} = k_p r - \mu p, \tag{11}$$

with k_p the translation rate per unit of r and degradation is usually not considered because in bacteria most proteins are not actively degraded. The term μp comes from dilution in a similar way to (10). In a similar way to (9), the number of proteins (n_p) inside the cell follows:

$$\frac{dn_p}{dt} = k_p n_r. \tag{12}$$

This implies that translation is a Poisson process with occurrence rate $k_p n_r$. During division, the proteins can be approximated as splitting evenly because the number of proteins inside the cell is high (≈ 10000), so the binomial noise from this process is small.

2.5 Simulation of stochastic processes

Monte Carlo methods are based on the frequentist approach of probability which defines an event probability as the limit of its relative frequency in many trials. The basic algorithm can be summarized as follows:

(1) Initialize the samples in a given state x_i.

(2) Define a time interval $\Delta t \ll \tau$.

(3) Compute the transition rates $k_{x_i \to x_j} = k_{ij}$. For instance, transcription rate is k_r jumping from the state of n_r to the state $n_r + 1$.

(4) Calculate the probability of occurrence of all the stochastic events during Δt by using the occurrence rates k_{ij} and integrating $P(ocurrence) = 1 - \exp(-\int_0^{\Delta t} k_{ij}(t')dt')$.

(5) Generate a random number (R) uniformly distributed in (0,1). If this number is less than $P(ocurrence)$, the event happens and the sample jumps from $x_i \to x_j$.

(6) Deterministic variables evolve. For instance, the time $t = t + \Delta t$.

(7) Repeat steps (2)-(6) for all the cells until the maximum simulation time ($t = t_{max}$).

Statistical properties of the given distibution such as mean values $\langle x \rangle = \sum_x xP(x)$ and variance $\text{var}(x) = \sum_x (x - \langle x \rangle)^2 P(x)$ can be approximated using their non-parametric equivalents. If we obtain N_x data each with value x_i:

$$\langle x \rangle \approx \frac{1}{N_x} \sum_i^{N_x} x_i; \quad \text{var}(x) \approx \frac{1}{N_x} \sum_i^{N_x} (x_i - \langle x \rangle)^2, \quad (13)$$

which are close to the expected values as the number of trials are sufficiently large. The variability of the ramdom variables is usually described by the coefficient of variation $C_v^2(x) = \frac{\text{var}(x)}{\langle x \rangle}^2$.

We chose typical biochemical parameters. The growth rate $\mu = \ln(2)$ defines our time unit ($\tau = 1$), the division rate is chosen as $k_d = \mu$, the mean added size is $\langle \Delta \rangle = \frac{\mu}{k} = 1$, defining our unit of size. The mRNA degradation rate is $\gamma_r = 5\mu$, and k_r was set such that mean concentration of mRNA was ten per cell. k_p was set such that the mean concentration of proteins was 1000 per cell. We fixed the time step on $\Delta t = 0.001$.

Fast growth was considered. In this regime the formation of up to three forks simultaneously during replication has been observed. This can be obtained if growth rate is faster than replication rate and if the size per origin is relatively small with respect to the mean size of bacteria. Under these conditions, we set $k_R = 0.7\mu$ and $s_0 = 0.6\frac{\mu}{k_d}$. With these parameters, the average copies of the gene promoter is $\langle n_c \rangle = 1.97$. Finally, considering that replication can be a stochastic process with many steps (each step associated to a process of DNA replication) we set, arbitrarily, $N_R = 20$.

3. RESULTS

3.1 Noise transmission by the processes in gene expression

The behavior of all the variables in this work can be described by either deterministic ODEs or stochastic master equations. For instance, division has a deterministic solution using (2) or a stochastic one if (5) is used.

To determine how the total noise in gene expression can be produced by the sources considered, we sequentially set each process to be stochastic and simulate the mean noise

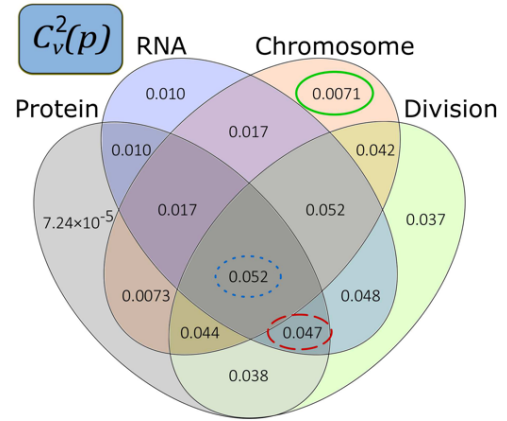in protein numbers over ten cycles. This is shown in Fig. 3

Fig. 3. Noise in protein concentration ($C_v^2(p)$) setting one or many cellular processes as stochastic. The simulation was done over 4000 cells such that the 95% confidence interval affects the significant figures after the last figure shown.

For instance, if only chromosome duplication is stochastic and the other three processes are deterministic, the resulting noise in protein concentration is $C_V^2(p) = 0.0071$ (Green continuous line in Fig. 3), if mRNA production, division and protein synthesis are stochastic and chromosome duplication is deterministic, the noise is $C_V^2(p) = 0.047$ (Red dashed line in Fig. 3). When all the sources are considered stochastic, the resulting noise is $C_V^2(p) = 0.052$ (Blue dotted line in Fig. 3).

The main result of this diagram is that when division is set as stochastic, the noise in protein concentration is relatively higher than when division is deterministic. This shows how important the division process could be as a source of noise in gene expression

3.2 The role of chromosome replication in transmitting division noise

To see how noise in division is transmitted, we explore three possible scenarios: stochastic replication following H & C, deterministic replication following this same model and fixed number of chromosomes per cell. This last mechanism can be interpreted as an exact replication when the cell divides such that the same number of chromosomes per cell is guaranteed. All of the parameters for the simulation were fixed to maintain the same average protein, RNA, size and chromosomes per cell.

Noise in division was modified changing the number of steps (M) needed to trigger division using the result from equation (3) that the noise in added size is given by:

$$C_V^2(\Delta) = \frac{1}{M} \quad (14)$$

Noise in the size does not have such as simple closed expression but is known to be proportional to $C_V^2(\Delta)$.

We ran a simulation on 1000 cells for the three models changing the division steps M estimating the noise in

both size $(C_V^2(s))$ and protein concentration $(C_V^2(p))$. The results of these simulation ares shown in Fig. 4.
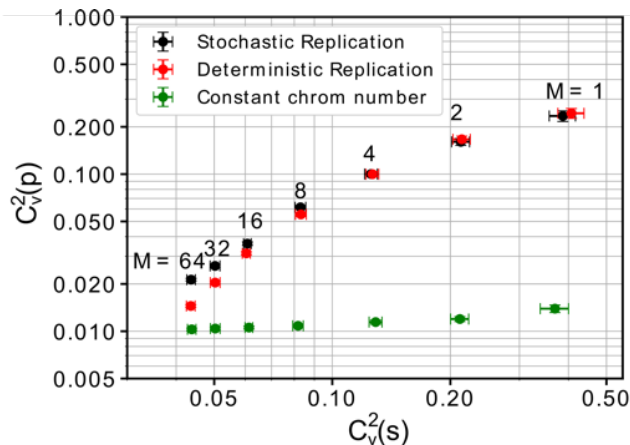

Fig. 4. Noise of protein concentration $(C_V^2(p))$ as function of the noise in size $(C_V^2(s))$ obtained changing the number of steps for division M from 1 to 64 in powers of 2. Error bas corresponds to the 95% confidence interval.

In Fig. 4, it can be seen how the noise in protein concentration, using the parameters listed bellow, is not very dependent on the stochasticity in chromosome replication because this $C_V^2(p)$ does not change appreciably when stochastic model is changed from stochastic process (black dots in Fig. 4) to deterministic one (Red dots in Fig. 4), this happens, unless the noise in division is enough suppressed ($M > 10$).

Protein noise, is highly suppressed when constant chromosome number is considered (green dots in Fig. 4) even if division is relatively noisy. In fact, the level of noise, with constant chromosome, is similar to that observed when division and chromosome are deterministic in Fig. 3 (≈ 0.01) as classical model expects. This effect shows how division stochasticity could be an important source of fluctuation in protein concentration mainly due to its transmission through replication process even if this is deterministic.

4. DISCUSSION

In this article, we have performed stochastic simulations of a model including four of the main processes involved on gene expression: transcription, translation, chromosome duplication and cell division. We explore how much noise can be transmitted to protein concentration from each of these processes. We see that considering stochastic divisions, with a noise level similar to those found in nature, can increment the noise in protein concentration almost five times when fast chromosome replication is considered. This prevision is dramatically different to the classical model of gene expression which does not take into account the division once chromosome dynamics and division have not been described as continuous Markov chain as we consider in this work.

We also explored how this division noise is transmitted to protein concentration. To do that, we compare the noise

in protein concentration changing the division noise for both a number of promoter in a replicating chromosome and a fixed number of this promoters. We observed that division noise is not transmitted to protein concentration if chromosome number does nos change along the time. This chromosome dynamics is not considered in classical models but has proven to happen experimentally (Wallden et al. (2016)).

This framework is versatile. We do not have any particular supposition of the specific nature of the chemical properties of molecules involved in all the processes. This means that this model could be used in genetically different organisms sharing similar gene expression dynamics: *E. coli* and *B. subtilis* for instance (Taheri-Araghi et al. (2015)). Adding other details to the model such as different division strategies and regulatory networks architecture can be possible for future studies bringing more features to the design of new processes involving protein production not only in bacteria but other rod-shaped organisms like yeast and archea.

Experimental Control of the the processes involved in gene expression is known to be possible (Si et al. (2018)). Translation and transcription rates can be modified by changing the chemical affinities of gene promoters ans ribosome binding sites to DNA polimerase and ribosome respectively. Division noise depends on the growth conditions and can be changed manipulating Ftsz concentration. Chromosome replication is also known to be dependent to the concentration of the replication initiator DnaA thus, an experimental study of how division noise affects the noise in gene expression through gene expression is doable.

Although chromosome replication and division noise are dependent on the growth conditions (Taheri-Araghi et al. (2015)), the effects are not well studied. For instance, in a minimal growth medium is known that division is more noisy but the number of chromosomes as seen in Fig 2, is less dynamic. This means that the noise transmitted can be very similar in both conditions: slow and fast growth. Only careful measurements can confirm our previsions.

REFERENCES

Amir, A. (2014). Cell size regulation in bacteria. *Physical Review Letters*, 112(20), 208102.

Blake, W.J., Kærn, M., Cantor, C.R., and Collins, J.J. (2003). Noise in eukaryotic gene expression. *Nature*, 422(6932), 633.

Bressloff, P.C. and Newby, J.M. (2013). Stochastic models of intracellular transport. *Reviews of Modern Physics*, 85(1), 135.

Cooper, S. and Helmstetter, C.E. (1968). Chromosome replication and the division cycle of escherichia colibr. *Journal of molecular biology*, 31(3), 519–540.

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.

Kaern, M., Elston, T.C., Blake, W.J., and Collins, J.J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics*, 6(6), 451.

Levin, P.A. and Taheri-Araghi, S. (2019). One is nothing without the other: theoretical and empirical analysis of cell growth and cell cycle progression. *Journal of molecular biology*.

Misteli, T. (2001). The concept of self-organization in cellular architecture. *J Cell Biol*, 155(2), 181–186.

Modi, S., Vargas-Garcia, C.A., Ghusinga, K.R., and Singh, A. (2017). Analysis of noise mechanisms in cell-size control. *Biophysical journal*, 112(11), 2408–2418.

Nieto, C., Arias-Castro, J., Sánchez, C., Vargas-García, C., and Pedraza, J.M. (2020). Unification of cell division control strategies through continuous rate models. *Physical Review E*, 101(2), 022401.

Paulsson, J. (2005). Models of stochastic gene expression. *Physics of life reviews*, 2(2), 157–175.

Si, F., Le Treut, G., Sauls, J.T., Vadia, S., Levin, P.A., and Jun, S. (2018). Mechanistic origin of cell-size control and homeostasis in bacteria. *bioRxiv*, 478818.

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. *Current Biology*, 25(3), 385–391.

Vargas-García, C.A. and Singh, A. (2018). Elucidating cell size control mechanisms with stochastic hybrid systems. In *2018 IEEE Conference on Decision and Control (CDC)*, 4366–4371. IEEE.

Wallden, M., Fange, D., Lundius, E.G., Baltekin, Ö., and Elf, J. (2016). The synchronization of replication and division cycles in individual e. coli cells. *Cell*, 166(3), 729–739.

Yi, T.M., Huang, Y., Simon, M.I., and Doyle, J. (2000). Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proceedings of the National Academy of Sciences*, 97(9), 4649–4653.