Politecnico
di Torino

First Passage Time statistics in cell trajectories
Politecnico di Torino
Physics of complex systems

a.a 2022/2023

Author: Andrea Siciliano
Supervisor: Carla Bosia

February 2023

1 Abstract

Timing is crucial in several cellular processes that rely on precise temporal organisation, ranging from cell-cycle control, to cell differentiation and cellular development. Here, in order to have a certain temporal order of gene expression, there are many regulatory mechanisms that not only need to be temporally organized, but whose timing fluctuations require to be tightly controlled. Indeed, since gene expression is a stochastic process, fluctuations in the amount of mRNAs or proteins naturally induce fluctuations on the time needed to reach the expression values necessary to induce downstream processes. Different regulatory strategies can be implemented in order to reduce these timing fluctuations. In this work, we are interested in the role played by microRNAs (miRNAs) in the control of timing fluctuations. miRNAs are small non-coding RNA molecules whose regulatory role in gene expression is widely recognised. In particular, by studying a stochastic model for the interaction between miRNAs and targets with Gillespie simulations, we showed that timing fluctuations can be minimized by the interplay of different sources of noise depending on the initial size of the miRNA pool. We showed that (i) if a miRNA is produced together with its target, while the production of the target increases over shorter time than a constitutive unregulated gene, timing fluctuations increase; (ii) if a pool of miRNA is already present when the miRNA-target is produced, then target expression is delayed with respect to constitutive regulation but with reduced timing fluctuations. We then compared our theoretical predictions with single-cell experimental data. In these experiments cells were transfected with bidirectional synthetic constructs containing sequences for a miRNA-target and its constitutive counterpart by means of two different fluorophores. By following their fluorescence over time at the single cell level, we obtained single-cell time trajectories for both a miRNA-target and its constitutive expression. From these single-cell time-trajectories we could then measure first-passage times and coefficient of variations, to be compared with simulations. Experimental data showed only a partial agreement with our theoretical model, thus informing on different possible alternatives to our modelling strategies.

2 Introduction

This work is a crossover between physics and biology, since we are studying the problem of the miRNA regulatory action on the timing noise in gene expression, by using some tool, most often used for physical problems, as the first-passage time. In addition, a simplificative model of protein production, mediated by miRNA, was implemented to do simulations . Then the simulations produced have been analyzed statistically in order to undestand how the noise does behave under certain conditions. As a final point the simulations results were compared with the experimental counterpart.
As already said, the first-passage time is most often exploited for physical studies, but in this case is interesting to inspect the noise characterizing the FPT in gene expression. First passage time is the first time at which a gene reaches a certain level of expression. For each simulation, due to the sthochastic nature of gene expression, one will have a different first passage time. More in general, is not well known how the cells manage to mantain perfect timing in sequentially ordered biochemical processes, for example in cellular development. The cellular control machinery is very complex and not yet fully understood.
MicroRNAs are among the regulators of gene expression , and because they play a role in regulating the noise of the number of proteins synthesized, they will play a role in decreasing timing fluctuations. The aim of the work is to find a model that well reproduces the data collected througth experiments, and to find if, under certain conditions, the noise of the first passage time reaches a minimum. The first objective is important to better understand how the miRNA regulatory machinery does work, since it is not yet fully understood. The model that is further presented tends to simplify the beheviour of the miRNA but, as one can read in the next chapters, it well reproduces experimental results in some specific cases. Despite this, the model fails to accurately represent the experimental data, and many questions remain unanswered. One idea for future studies may be to add a new chemical species to the model, ribosomes .
The second objective is to undertand if the noise in timing, has a minimum under certain condition. In particular is interesting to understand if fluctuations in time are somehow controllable and which conditions are required to reduce significantly the noise.

A gene can reach a target level of expression with a certain cell-to-cell variability, even in genetically identical populations of cells exposed to the same stimulus, due to the intrinsic noise that characterizes chemical reations. In particular if a gene has to reach a certain level of expression, what will

vary is the time in which the expression reaches it. In many cellular processes the cell need a specific protein to be expressed up to a certain threshold level, so since this threshold level is a fixed value, it is interesting to see how does varies the crossing time.
Understanding the temporal variability of gene expression is useful to better understand the basic biological mechanisms at the single cell level. Isolating the possible regulatory strategies able to control this variability can be useful to understand how the regulatory networks associated to cellular timing does work. [2]
A wrong timing in protein synthesis could also bring to disastrous effects, since many cellular processes need a set of action that must be performed in the right order and following a precise timeline.
Here we develop simulations and then analyze experimental data to study the fluctuations in time necessary to reach a target gene threshold. In particular, one of the aims of this work is to understand for which condition the model exploited in simulations is in more agreement with experimental data. The model under study has been tested for two different initial conditions.In the first case, it was assumed that miRNA is not present at time t=0, but can be synthesized from the initial time of the simulation. In the second case we assume that a nonnegligible pool of miRNAs is already present in the system at time t=0. Clearly, the cellular environment is dense and chaotic, and it cannot be predicted a priori whether a nonnegligible amount of miRNA is already present. Despite this, it is still interesting to try to see if one of the two possibilities is more inflated than its counterpart. It is also interesting to understand whether temporal noise has different behaviors depending on the initial conditions of the system, and whether it can be controlled in some way.
Another aim of the research is to understand the role of miRNA on FPT noise and if there are certain conditions for which the variability is minimal.

In the first section there will be an overview on the miRNA genesis and role in the cellular environment, in particular by focusing on its regulatory activity in gene expression.
In the second section will be presented the model that have been exploited in the simulations by a mathematical point of view. Next, individual typical trajectories are depicted, helping to better understand how the system evolves as initial conditions change. .
Then in the third section has been described the experimental setup with which has been measured the level of expression of proteins. The second section explains the work done on the experimental data, particularly how the trajectories were selected. In addition, the most significant trajectories

are illustrated and commented on.
The final section shows the results obtained after doing statistical analysis on both theoretical and experimental trajectories. in particular, statistical analysis was done on the first passage times calculated from the individual trajectories.

3 MicroRNAs and their regulatory role

3.1 microRNA biogenesis and regulation of gene expression

A cell is mainly composed by proteins, which are, not only its building blocks, but also the excecutors of cells main activities. It is clear that protein synthesis is the most important process in cell life. [1]
The genetic information of the cell is preserved in the cellular DNA, that in eukaryotic cells, is "protected" in the nucleus. When the cell needs specific proteins, in response to an external stimulus , the genetical information contained in a small portion of the DNA, called gene, is passed to the messenger RNA(mRNA), and subsequently, from mRNA, protein are syntetized.[1] In particular the process that passes genetic information from DNA to RNA is called transcription, since the genetic information contained in a portion of the DNA is copied in a different chemical species (DNA is a sequence of single pieces of deoxyribonucleic acid, RNA is a sequence of ribonucleic acid), which has almost the same language.[1] The set of enzymes that perform transcription are the RNA-polymerases.
The process that passes information from RNA to proteins is called translation and is performed by ribosomes, which is a complex made by proteins and some specific RNAs.[1]
most of the transcribed genes are mRNAs, synthetized starting from coding genes,but a portion of genes is also non-coding, and from the latter other types of RNA are produced, as the microRNAs (miRNAs). The miRNA is a very short RNA, tipically it is $\approx 20 - 22$ nucleotides long.[1] They are very important in eukaryotic cells expecially for their regulatory activity, both transcriptional and post transcriptional (in this research has been inspected the post trancriptional regulation.[1] Over 1000 different miRNA are present in human cells and regulate one third of the protein coding mRNA.[13]

MiRNAs have been found in all animal cells and some of them are highly conserved across many species (from sea urchins to humans).[1]
Indeed, has been discovered that miRNAs are present also in the nucleus and can regulate gene expression at the transcriptional level. The regulation performed by miRNA can be on transcriptional gene activation(TGA) or on transcriptional gene silencing(TGS), this means that miRNA can either initiate gene expression, or can prevent it. The first miRNA discovered to perform trancriptional gene activation was the miR-373, present in human cells, which activates the transcription of proteins CDH1 and CSDC2. Mechanisms througth which miRNA regulates gene expression at a transcriptional level is not yet fully understood. [14]

For what concerns post transcriptional regulation, in cytoplasm, miRNA regulates gene expression by linking to some specific binding sites presents on the mRNA, both preventing translation or by favourating mRNA degradation.[1] The miRNA precursor is synthetized by RNA polymerase II after which they are capped and polyadelinated, in order to increase its stability. After this the miRNA couples with some specific proteins in order to form an RNA induced silencing complex(RISC)(as depicted in Figure 1). Such complex then binds to the 3' untraslated region (3'UTR) of target mRNAs, if the latter has complementary binding sites for the specific miRNA, to induce target degradation and translational repression.[7]

If the base pairing is extensive(more often in plants cells), the RISC, once bound to a specific binding site of a mRNA, cleaves it and then rapidly unbounds and as a consequence the mRNA is rapidly digested by cellular degradation machineries .[1] Extensive base pairing means that there is a very high complementarity between RISC and binding sites present on the target mRNA. Once decoupled from the cleaved target, RISC is free to bind to another free target mRNA. In such a way, the RISC can rapidly destroy a big set of the target mRNA. [1]
If the the base pairing is not so extensive (as in animals cells) it tipically repress mRNA translation and in some rare cases it can bring to degradation. Moreover a single miRNA can bind to houndreds of different mRNAs .[1]
MiRNA-target reaction is dynamic and depends on miRNA concentration, target concentrations, miRNA-target affinity. Indeed miRNA suppression of mRNA targets is not ubiquitous between cells, even if cells are identical. RISC is supposed to move in the cytoplasm through diffusion so it is not possible a priori to know where it is located in cell. For this reasons miRNA regulatory activity can be different in two identical cells. Moreover the understanding of when and how the miRNAs exerts their functions in the nucleus is not so clear.[7]

For this many reasons also miRNA-target binding event is a stochastic process itself.

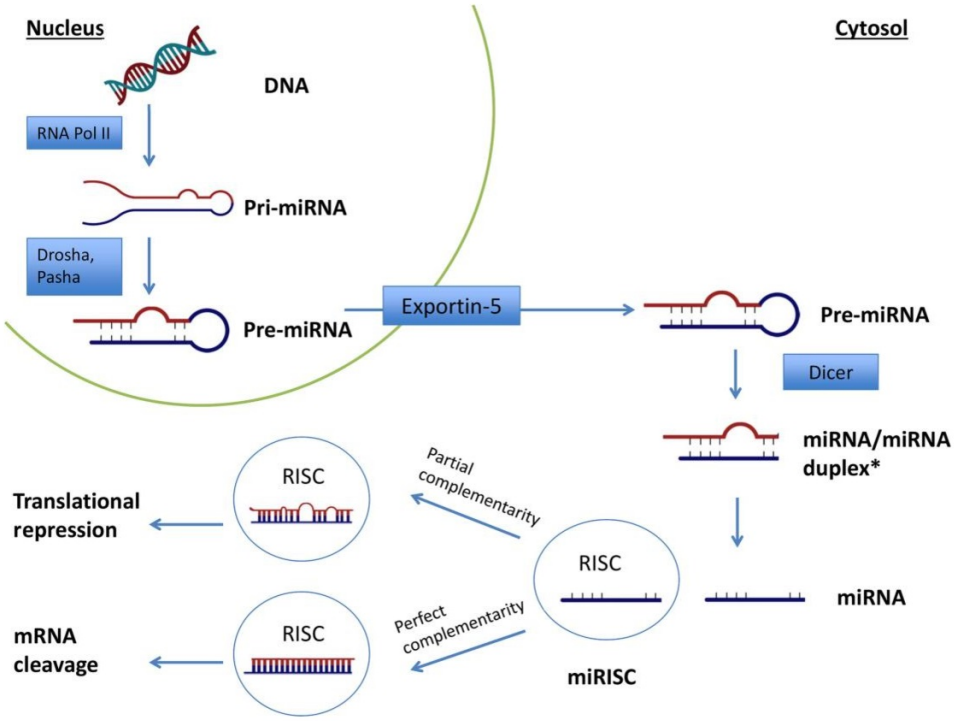

Figure 1: MiRNA transcription and mechanism of action.MiRNA is transcribed by RNA-polymerase II into the pri-miRNA. Then the pri-miRNA is again processed by the Drosha and Pasha to the pre-miRNA stage. Then the miRNA is exported in the cytoplasm by the exportin-5 througth nuclear pores. The pre-miRNA is then further processed by the Dicer into two strands, which are 20-22 nucleotides long. On of those strands then couples to other proteins to form the RISC. AS a final point the RISC will bind to the target mRNA. If the complenentarity between miRNA and mRNA is perfect the mRNA is cleaved, on the contrary if the complementarity is not perfect, mRNA is subject to translational repression. Image adapted from "miRNA in spinal muscular atrophy pathogenesis and therapy", Francesca Magri , 2018 [10]

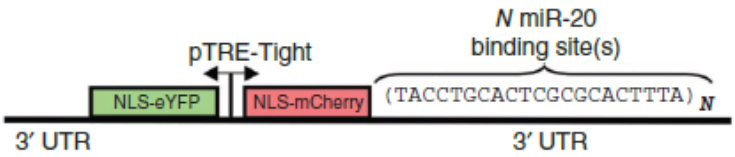

Figure 2: A simple representation of the bidirectional Tet promoter.The green block is the set of genes that express the eYFP and the red box is for the mCherry. One can notice that on the rigth of red box,attacked to it, there is the sequence of the binding sites for the miRNA(in this specific case miR-20),Image adapted from Mukierji,2011[5]

3.2 MicroRNA can generate threshold in target gene expression

As already said, miRNA regulate gene expression in the cytoplasm by promoting target mRNAs' degradation or inhibiting their translation. In particular, to prove gene expression repression performed by miRNA, is performed a single cell measurements by using quantitative fluorescence microscopy. In Mukukherji experiment has been used a two color fluorescent reporter system in order to simultaneously observe gene expression in presence and absence of miRNA mediation. Indeed they ingegnerized a bidirectional Tet-inducible promoter driving two genes expressing the fluorescent proteins mCherry and enhanced yellow fluorescent protein(eYFP). [5]
The 3'UTR of mcherry has been ingegnerized to contain N binding sites complementary to the specific miRNA added in the environment. The eYFP has been left unchanged and has no complementary binding sites for the specific miRNA already present in the cellular environment(miR-20) . Then the prepared DNA sequence has been tranfetted in the host cell. This DNA sequence is represented in figure 2.[5]

The first measure is made for 0 binding sites on the 3'UTR and as expeted the quantity of mCherry has grown accordingly to the increase of eYFP. If the binding sites are increased to 1 has been observed that the initial increase in eYFP has no corresponding increase in regulated mCherry as shown in figure 3.[5]

Those results highlights the presence of a non-linearity between mCherry level and eYFP level, since a sort of treshold appears in the expression for mCherry. Below this threshold the mCherry is highly repressed, above the the threshold the expression return similiar to the unregulated case(eYFP).

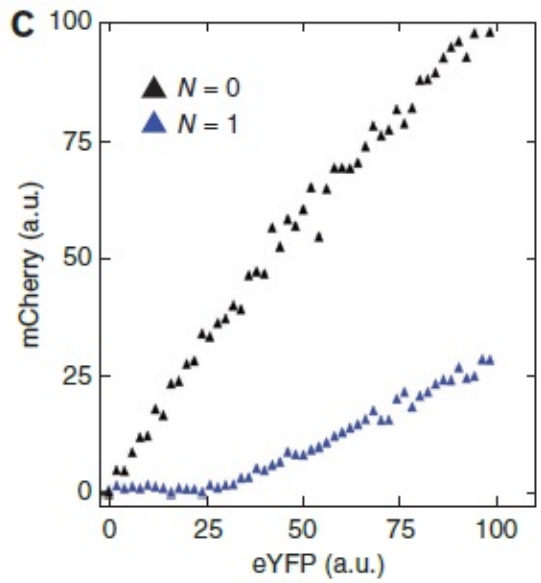

Figure 3: Scatter plot relating eYFP concentration on x axis and mCherry on y axis.The eYFP has been binned and the corresponding mCherry has been mediated for each bin. The variable N indicacetes the number of binding sites present on the target mRNA, complementar to the miR-20. As one can observe there is a delay in the expression of the mCherry(in the case of 1 binding site) with respect to the one of the eYFP highlighting the existance of the threshold behaviour,Image adapted from Mukierji,2011[5]

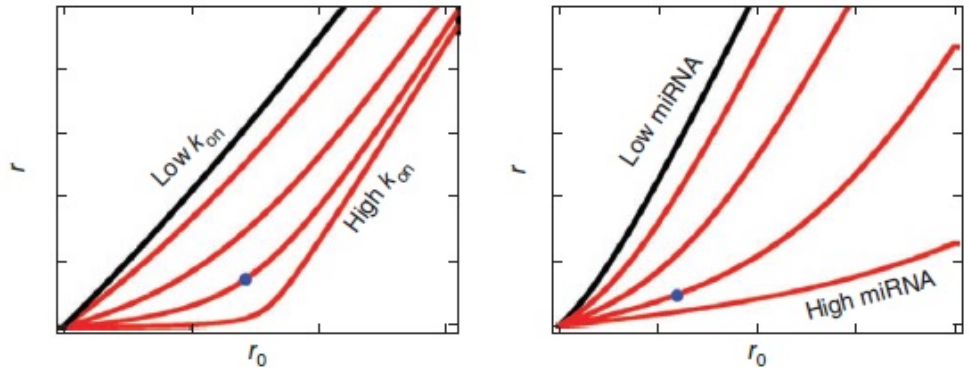

Figure 4: Steady state solution for eYFP concentration on the x axis,cslled r_0 and mCherry concentration on y axis called r.In the first graph each curve correspond to a different value of k_{on}, with k_{on} that increases from the left curve to the last curve on the right, in the second one each curve correspond to a different concentration of miRNA,Image adapted from Mukierji,2011[5]

In order to explain this behaviour had been developed a model that assumes the interaction between miRNA and mRNA to be tritative. In this experiment the miRNA level has been keep constant.[5]
The treshold level is sharpened by the increase of the number of binding sites present on the regulated portion of mRNA. In the mathetematical model the number of binding sites present on the regulated portion is modelled by the coupling rate k_{on}, which is the rate at which mRNA and miRNA couples between themselves. In this article have been plotted the concentration of the regulated gene in function of the unregulated genes for different values of k_{on}. As one can see the treshold at which the regulated gene start to be expressed, grows in function of k_{on} and also in function of miRNA concentration as depicted in figure 4. [5]

So the miRNa driven repression is strong for low target expression(consisting in very low expression below the treshold) and weak target repression for high target expression(consisting in a good level of expression above the treshold). This tritative machanism is the base on which the models implemented to perform simulations stands as we can see in next section.
As a conclusive remark this threshold behavior could be one reason for cell's phenotipic differentiation. Indeed during a developmental transition in an organims, the presence of miRNA can favour the expression of some gene despite other ones specializing the cell.

3.3 First Passage Time

First-Passage time (FPT), or first hitting time, is the time at which a stochastic or a random process first crosses a threshold. In our case, the threshold to be reached is a protein concentration fraction. In particular during the protein synthesis a steady state in protein concentration is reached after a certain time, called p_{ss}. So the crossing time at which is reached a selected fraction of the p_{ss} is the FPT.
Since gene expression is a stochastic process, the trajectories of the protein concentrations fluctuate around its mean behaviour, as illustrated in figure 5. For many different trajectories, as a consequence, will correspond many different first passage times.

One of the first work calculating the first passage time for the biological framework of our interest, is the one done by the group of research headed by Alma Dal Co. In Dal Co's paper was done something similiar to what has been done for this work, but it was inspected mainly the unregulated case. In particular, in their work, was exploited a model in which it was not present any regulation. The gene that must be expressed is switched on at some time $t_0 = 0$, from which is calculated the FPT.[2]
It was used the standard model of stochastic gene expression in which are present only two chemical species, mRNAs and proteins. The time evolution for this two chemical species is governed by two differential equation, which are the following:

$$\frac{d[P]}{dt} = k_1[m] - \gamma_1[P]$$

$$\frac{d[m]}{dt} = k_2 - \gamma_2[m]$$

Here the k_1 and γ_1 are respectively the translation rate and the protein degradation rate.k_2 and γ_2 are respectively the transcription rate and the mRNA degradation rate. The rates were choosen with the condition $\gamma_1 >> \gamma_2$ since tipically protein life is higher than the mRNA one.[2]

In the figure 5 is depicted the model schematic representation. Clearly in the simulation the initial condition for protein and mRNA is $[P](0) = 0$ and $[m](0) = 0$. Moreover the the protein expression level at the steady state is given by

$$p_{ss} = \frac{k_1 k_2}{\gamma_1 \gamma_2}$$

As a final remark, the threshold level $\tilde{(p)}$ to be crossed is expressed also as

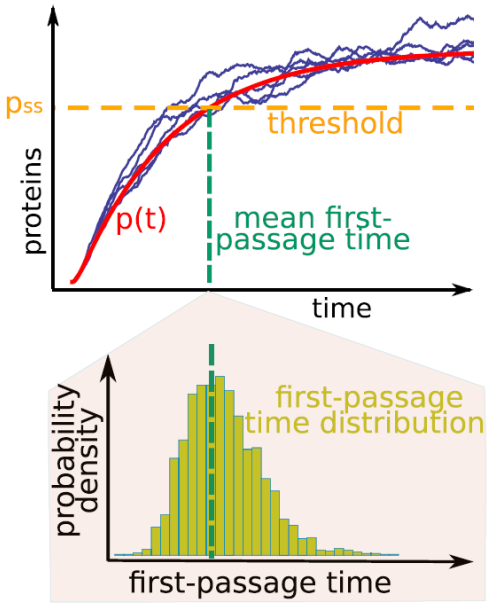

Figure 5: Definition of the threshold crossing problem. The gene is switched on at time t=0. After some time the trajectory of the protein expression reaches the steady state p_{ss}. Since the gene expression is a stochastic process, all trajectories will fluctuate around their mean behaviour. The stochastic trajectories are depicted in blue in the top image, their mean behaviour is drawn in red. The first passage time is the time at which the trajectory crosses the threshold p_{ss}. All first passage times can be collected to calculate their probability distribution depicted in the bottom image. Image adapted from Dal Co ,2017

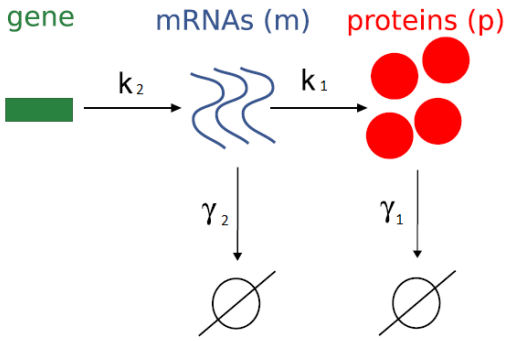

Figure 6: In the image is represented the standard model of stochastic gene expression. In particular the gene is switched on at time $t = 0$. As one can observe, starting from the initial time the gene can be transcribed to a mRNA with rate k_2, the latter is degraded with rate $\gamma_2[m]$.The protein is translated with rate$k_1[m]$ and degraded with rate $\gamma_1[P]$,Image adapted from Dal Co ,2017[2]

an adimentional parameter which depends on the p_{ss} defined as

$$\alpha = \frac{\tilde{(}p)}{p_{ss}}$$

In the paper was evaluated the coefficient of variation of the first passage time as a function of the threshold α. The coefficient of variation is the ratio between the standard deviation and the mean value, in this case, of the FPT. The CV is useful, since it is adimentional, to compare data variability, even if data sets have different dimensionalities or different mean values.
In this work does emerges the fact that the coefficient of variation of the first passage time does exhibit the presence of a minimum for certain condition. In particular the minimum is appoximatively at half of the threshold for $\alpha \approx 0.5$. In order to test the algorithms implemented for this work the same calculus was reproduced and the very same results were found. The results of the simulations are depicted in the subsequent picture.

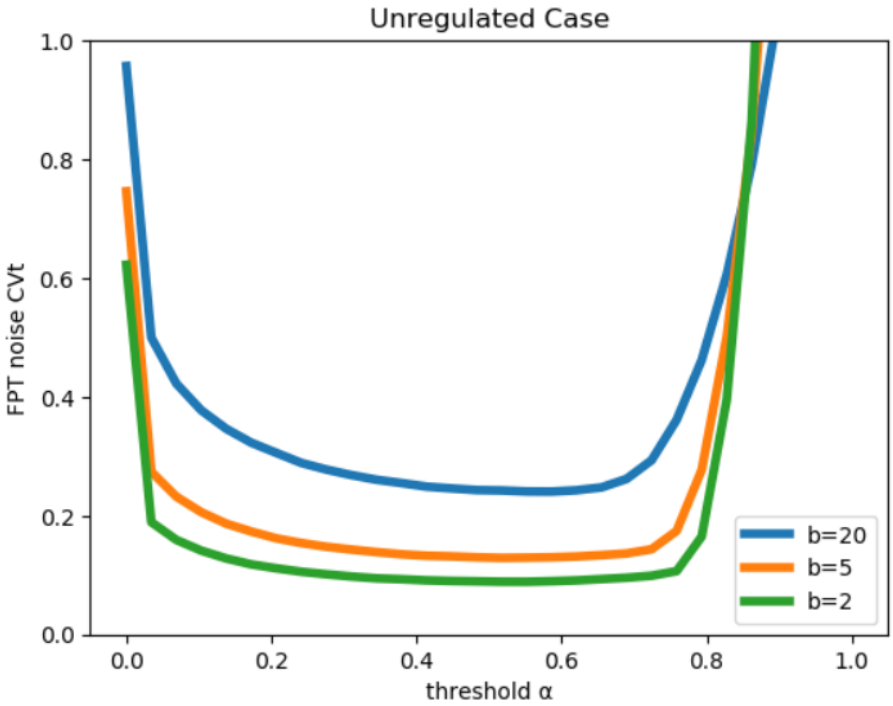

Figure 7: CV of the FPT in function of α.As one can observe there is a minimum in the CVt of the FPT at a certain value of the threshold α. In this case the minimum is at $\alpha \approx 0.5$. This result has been obtained by selecting the same rates of the article of Dal Co.

4 Algorithms and miRNA mediatied model for gene expression

4.1 Gillespie algorithm

The Gillespie algorithm has been first developed by Doob in the mid-1940s, and then lately riderived by Gillespie in 1970. After that, the algorithm became widely used. It is an algorithm mainly used to simulate the kinetics of chemical reactions and as such allows to know in detail the time evolution of the different chemical species involved in a biochemical process. [20] [19]

The main structure of the algorithm is the following:

1. Initialisation of the system. For example for biochemical reactions the initial concentrations of reactants must be initialized, the time must be initialized to zero, and a stopping time must be declared;

2. Monte Carlo. For each reaction to simulate, a putative time is randomly selected from a distribution that depends on the process one wants to simulate, then the reaction associated to the lowest time is chosen;

3. Update. The state of the system must be updated according to the reaction selected in (2);

4. Repeat. Repeat step (2) and (3) until the stopping time is reached.

For our simulations, we implemented the original Gillespie algorithm.

4.2 MiRNA-mediated regulation: the standard model

In order to analyze the experimental data, and thus better understand miRNA-mediated regulation, we studied a minimal model for miRNA-target interaction. This model is a simple evolution of the minimal model for gene expression, that accounts for gene transcription and translation, and mRNA and protein degradation. On top of these reactions, we added the miRNA regulatory activity, modelled via a titration mechanism (i.e. we added the possibility of having miRNA-target complexes that can either unbind or be degraded. As long as an mRNA is bound in complex with a miRNA, it cannot be translated). We then explored different initial conditions for this model in terms of miRNA pool size. Indeed, we account for transcription and degradation reactions for the miRNA, which can be already present in the system with a pool of a given size. The presence of a pool of miRNA

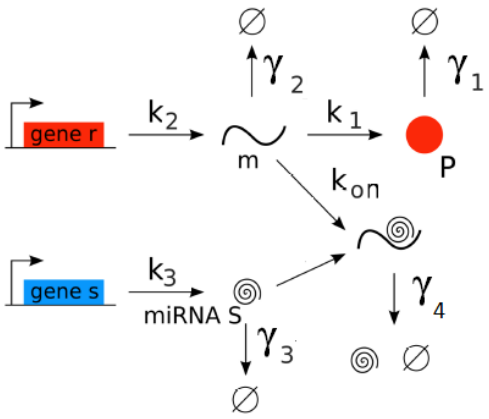

Figure 8: Schematization of the model simulated with the Gillespie algorithm. Two different genes code for an mRNA (gene r, in red) and a miRNA (gene s, in blue), which can be transcribed with rates k_2 and k_3 respectively. miRNA s and mRNA r can either be degraded with rates γ_3 and γ_2 respectively, or bind into a complex with rate k_{on}, which can in turn be degraded with rate γ_4 releasing a miRNA. Only mRNA r not bound to miRNA can be translated with rate k_1 into protein p, which can eventually be degraded with rate γ_1. Image adapted from Del Giudice, 2018. [6]

gives interesting and different out-of-steady-state outcomes with respect to a case in which the initial amount on miRNA is zero.

Since in our experimental set-up it is not known a priori if a pool of miRNA regulating the target is already present, we are interested in exploring the different predicted out-of-steady-state scenarios. In the design of this model the production of mRNA and proteins has been strongly simplified, avoinding details that account for example for proter activation/inactivation or polymerase or ribosome binding/unbinding reactions. Our miRNA/target interaction model is represented schematically in the Figure 8.

As one can notice, the miRNA has its own production rate, k_3, so the idea is that the miRNA is produced in the whole simulation. The ordinary differential equations describing this model are the following:

$$\frac{d[P]}{dt} = k_1[m] - \gamma_1[P]$$

$$\frac{d[m]}{dt} = k_2 - \gamma_2[m] - k_{on}[m][\mu] + k_{off}[C]$$

$$\frac{d[\mu]}{dt} = k_3 - \gamma_3[\mu] - k_{on}[m][\mu] + k_{off}[C] + \gamma_4(1-\gamma_5)[C]$$

$$\frac{d[C]}{dt} = k_{on}[m][\mu] - k_{off}[C] - \gamma_4(1-\gamma_5)[C] - \gamma_4\gamma_5[C]$$

Where [P],[m],[μ],[C] are respectively the protein P, mRNA m, miRNA μ and complex C concentrations.
Here many different rates have been introduced. In particular we have k_1 and γ_1 which are respectively the translation and degradation rates for protein P. k_2 and γ_2 are the transcription and degradation rates for mRNA m.
k_{on} and k_{off} are the rates of miRNA/target binding and unbinding.
k_3 and γ_3 are the transcription and degradation rates for miRNA.
Finally γ_4 and γ_5 are respectively the degradation rate for the complex C and the fraction of miRNA μ that is degraded together with complex C (so that $(1-\gamma_5)$ represents the probability that a miRNA is recycled after its interaction with the target).

The claim is to produce stochastic simulations in order to find the behaviour of the first passage time in many different regimes, depending on rates values.
As pointed out in [5], dependig on the relative amount of miRNA and target mRNA, we can distinguish three regimes: above the threshold (more mRNA than miRNA), at the treshold (mRNA and miRNA in similar amount), and below the threshold (more miRNA than mRNA). The system can be brougth to one of those regimes by tuning the target transcription ratek_2 while keeping fixed the other parameters (and in particular the miRNA transcription rate k_3). Indeed, it has been verified in [30] that the treshold level for k_2 is given by:

$$k_2 = \frac{k_3}{\gamma_5}$$

The single simulation in the "over the treshold" regime is the most similar to the constitutive (unregulated) case (a gene which is simply transcribed and translated).

Indeed in this regime the transcription rate is much higher than the one for miRNA and so in this case its regulatory action is not so evident. It is interesting the fact that by systematically decreasing the transcription rate k_2 toward the treshold level becomes more evident an overshoot.
An overshoot in an increase of the protein concentration that exceed evidently the p_{ss} in the starting times of the trajectory, then the expression level return to its steady state.
This is due, to the fact that the transcription rate is comparable but still higher than the one for miRNA, k_3, and so miRNA reaches a significative level of expression only after the mRNA. This brings to a sort of delay in

regulatory action that brings to the overshoot. Also in some trajectories taken from the experimental data this behaviour is evident, confirming this. At the treshold the overshoot is very evident, but the regulatory action of the miRNA reduces significatively the level of expression of the protein with respect the above the threshold behaviour.
Finally in the sub treshold regime the protein is almost unexpressed since its concentration oscillate between zero and very low level of expression.
Depending on γ_5 one can distinguish three different regimes.
For $\gamma_5 \approx 1$ we are in the stechiometric case in which the miRNA is almost never recycled, but is destroyed when the miRNA and mRNA decouples for the complex destruction rateγ_4.
For $\gamma_5 \approx 0.5$ we are in the intermediate case in which the miRNA has the same probability to be recycled and to be destroyed.
Finally for $\gamma_5 \approx 0$ we are in the catalytic case in which almost all the miRNA is recycled.
In those three cases the treshold varies since it does depends on $gamma_5$.
Moreover the simulations are made also in the unregulated case, but this time the stochastic data are obtained by simply putting the trancription rate of miRNA to zero, $k_3 = 0$.By doing this the miRNA is kept to zero so there is not any regulatory action.
The rates have been changed in order to keep the p_{ss} at the same level for both regulated and unregulated case. In particular the the protein mean value at the steady state, p_{ss}, in this regulated case has an explicit form, which is the following

$$p_{ss} = \frac{1}{2\gamma_1}\Big(\frac{k_1k_2}{\gamma_2} - \frac{k_1k_3}{\gamma_2\gamma_5} - \frac{\gamma_3k_1}{\gamma_5k_{on}} - \frac{\gamma_3k_1k_{off}}{\gamma_4\gamma_5k_{on}} + \frac{k_1}{\gamma_2\gamma_4\gamma_5k_{on}} *$$

$$* \sqrt{4\gamma_2\gamma_4\gamma_5k_{on}(\gamma_3\gamma_4k_2 + \gamma_3k_2k_{off}) + (-\gamma_2\gamma_3\gamma_4 - \gamma_2\gamma_3k_{off} + \gamma_4\gamma_5k_2k_{on} - \gamma_4k_3k_{on})^2}\Big)$$

The first passage time is calculated when the protein conconcentration reaches a fraction of this steady state. In figure 9 are depicted some trajectories which are the result of a single simulation for each regime mentioned earlier.

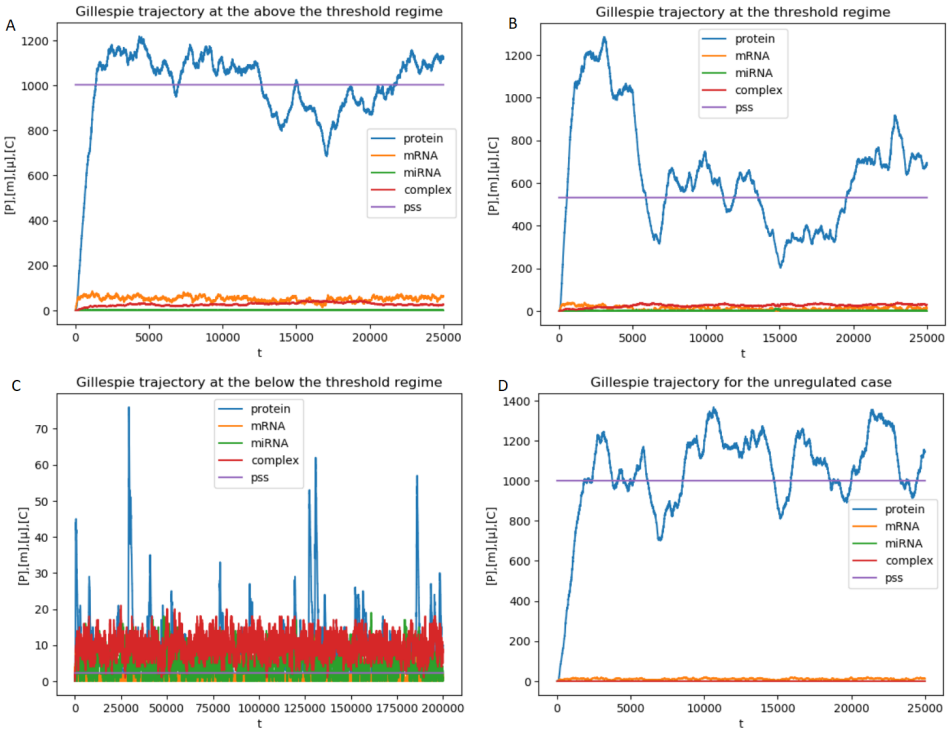

Figure 9: In the following pictures are depicted on the y axis the concentration of the four chemical species, which arein particular [P] for protein concentration, [m] for mRNA' s one, $[\mu]$ for the miRNA, and [C] for the miRNA-mRNA complex. In this case $[\mu](0) = 0$. On the y axis is also depicted the protein mean value at the steady state p_{ss} and protein concentration, after a "growth stage" , oscillates around this value. On the x axis there is time t. As one can observe image A and D are very similiar. Indeed in image A the system is above the threshold and in this case the transcription rate of the mRNA is much higher than the miRNA's one. For this reason the miRNA regulatory activity is not appreciable at all and as a consequence the two trajectories are very similiar. The main difference is the fact that the miRNA's concentration$[\mu]$ and the complex one [C] are different from zero in the case A and are zero for case D. Then the image B represent the system at the threshold. In this case is very evident the presence of the overshoot that consist in a protein overexpression in the first quarter of the trajectory since the mRNA and miRNA trancription rates are similiar but the mRNA's one is still higher than the miRNA's one. As a consequence, miRNA regulatory activity becomes evident only after a certain time . As a final point, image C represent the above the threshold regime in which protein concentration is comparable with the one of the other chemical species, and the miRNA regulatory activity is really evident.

The second set of simulations were made by changing the initial condition of miRNA amount. Indeed, in this case a pool of miRNA is already present and has a nonnegligible quantity . This new model is interesting because is not known a priori if a certain quantity of miRNA is already present in the cytoplasm.
This small difference with respect to the previous model will consist in great differences in simulations. Indeed in this case, the threshold like behaviour discussed in the Mukherji paper[Mukherji], becomes evident, and the protein starts its expression only after a certain time. Since miRNAs are already present in cellular environment in a great number, when at time $t = 0$ the gene is switched on, the protein production is stopped by this abundance. In this case, as a consequence proteins synthesis is delayed.
Another important consequence of this model is the fact that is not possible to obtain anymore an overshoot. As already said when this phenomenon was presented, the overshoot is a consequence of the fact that the mRNA trancription is higher than the miRNA's one. In this model, a pool of miRNA is already present and is not possible to observe it since it is more probable that the miRNA binds to its target mRNA, preventing the translation of proteins. As a consequence, mRNAs cannot reach a significant level of expression before the miRNA's one and the overshoot phenomenon is not observable anymore.

The protein concentration mean value at the steady state, p_{ss} is the same of the former case.
In figure 10 are depicted some trajectories characterizing the three regimes already anticipated.

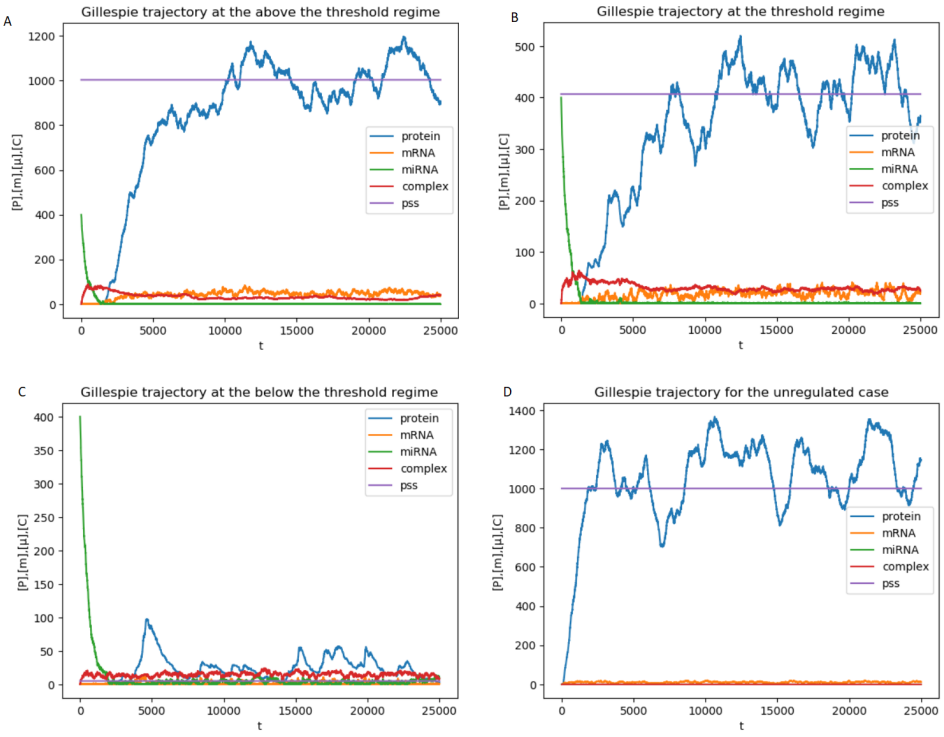

Figure 10: Characteristic trajectories produced by simulation for the case with the pool of miRNA. At the time $t = 0$ there was a pool of 400 miRNA already present in the system. As one can observe for all the three regimes, above the threshold, at the threshold and below the threshold (respectively figures A, B, C) there is a clear delay in protein expression that start to be appreciable only when the concentration of "free" miRNA present in the system becomes very small and comparable with the concentration of mRNA. Moreover, oberving figure A and B one can notice, as already said, that the overshoot effect is not possible anymore at the threshold regime(figure B). What changes between this two regimes is the effectiveness of the regulatory action of the miRNA. Indeed in the figure B the protein is more repressed than figure A. as k_2 decreases, the protein becomes less expressed. Figure C represent the below the threshold case and, as in the previous case, the protein expression highly repressed and comparable with the one of the other chemical species.

5 Experimental setup and data analysis

The experimental setup is described in detail here below.
A bidirectional synthetic construct with a Tet-inducible promoter driving the expression of two genes has been engineered by the research group I did my thesis with. As in [5], the two genes code respectively for the mCherry fluorescent protein and the enhanced yellow fluorescent protein (eYFP). The 3'UTR of mCherry may contain N binding sites complementary to miRNA miR-20a, while the 3'UTR of eYFP is left unchanged. The synthetic construct was then transfected into Human Embryonic Kidney 293 (HEK-293) cells, a cell line expressing miRNA miR-20a at high levels.
Six hours after transfections, cells were time-lapse monitored at a fluorescence microscope for 72 hours, with one frame every 30 minutes. Since $N = 0, ..., 9$, there were ten different experimental conditions monitored in parallel. A tracking algorithm developed within the research unit has then been used to track single cells over time and generate single-cell fluorescent time trajectories (with the two fluorophores measured over time for every single cell).
We obtained $\sim 20 \times 10^3$ single cell trajectories per each experimental condition, whose duration could vary from a few minutes to several hours. Out of these trajectories, we selected for further analysis only those whose level of fluorescence was below the saturation of the signal for both mCherry and eYFP, and that could contain gap with no eYFP fluorescence for no longer than 1.5 hours. Indeed, we reasoned that while a very low/almost zero fluorescence on mCherry could imply a regulatory miRNA-mediated effect, eYFP should be always present and thus sequences of zero values of fluorescence for eYFP longer than 1.5 hours could be due to bad performances of the tracking algorithm.
Single trajectories belonging to the different experimental conditions ($N = 0, ..., 9$) were collected in different matrices, each one corresponding to a different N on the mCherry sequence, and then analyzed. In Figure 10 we show a sample of representative single cell trajectories, to be compared to stochastic simulations.

Since in experiments with transient transfections it is usually assumed that the system reaches the steady state 48 hours after transfection, we evaluated the mCherry and eYFP steady states by mediating the last 9 time points of each cell trajectory per each experimental condition ($N = 0, ..., 9$). The scatter plot of these mean steady-state values for the two fluorescent proteins is shown in Figure 11.

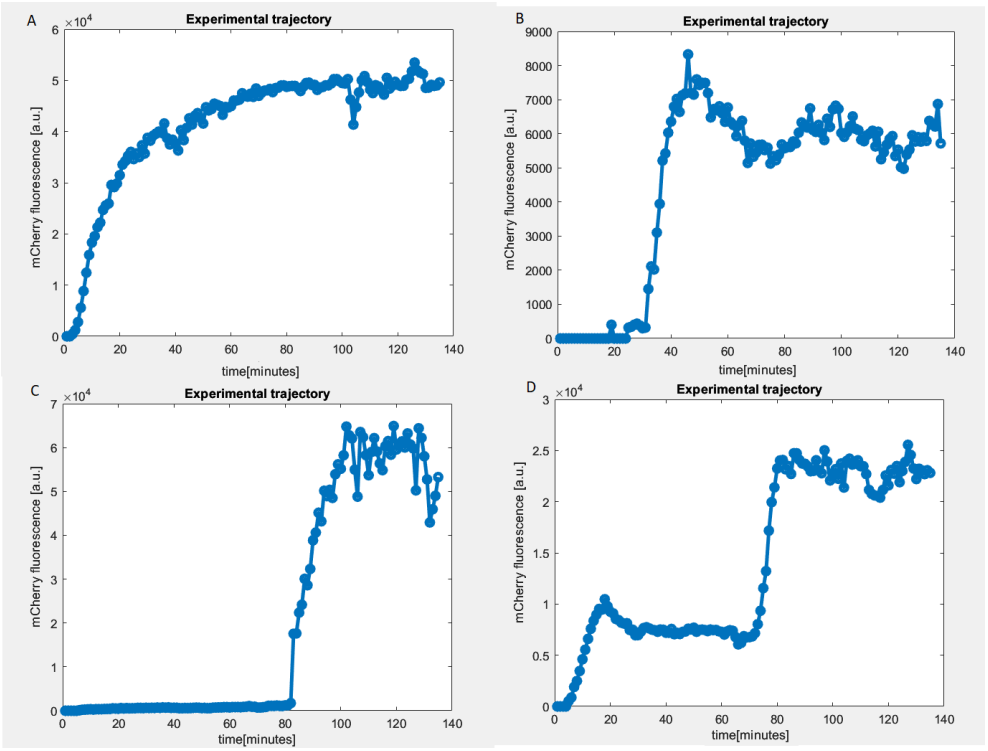

Figure 11: Characteristic trajectories taken from the experimental dataset. The trajectories selected are the most significative and the model we simulated well reproduces some of them. In particular in panel (A) we show a trajectory that is thougth to belong to the "above the threshold" regime. The protein synthesis starts very soon in time so it could be possible that a large amount of the transfected construct is entered into the cell. The trajectory belongs to the dataset with mCherry with N=1, so that we can argue that either protein translation is mildly regulated or the pool of miRNA is very small. Trajectory in panel (B) shows an overshoot. We can also observe that protein fluorescence begins to rise with a little delay, a scenario partially compatible with the presence of a pool of miRNA. In panel (C) the delay in fluorescence rising is even more evident than in (B), suggesting that the pool of miRNA keeps mCherry concentration very low until a threshold is crossed. Panel (D) is important for futher studies and new models to be implemented. Indeed, we can notice a step-wise growth in protein expression, suggesting that some rates are changing over time. We argue that this scenario is compatible with a competition mechanisms, where miRNAs and ribosomes compete for binding to the same mRNA.

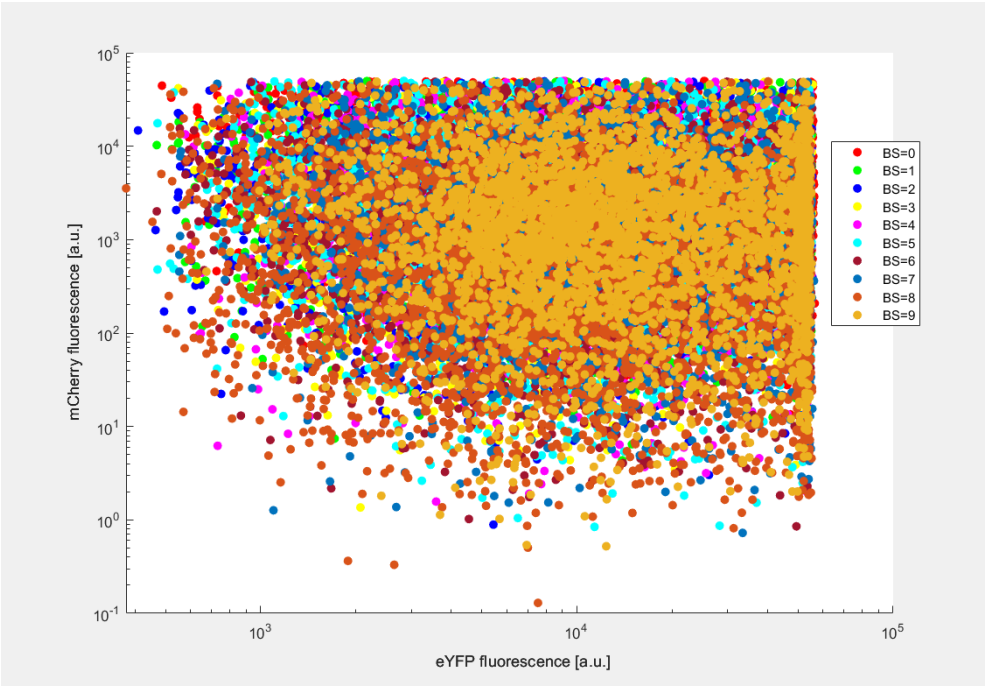

Figure 12: Scatter plot of the concentration of proteins at the steady state. On the x-axis there is the fluorescence of eYFP, the unregulated protein, on the y-axis the fluorescence of mCherry. As one can notice, the expression of mCherry decreases as the number of binding sites increases. Each point corresponds to a steady state, and each color corresponds to a different number N of binding sites.

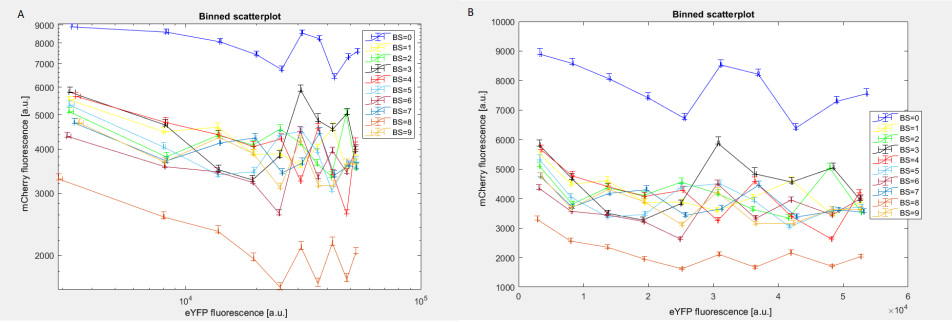

Figure 13: Steady states for the eYFP have been divided in 10 bins and the average value of the steady states contained in a bin has been evaluated together with its standard error. (A) Scatterplot with mean eYFP values on the x axis, and mean mCherry values on the y axis. (B) The same as in (A) in linear scale.

Datapoints were then grouped in 10 bins over eYFP fluorescence and then both eYFP and mCherry were mediated over these bins. Figure 12 shows the result of this binning. However, it seems that the varibility of mCherry upon varying eYFP is negligible.

6 Results

As a final point of the work, have been calculated the first passage time for both the two models and for experimental data. In models, the first passage time is the first time at which the protein's concentration cross the threshold α starting at time t=0. For what concerns experimental data, first passage time is again the first time at which proteins's expression crosses the threshold starting from the first time at which the protein's luminescence is different from zero. In this case indeed the transfetted mRNA begin to be translated at random times, since cells can also divide in two daughter cells in which the construct can also enter. So after cell division the target mRNA start to be translated only after a certain time. For the models have been performed simulation for which has been calculated the first passage time. Then, with the values of first passage time calculated one can calculate the coefficient of variation defined as the ratio between the standard deviation of the FPT and its mean value. The CVt has been calculated for 20 different and equispaced values of the threshold α, from 0 to 1. For each value of α have been performed 100000 simulations. For the first model the coefficient of variation has been performed for all the regimes introduced in its section . The obtained results are depicted in figure 14 and 15, the result obtained for the experimental data are the ones in figure 16

Another calculation performed in both cases is the average first passage time as a function of the threshold α for both the model and experimental data. here the experimental and theoretical results are in disagreement. The simulation's results are depicted in figure 17, the experimental counterpart is depicted in figure 18.

As a final point it was calculated the mean first passage time as a function of the binding sites on the mCherry's mRNA. The trajectories were binned in three different bins, depending on their steady state values in fluorescence. The curve shows a maximum for small values of binding sites. Also this phenomenon has not yet a theoretical interpretation. This is depicted in figure 19.

Results obtained suggests that the model exploited in this work is still uncomplete. Despite this, several interesting results were obtained, which could be useful in understanding how to improve the model in further works. An idea could be the introduction in the model of a new chemical species, ribosomes, that compete with miRNA for the binding of target mRNA. In this model protein can be translated only if ribosomes and mRNA are bound. Moreover, another interesting result is the fact that if miRNA is produced simultaneously with its target, the translation of target take shorter times with respect to its unregulated counterpart. In this case, timing fluctuations

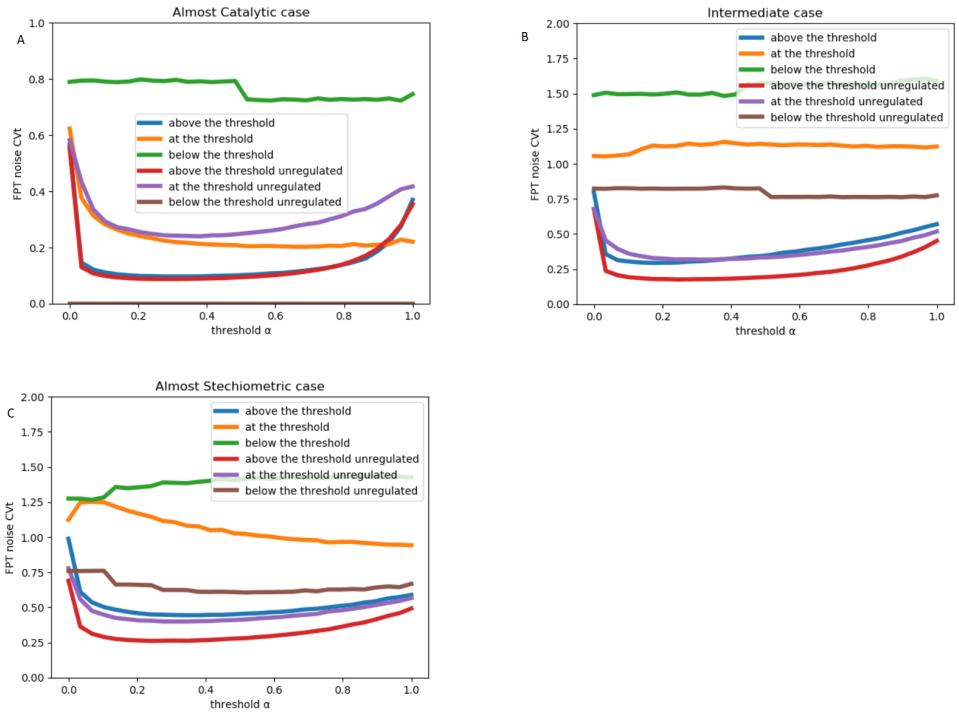

Figure 14: In this figure are depicted the results obtained for the coefficient of variation as a function of the threshold α. Figure A is in the almost catalytic case, in which one has an high recycle of miRNA. The result that we wanted to obtain, i.e. a minimum in the coefficient of variation as a function of threshold. There are numerical issues for the below the threshold regime. A possible explaination of this, is the fact that the p_{ss} in those cases is approximatively 0, since in this case the protein is highly repressed. So the FPT in this case is just a random number. In figure B the minimum of the CVt as a function of α has been obtained only in the above the threshold behaviour. This happens because the noise increases as the γ_5 rate increases. Indeed the CVt at the threshold is very high since it is higher than 1. It means that the noise in data is too high. The same happens in image C. The simulations are very noisy and the outcome are good only in the above the threshold case. What we can conclude from those plot is the fact that the a minimum in the CVt is always present in the above the threshold case. Stochastic simulations are too noisy in the intermediate case and the almost stechiometric case, highlighting the fact that a low recycle of miRNA result in an great increase of noise. Moreover the below the threshold case is out of our interest since it has not an experimental counterpart since the protein is almost not expressed.

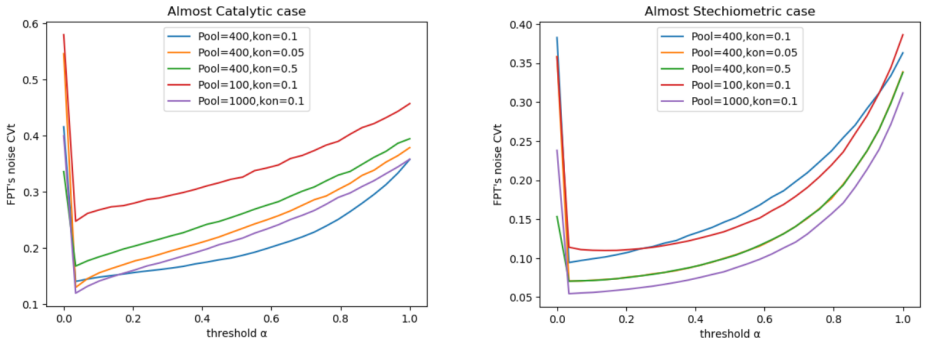

Figure 15: Coefficient of variation as a function of *alpha* for the pool of miRNA model. As one can observe this case is less noisy than the previous model. Moreover the minimum in noise is for very low values of the threshold α. This means that protein can be produced in very high precision in timing but only in small amounts.

grow. On the other hand, when a pool of miRNA is already present in cellular environment, the gene expression takes longer times to reach the threshold level, but the timing noise is reduced. So there is an interplay between the velocity and the fluctuations in timing in which a certain level of protein expression is reached.

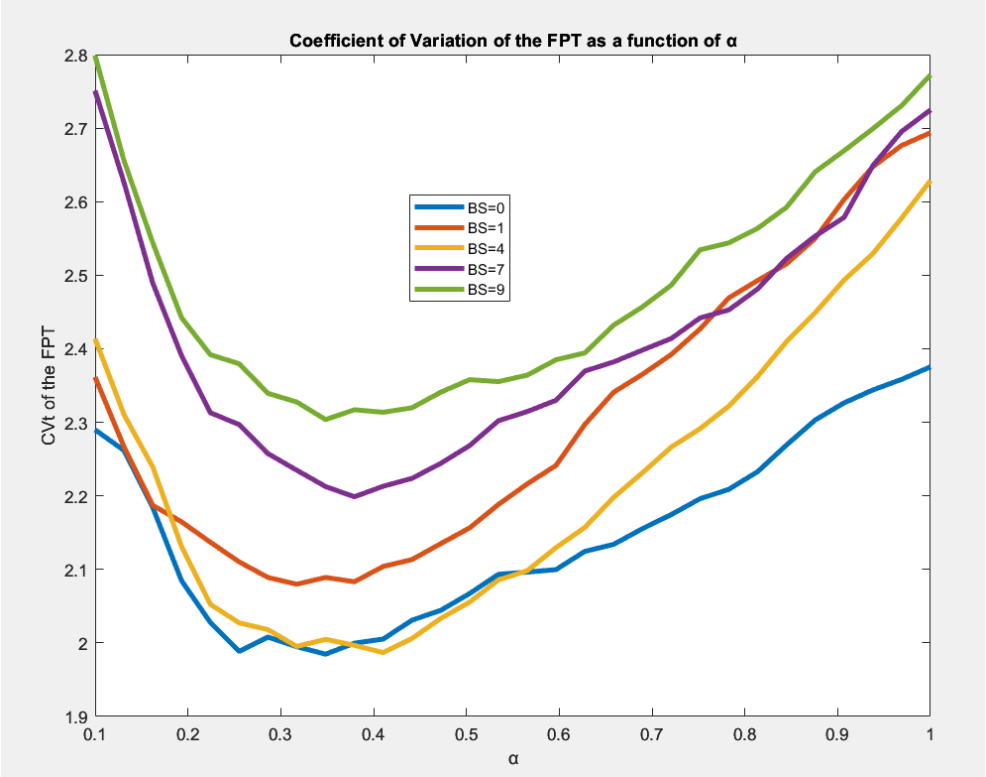

Figure 16: In the picture is depicted the coefficient of variation as a function of α for five different values of binding sites. As one can observe even in this case is evident the presence of a minimum in the coefficient of variation. As one can observe it seems that there is not any particular reltion between the number of binding sites and the value of the threshold at which is present the minimum. Moreover it doesn't seem to exist a relation between noise(CVt) and the number of binding sites at fixed α since the curves crosses between themselves

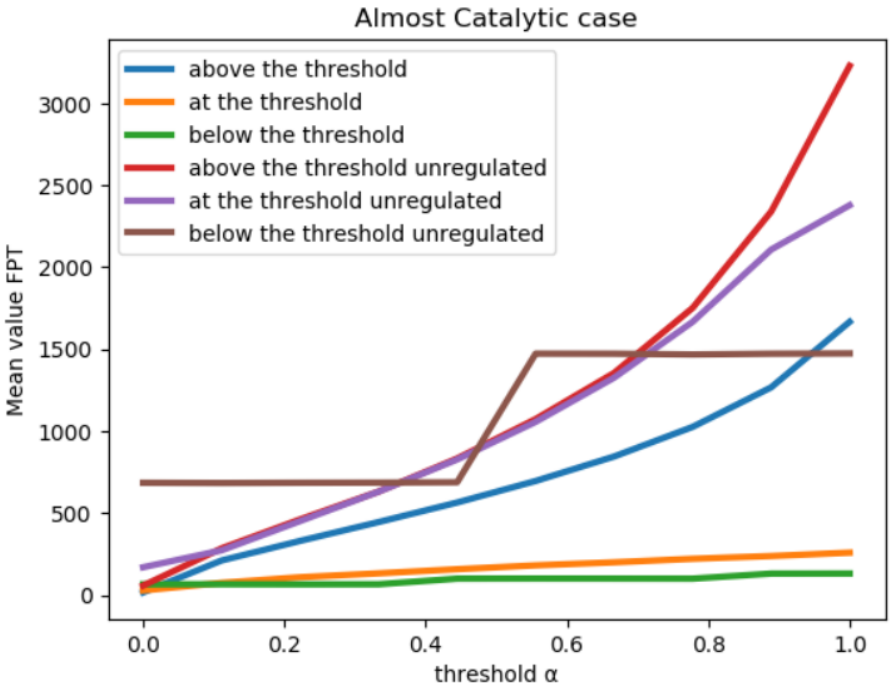

Figure 17: In the picture is depicted the average first passage time as a function of α in the theoretical case. All the regimes depending on the threshold were inspected. The average first passage time increases as α increase. so the function is monotone increasing with respect to α

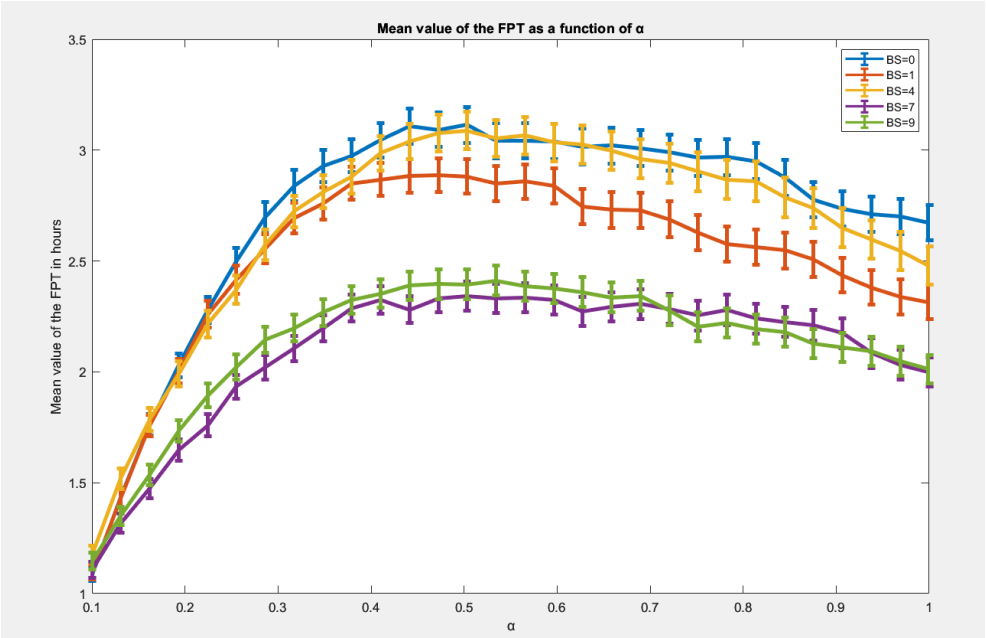

Figure 18: In the picture is depicted the average first passage time as a function of α in the experimental case. The curves correspond to 4 different values of binding sites present on the target miRNA. In this case the function seem to be a sigmoid, so it seems that there is a saturation phenomenon in the mean FPT as a function of *alpha*. No justification for this phenomenon has been found. Our hypothesis is that this phenomenon depends on the sensitivity of the instrument used to collect the experimental data.

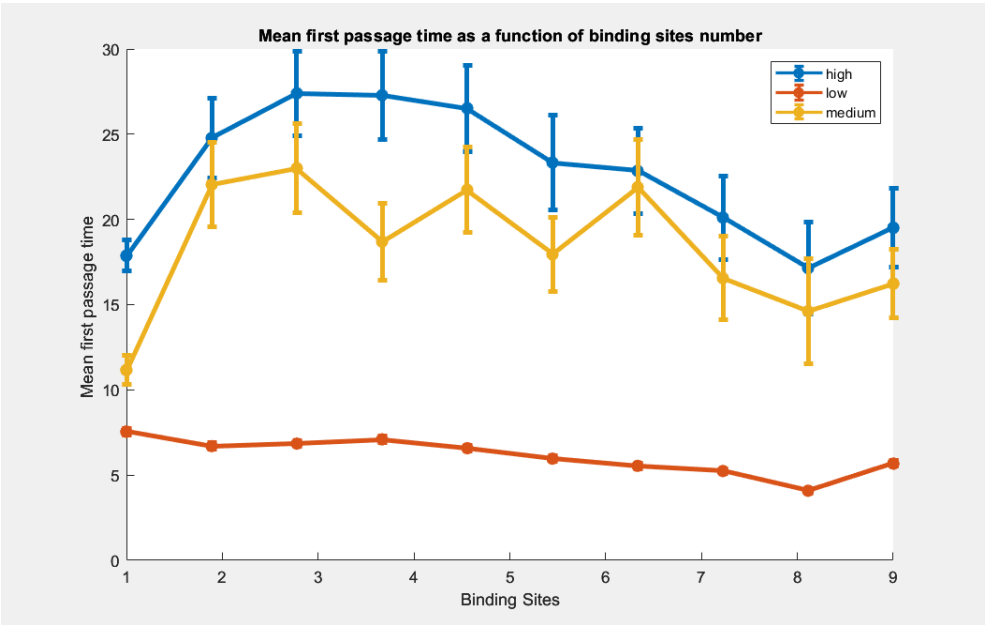

Figure 19: In the picture is depicted the average first passage time as a function of the number of binding sites on mRNA in the experimental case. The mCherry steady state were divaded in 3 different bins. In arbitrary units of fluorescence, "high" bin contains steady state with fluorescence higher than 40000. The "medium" bin correspond to values which go from 25000 to 40000. The "low" bin correspond to steady state values lower than 25000. The function shows a maximum for lower values of binding sites. This behaviour has not yet any theoretical justification.

References

[1] Bruce Alberts, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts, Peter Walter (2017),p.429-432,1180-1182 *Molecular biology of the cell*, CRC Press.

[2] Alma Dal Co,Marco Cosentino Lagomarsino,Michele Caselle,Matteo Osella (2017) *Stochastic timing in gene expression for simple regulatory stategies*, Nucleic Acids Research,Vol.45, N.3, p.1069-1078.

[3] Vahid Shaherezaei, Peter S.Swain (2008) *Analytical distributions for stochastic gene expression*, PNAS, Vol.105, N.45, p.17256-17261.

[4] Khem Raj Ghusinga, John J. Dennehy and Abhyudai Singh (2017) *First passage time approach to controlling noise in the timing of intracellular events*, PNAS, Vol.114, N.4, p.693-698.

[5] Shankar Mukherji,Margaret S. Ebert,Grace X Y Zheng,John S.Tsang, Phillip A.Sharp,Alexander van Ouedenaarden (2011) *MicroRNA can generate threshold in target gene expression*, Nature Genetics, Vol.43, N.9, p.854-860.

[6] Marco Del Giudice,Stefano Bo,Silvia Grigolon,Carla Bosia (2018) *On the role of extrinsic noise in microRNA-mediated bimodal gene expression*,PLOS Computational Biology.

[7] Jacob O'Brien,Heyam Hayder,Yara Zayed,Chun Peng (2018) *Overview of MicroRNA Biogenesis,Mechanism of Actions, and Circulation*, Frontiers in Endocrinology, Vol.9, N.402.

[8] Brian Munsky,Gregor Neuert,Alexander van Oudenaarden (2013) *Using Gene Expression Noise to Understand Gene Regulation*, National Institutes of Health, Vol.336, N.6078, p.183-187.

[9] Nicholas F. Polizzi, Michael J. Therien, and David N. Beratan (2016) *Mean First-Passage Times in Biology*, PubMed Central, Vol.56, N.9-10, p.816-824.

[10] Francesca Magri, Fiammetta Vanoli, Stefania Corti(2018) *miRNA in spinal muscular atrophy pathogenesis and therapy*, Journal of Cellular and Molecular Medicine, Vol.22, N.2, p.755-767.

[11] Witold Filipowicz, Suvendra N. Bhattacharyya Nahum Sonenberg(2008) *Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?*, Nature Reviews Genetics, Vol.9, p.102–114 .

[12] Lyudmila F. Gulyaeva, and Nicolay E. Kushlinskiy(2016) *Regulatory mechanisms of microRNA expression*, Journal of Translational Medicine, Vol.14, N.143 .

[13] Brian C. Schanen, Xiaoman Li(2011) *Transcriptional regulation of mammalian miRNA genes*, Genomics, Vol.97, N.1, p.1-6.

[14] Caterina Catalanotto, Carlo Cogoni, and Giuseppe Zardo(2016) *MicroRNA in Control of Gene Expression: An Overview of Nuclear Functions*,International Journal of Molecular Sciences, Vol.17, N.1712.

[15] Gregor Obernosterer, Philipp J.F. Leushner, Mattias Alenius, and Javier Martinez(2006) *Post-transcriptional regulation of microRNA expression*,RNA Society, Vol.12, p.1161-1167.

[16] Danlel T. Gillespie(1976) *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions* , The Journal of Physical Chemistry, Vol.22, p.403-434.

[17] Danlel T. Gillespie(1977) *Exact Stochastic Simulation of Coupled Chemical Reactions* , The Journal of Physical Chemistry, Vol. 81, N.25, p.2340–2361.

[18] Doob, Jacob L. (1942) *Topics in the Theory of Markoff Chains* , Transactions of the American Mathematical Society, Vol. 52, N.1, p.37–64.

[19] Doob, Jacob L. (1945) *Markoff chains – Denumerable case* , Transactions of the American Mathematical Society, Vol. 58, N.3, p.455–473.

[20] Cai L., Friedman N., Xie X.S. (2006) *Stochastic protein expression in individual cells at the single molecule level* , Nature, Vol.440, p.358–362.

[21] Raj A., van Oudenaarden A. (2008) *Nature, nurture, or chance: Stochastic gene expression and its consequences* , Cell, Vol.135, p.216–226.

[22] Yurkovsky E., Nachman I. (2013) *Event timing at the single-cell level* , Cell, Vol.12, p.90-98.

[23] Van Kampen N. (2011) *Stochastic Processes in Physics and Chemistry* , Elsevier,Amsterdam.

[24] Redner S. (2001) *A Guide to First-Passage Processes*, Cambridge University Press, Cambridge,UK.

[25] McAdams H.H., Arkin A. (1997) *Stochastic mechanisms in gene expression*,Proc Natl Acad Sci USA, Vol.94, p. 814–819.

[26] Amir, Kobiler, Rokney , Oppenheim and Stavans.(2007) *Noise in timing and precision of genev activities in a genetic cascade*,Mol. Syst. Biol., Vol.3, N. 71.

[27] Amir,A., Kobiler,O., Rokney,A., Oppenheim,A.B. and Stavans,J.(2007) *Noise in timing and precision of genev activities in a genetic cascade*,Mol. Syst. Biol., Vol.3, N. 71.

[28] Paulsson,J. (2005) *Model of stochastic gene expression*,Phys. Life Rev., Vol.2, p. 157-175.

[29] Friedman,N., Cai,L. and Xie,X.S. (2006) *Linking stochastic dynamics to population distribution: an analytical framework of gene expression*,Phys. Rev. Lett., Vol.97.

[30] Carla Bosia, Andrea Pagnani, Riccardo Zecchina (2013) *Modelling Competing Endogenous RNA Networks*,PLOS ONE.