

POLITECNICO DI TORINO

Master Degree in Physics of Complex Systems



MASTER's Degree Thesis

**Extrinsic vs. Intrinsic noise-induced
bimodality in post transcriptional
regulation experiments**

Supervisor

Dr. Carla BOSIA

Candidate

Valeria FAMÀ

JULY 2021

To my father
To my grandmother Angela

Summary

MicroRNAs (miRNAs) are a class of nearly 22 nucleotide-long, non-coding RNA molecules involved in the post-transcriptional regulation of their target RNAs. MiRNAs and their targets interact via a titration-like mechanism characterized by threshold effects and cross-talk among targets. Threshold has to be understood as a value of mRNA transcription rate such that above this value many target molecules are available for translation while below it mRNA are bound to miRNA, rapidly degraded and cannot be translated. It has been discovered that different RNA species may compete for miRNAs binding with this inducing indirect interactions amongst the miRNA targets which reciprocally influence their expression levels. Moreover, also the fluctuations on their levels of expression are coupled through miRNAs. It is of particular interest that an increase in noise leads to an increase in cell variability and this may lead to bimodal cell population distributions with high and low expression states of specific miRNA targets. This mechanism is particularly relevant near the threshold, where stochastic fluctuations play an important role for the cell fate. It is worth noting that bimodal expression of genes may lead to very different phenotypes, where the modes of the distribution underlies healthy or sick cells or two different stages of differentiation. An open question is whether these bimodal distributions at the target level are due to extrinsic fluctuations on miRNA pools or to intrinsic fluctuations on miRNA-target interaction strength. My thesis aims to answer this question combining both experiments and theoretical modelling. The experimental part consists of transfection experiments of bidirectional plasmids in epithelial human cells (HEK293). The plasmids, circular DNA molecules, code for two fluorescent proteins, a yellow fluorophore named eYFP and a red one named mCherry. The sequence coding for mCherry contains a varying number of miRNA binding sites, while the eYFP sequence is left unchanged. mCherry and eYFP are thus respectively proxies for target expression and its constitutive activity. Thanks to this genetic device it is possible not only to distinguish between cells hosting plasmids or not but also to track gene expression in presence or in absence of miRNA regulation by measuring cell fluorescence. To understand if bimodal distributions are due to extrinsic noise in the miRNA pool or to intrinsic noise in the miRNA-target interaction strength we need to distinguish between two

scenarios: (i) a single cell scenario, in which bimodality is given by single cells near the threshold that jump from one state to the other over time (this scenario is due to intrinsic noise on the miRNA-target interaction strength); (ii) a population scenario, for which bimodality is given by cells whose target may be in one state or the other because of different miRNAs basal values (this is a situation of extrinsic noise in the miRNA pool). These two situations can be distinguished by measuring transition times between the two states when observing time trajectories of the target. Cells are therefore observed over time with time-lapse microscopy experiments that allow to record their fluorescence. Acquired images should be analysed to detect every single cell, track it over time and obtain temporal trajectories of its fluorescence. It can be expected that single cell transition times are distributed according to $\exp(-kt)$: therefore we can estimate k for every time trajectory and, depending on the distribution of k found, be able to understand which scenario underlies the measurements.

Acknowledgements

I would like to thank my supervisor Dr. Carla Bosia who have guided me throughout this project and sparked my interest in physics and in research despite the difficulties of this year. I would also like to acknowledge Chiara Enrico Bena, Elsi Ferro and Simone Tealdi for the collaboration and the insightful discussions about the project. In addition, I would like to thank my mother for giving me the possibility to study in another city - I know how it cost her. Thanks to my brother for the support and advices and for beign a firm reference in my life.

Thanks to Patrizia, Alessandro, Clelia and Jacopo for welcoming me in their home in my diffucult times.

A big aknowledgment to Matteo, without whom this journey would have not been the same, for his loving support in good and bad times. Last but not least, I would like to thank Jessica and Donato for their true friendship, precious gift and not granted.

Table of Contents

List of Tables	VIII
List of Figures	IX
1 Introduction	1
1.1 At the interface between physics and biology	1
1.2 MiRNA biogenesis	2
1.3 MiRNA importance and functions	3
1.4 Threshold effects	4
1.5 Intrinsic and extrinsic noise effects in gene regulation	6
2 Experimental work	8
3 Models and methods	12
3.1 Intrinsic noise-induced bimodality	12
3.1.1 Bimodal population distributions	14
3.1.2 Single-cell trajectory	19
3.1.3 Permanence times	23
3.2 Extrinsic noise-induced bimodality	38
3.2.1 Permanence times	41
3.3 Experimental data	46
3.3.1 Bimodal distributions	46
3.3.2 Permanence times	49
4 Discussion and conclusions	55
Bibliography	57

List of Tables

3.1	Values of parameters used to generate reference bimodal population distribution. Gillespie algorithm works with number of molecules and propensity rates. [11]	15
3.2	Values of the parameters estimated from the fit of bimodal population distributions with relative errors. The fit has not been performed for transcription rate $k = 104s^{-1}$ and $k = 124s^{-1}$ in that these are not bimodal population distribution so we cannot compute permanence times in the two states for them.	18
3.3	Values of the parameters estimated from the fit of bimodal population distributions with relative errors. The fit has not been performed for a miRNA-target interaction strength $g = 7 \cdot 10^{-4}s^{-1}$ because its population distribution is not bimodal.	18
3.4	Values of decay rates, i.e. the inverse of permanence times of cell population distribution in the repressed (k_1) and expressed (k_2) states, for different target transcription rates.	35
3.5	Values of decay rates, i.e. the inverse of permanence times of cell population distribution in the repressed (k_1) and expressed (k_2) states, for different values of miRNA-mRNA interaction strength.	36

List of Figures

1.1	Scheme of the possible biogenesis processes, adapted from [13]. Biogenesis pathways are distinguished in canonical and non canonical. The difference between the two is that while canonical pathway undergoes two maturation steps (performed by Drosha and Dicer), non canonical biogenesis pathways bypass certain steps [14].	3
1.2	Figure adapted from [8]. Under threshold there are more free miRNAs than targets. Over threshold, miRNAs are almost all bound in complexes and the number of free targets is much grater than that of free miRNAs. At the threshold the number of free targets is comparable to that of free miRNAs.	5
1.3	(a) Two color fluorescent reporter system that co-regulates eYFP, yellow fluorescent protein, and mCherry,red fluorescent protein, which are proxies for transcription activity in absence and in presence of regulation. (b) Sample fluorescence microscopy data. (c) Trend of mCherry as a function of eYFP for N=0,1 miRNA binding sites showing the appearance of a non-linear behaviour when increasing the number of miRNA binding sites.	6
1.4	Intrinsic and extrinsic noise-induced bimodal population distributions. Figures adapted from [8].	7
2.1	Bimodal mCherry distribution obtained from experimental data with 7 miRNA binding sites.	10
2.2	Schematic diagram of code for cells tracking.	11
3.1	Scheme of the reactions involved in the model for intrinsic noise. Gere r codes for target mRNA R while gene s codes for miRNA S . A fraction α of the latter interact with R with an interaction strength g . Both R and S can be degraded at rate g_R and g_S . Gene R can be translated at rate k_P into protein P , which can be degraded at rate g_P . Figure adapted from [8].	13

3.2	Protein trend at the steady state as a function of the transcription rate k_R . Under the threshold, target mRNA are bound in complexes with miRNAs and the concentration of free miRNA is higher than that of mRNAs (we will say that the system is repressed). Over the threshold there are more free mRNAs than miRNAs (the system is expressed). At the threshold there is equimolarity between miRNAs and mRNAs: fluctuations are crucial to determine the behaviour of the system.	15
3.3	Reference bimodal population distribution obtained in case of pure intrinsic noise with parameters in 3.1. The black line is the fitting obtained with expression 3.3	17
3.4	Probability distribution functions for target transcription rate k_R under threshold.	18
3.5	Different bimodal distribution for increasing transcription rate k_R . The higher target transcription rate is, the higher the number of free targets in the distribution and the more important the expressed state becomes.	19
3.6	Different bimodal population distribution for increasing miRNA-target interaction strength g . When this interaction strength is low, as in Figure 3.6b, the bimodal distribution can't be seen. Increasing the coupling between miRNA and target the distribution becomes bimodal and the expressed state begins to flatten.	20
3.7	Result of Kolmogorov-Smirnov statistical test performed on bimodal population distribution and single cell bimodal distribution. With 95% of confidence the null hypothesis, entailing that the compared datasets have been extracted from the same distribution, is verified.	21
3.8	Comparison between bimodal population distribution and bimodal single cell distribution. Performing a Kolmogorov-Smirnov statistical test it has been proved that a single cell trajectory summarize the behaviour of the entire population.	22
3.9	Trajectory of a single cell and its auto-correlation function.	22
3.10	First permanence times of the system with $k_R = 0.1095s^{-1}$ for repressed and expressed state in case of pure intrinsic noise with states defined as $\mu_i \pm b_i \cdot \sigma_i$	24
3.11	All permanence times of the system with $k_R = 0.1095s^{-1}$ for repressed and expressed state in case of pure intrinsic noise with states defined as $\mu_i \pm b_i \cdot \sigma_i$	25
3.12	First permanence times of the system with $k_R = 0.1095s^{-1}$ for repressed and expressed state in case of pure intrinsic noise with states defined as $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	25

3.13	All permanence times of the system with $k_R = 0.1095s^{-1}$ for repressed and expressed state in case of pure intrinsic noise with states defined as $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	26
3.14	First permanence times of the system with $k_R = 0.107s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	26
3.15	All permanence times of the system with $k_R = 0.107s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	27
3.16	First permanence times of the system with $k_R = 0.112s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	27
3.17	All permanence times of the system with $k_R = 0.112s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	28
3.18	First permanence times of the system with $k_R = 0.115s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	28
3.19	All permanence times of the system with $k_R = 0.115s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	29
3.20	First permanence times of the system with $g = 3.5 * 10^{-3}s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	29
3.21	All permanence times of the system with $g = 3.5 * 10^{-3}s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	30
3.22	First permanence times of the system with $g = 3.5 * 10^{-2}s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	30
3.23	All permanence times of the system with $g = 3.5 * 10^{-2}s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	31
3.24	First permanence times of the system with $g = 7 * 10^{-2}s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	32
3.25	All permanence times of the system with $g = 7 * 10^{-2}s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	32
3.26	Caption	33

3.27	Permanence times of cell population with transcription rate $k_R = 0.107s^{-1}$ in repressed and expressed states with relative fit. Time range at which we look at has been restricted to 100s.	33
3.28	Permanence times of cell population with transcription rate $k_R = 0.1095s^{-1}$ in repressed and expressed states with relative fit. Time range at which we look at has been restricted to 100s.	34
3.29	Permanence times of cell population with transcription rate $k_R = 0.112s^{-1}$ in repressed and expressed states with relative fit. Time range at which we look at has been restricted to 100s.	34
3.30	Permanence times of cell population with transcription rate $k_R = 0.115s^{-1}$ in repressed and expressed states with relative fit. Time range at which we look at has been restricted to 100s.	35
3.31	Permanence times of cell population with transcription rate $g = 0.0035s^{-1}$ in repressed and expressed states with relative fit. Time range at which we look at has been restricted to 100s.	36
3.32	Permanence times of cell population with transcription rate $g = 0.035s^{-1}$ in repressed and expressed states with relative fit. Time range at which we look at has been restricted to 100s.	37
3.33	Permanence times of cell population with transcription rate $g = 0.07s^{-1}$ in repressed and expressed states with relative fit. Time range at which we look at has been restricted to 100s.	37
3.34	Model of the system in presence of extrinsic noise. Gene r is transcribed at constant rate k_R into target mRNA R , which can be translated into protein P at rate k_P or be degraded with rate g_R or still can interact with a fraction α of microRNA S . This latter is transcribed by gene s at a fluctuating transcription rate, extracted from a Gaussian distribution. MiRNA can then be degraded at rate g_S and protein can be degraded as well at rate g_P [8]. Figure adapted from [8].	38
3.35	Extrinsic noise-induced bimodal population distribution generated with the same parameter of Table 3.1. Extrinsic noise has been taken into account by allowing miRNA transcription rate k_S fluctuations from cell to cell. The value of k_S has been drawn by a Gaussian distribution with mean $\mu = 0.085s^{-1}$ and standard deviation $\sigma = 5 \cdot 10^{-3}s^{-1}$	39
3.36	Comparison between two extrinsic noise-induced bimodal population distributions. They have been generated with the parameters in Table 3.1 but the standard deviations σ_i ($i=1,2$) of the distribution from which miRNA transcription rate has been drawn are different. Increasing σ corresponds to increasing the extrinsic noise.	40
3.37	Extrinsic noise can induce bimodality even when intrinsic noise cannot.	40

3.38	Moving miRNA transcription rate k_S allows to draw values in a large interval. On the left, threshold's location is represented for the highest possible k_S extracted, that is mean value plus a standard deviation. On the right, threshold's location is represented for the lowest possible k_S extracted, that is mean value minus a standard deviation.	41
3.39	First permanence times of the system in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	43
3.40	All permanence times of the system in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	44
3.41	First permanence times of the system in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	45
3.42	All permanence times of the system in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$	46
3.43	Scatter plots of mean mCherry values versus mean eYFP values for varying number of binding sites. In going from 0 to 7 binding sites, miRNA-mRNA interaction strength is increasing resulting in a non-linearity.	48
3.44	Figure showing the non linear behavior of mCherry as a function of eYFP for increasing number of binding sites.	49
3.45	mCherry distributions under, over and at the threshold for the cases of 0,1,4 and 7 binding sites. The higher the number of binding sites, the stronger the interaction strength between miRNA and target. A bimodal distribution can be seen in 7 binding sites picture.	50
3.46	Fit of 7 binding sites bimodal distribution using the combination of two Gaussian distributions.	51
3.47	mCherry probability distribution functions for the cases of 4 (green distribution) and 7 (orange distribution) binding sites. Vertical lines represent the end and the beginning of repressed and expressed states respectively.	52
3.48	Cells trajectories in presence of 7 miRNA binding sites. The dotted lines represent the two means extracted from the fit of the bimodal distribution in Figure 3.46	53
3.49	Cells permanence times in case of 4 miRNA binding sites. Both repressed and expressed states show peaks on long times.	53
3.50	Cells permanence times in case of 7 miRNA binding sites. Both repressed and expressed states show peaks on long times.	54

Chapter 1

Introduction

1.1 At the interface between physics and biology

This thesis work stands at the interface between physics and biology. One may wonder why a physicist should be interested in biological processes. Schrodinger cared about this issue in his book "What is life?" by saying that "We have inherited from our forefathers the keen longing for unified, all-embracing knowledge" [1].

In this work, we would like to explore physical properties of a biological system by combining theoretical knowledge and experiments. The focus of this thesis will be on microRNAs (miRNAs), a class of small non-coding RNAs, that exerts regulatory functions in both mammals and plants: miRNAs can induce translation repression or degradation on their target genes [2]. Being gene expression stochastic [3], we will make use of a probabilistic framework to study the behavior of a system composed by microRNAs, their target mRNAs and proteins. In particular, we will focus on protein expression. It is known that gene expression may vary across different individuals subject to different environmental conditions but it may also vary in a population of identical individuals subject to the same environment [4] [5]. We will refer to "extrinsic noise" in the first case and to "intrinsic noise" in the second one. Variability in gene expression is crucial in determining phenotypic diversity [6] and miRNAs play an important role in this phenomenon [7]. In previous studies it has been found out that miRNAs could lead to a bimodal expression of their target genes [8], associated to phenotypic diversity in a cell population that is genetically identical [9]. Moreover, miRNAs anomalous expression is associated with many diseases ranging from cancer to central nervous system diseases [10].

In this thesis work, we want to distinguish the contribution of intrinsic and extrinsic noise in generating protein bimodal distributions by looking at protein permanence times in the two modes of their expression distribution. For this purpose, Gillespie simulations [11] of the systems 3.1 and 3.34 will be performed. In the following,

we will talk about miRNA biogenesis and functions. Next, we will explain how miRNAs regulate protein expression via a threshold mechanism. Effects of different types of noise on gene regulation will be taken into account by means of two different models for microRNA-driven inhibition. The experimental set-up will be described, together with analysis performed on experimental data. Population permanence times in the two modes of bimodal expression distribution will be studied. For the case of intrinsic noise, we will introduce the concept of double potential well to justify the distribution of population permanence times in the two modes.

1.2 MiRNA biogenesis

MicroRNA (miRNAs) are single-stranded, non-coding RNA molecules with an average length of 22 nucleotides. They have been discovered in 1993 first in *Caenorhabditis Elegans*, a nematode, and some years later in various species of plants and animals[12]. MiRNA biogenesis begins by means of RNA polymerase II and III, which transcribe nucleotide sequences coding for miRNAs and later bind a localized promoter in proximity of this sequence. The possible biogenesis pathways can be classified in canonical and non-canonical.

The canonical pathway is the dominant way in which miRNAs are processed: they are transcribed from DNA sequences into primary miRNAs (pri-miRNAs) and processed into precursor miRNA (pre-miRNAs) by means of the microprocessor complex, made of DiGeorge Syndrome Critical Region 8 (DGCR8), a protein which recognizes precise motifs in pri-miRNAs, and Drosha, an enzyme which cleaves them producing a characteristic stem-loop structure of about 70 base pairs long, called pre-miRNAs. Once these latter have been generated, they are exported from the nucleus to the cytoplasm by Exportin 5 and processed by Dicer, a RNase III which removes the terminal loop forming a mature miRNA duplex. Mature miRNAs strands have a certain directionality which determines their name: if a strand derive from the 5' end of the pre-miRNA hairpin then it is called *5p*, otherwise if it derives from the 3' end it is called *3p*. Depending on their stability, either both strands can be loaded into the Argonaute family of proteins or only the one with lower 5' stability is loaded, taking the name of guide strand. The other strand, called passenger, will be either cleaved by AGO2 and degraded by cellular machinery (if it does not contain mismatches) or passively unwound and degraded in the other case. Non-canonical pathways make use of different combinations of previous proteins and they can be classified into Drosha/DCGR8-independent and Dicer-independent. The first group of pre-miRNAs is directly exported to the cytoplasm by mean of exportin 1 and does not need Drosha cleavage [13].

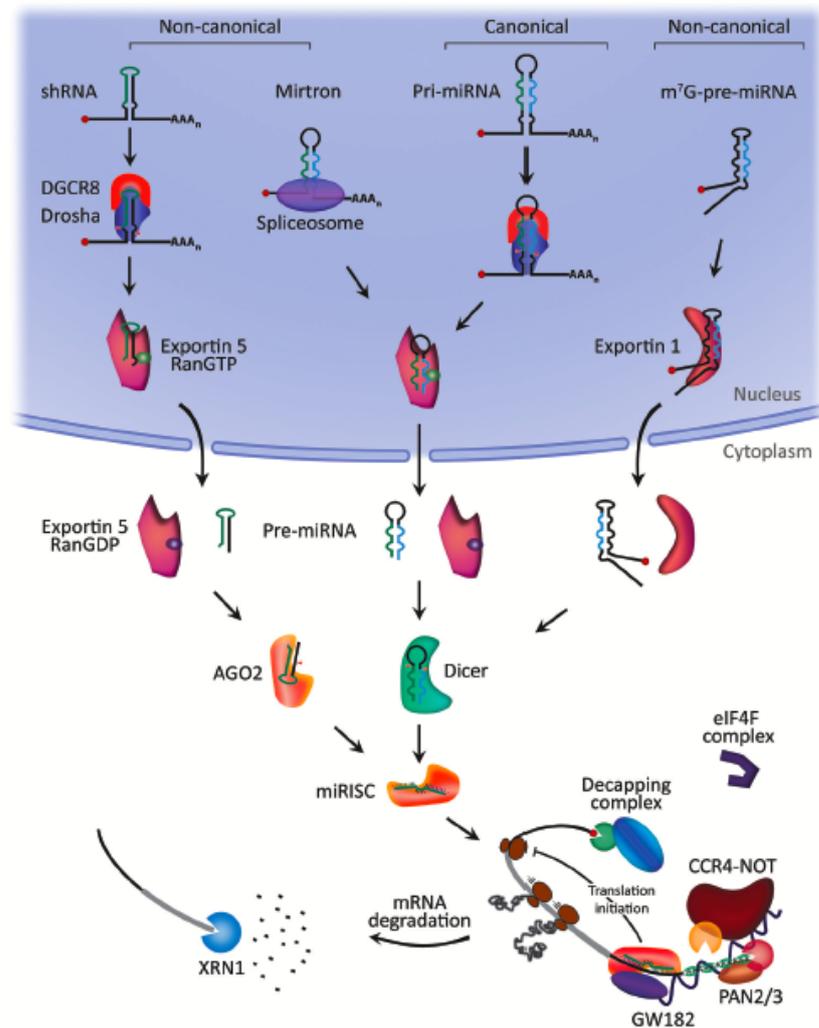


Figure 1.1: Scheme of the possible biogenesis processes, adapted from [13]. Biogenesis pathways are distinguished in canonical and non canonical. The difference between the two is that while canonical pathway undergoes two maturation steps (performed by Drosha and Dicer), non canonical biogenesis pathways bypass certain steps [14].

1.3 MiRNA importance and functions

Numerous studies have pointed out the crucial role of miRNAs in both biological and pathological processes. The latter are generated by changes in miRNA expression [10]. Thus, it is essential to know the different functions miRNAs can exert depending on their binding site. The most important ones are:

1. MiRNA-mediated gene silencing via minimal miRNA-induced silencing complex (miRISC), formed by the guide strand loaded into AGO, which is target specific and binds complementary sequences on target mRNA, called miRNA response elements (MRE). Depending on the degree of complementarity, there are two possible scenarios:
 - the complementarity between miRNA and mRNA is perfect leading to a degradation of target mRNA via a cleavage mechanism process, which is said to be slicer dependent [13] ;
 - there are mismatches facilitating miRISC-mediated translational inhibition and target mRNA decay [13].
2. MiRNA-mediated translational activation: it could arise from binding of miRNAs with 5' UTR of mRNAs and leads to an up-regulation of gene expression [13].
3. MiRNA-mediated transcriptional and post-transcriptional gene regulation within the nucleus: some studies have reported that low molecular weight miRISC can interact with mRNAs within the nucleus and induce nuclear mRNA degradation but they can also perform a direct regulation [13].

Some studies have proved that gene regulation mediated by microRNAs is a dynamic and robust process in helping to buffer gene expression to a steady state [13]. The focus of this work will be on gene silencing by mRNA target degradation.

1.4 Threshold effects

Having established the importance of gene regulation by miRNAs, a question arises spontaneously: how does a target's gene protein expression change in presence and in absence of miRNA regulation? Mukherji et al. developed a model [15] to show that under a certain threshold level of target mRNA, protein level is strongly repressed. Threshold has to be intended as a value of target transcription rate under which the concentration of free miRNA is much greater than target's one. We will call repressed state the one associated to miRNA's abundance with respect to targets. Over the threshold, instead, the concentration of free target is high and miRNA are almost all bound in complexes with their targets. When the system is in this situation, we will say that it is in the expressed state. At the threshold there is a situation of equimolarity which makes stochastic fluctuations of crucial importance (See Figure 1.2). Mukherji et al. model is based on protein-protein titration reactions [16], where titration is defined as "the process of determining the quantity of a substance A by adding measured increments of substance B, the titrant, with which it reacts until exact chemical equivalence is achieved (the

equivalence point) (Worsfold,2005) [17]. It describes the concentration of free target mRNA under the regulation of miRNA as a function of non-regulated transcriptional activity, under the assumption that proteins can be produced solely by the target [15]. The crucial point is the formation of a complex miRNA-target. From an experimental point of view, a construct allowing for monitoring of the system in presence and in absence of regulation has been obtained with a bidirectional promoter driving two genes that code for two fluorescent proteins, named mCherry and eYFP. mCherry is a red fluorescent protein that accounts for regulation and has its 3'UTR engineered to contain a certain number $N = 0,1,4,7$ of binding sites, while eYFP accounts for unregulated transcriptional activity. The model strongly depends on two parameters: g , that is the rate at which miRNA binds target, and λ , which gives a measure of the sharpness of the threshold and it inversely proportional to g [15]. According to this model, the threshold can be sharpened by increasing g that is equivalent to adding miRNA binding sites to the system [15]. Increasing miRNA binding sites entails that the probability of a miRNA binding a target is higher so a higher level of targets are necessary to obtain the expression we would have in absence of regulation. The increasing in the sharpness of the threshold is greater in going from 1 to 4 binding sites than in going from 4 to 7 binding sites [15].

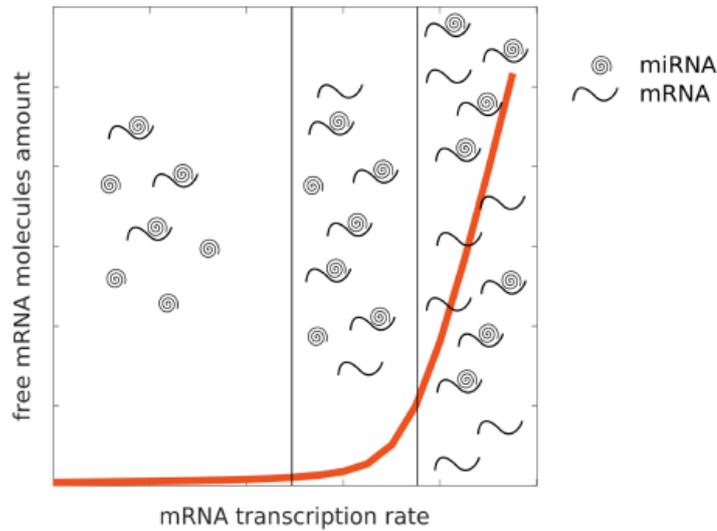


Figure 1.2: Figure adapted from [8]. Under threshold there are more free miRNAs than targets. Over threshold, miRNAs are almost all bound in complexes and the number of free targets is much greater than that of free miRNAs. At the threshold the number of free targets is comparable to that of free miRNAs.

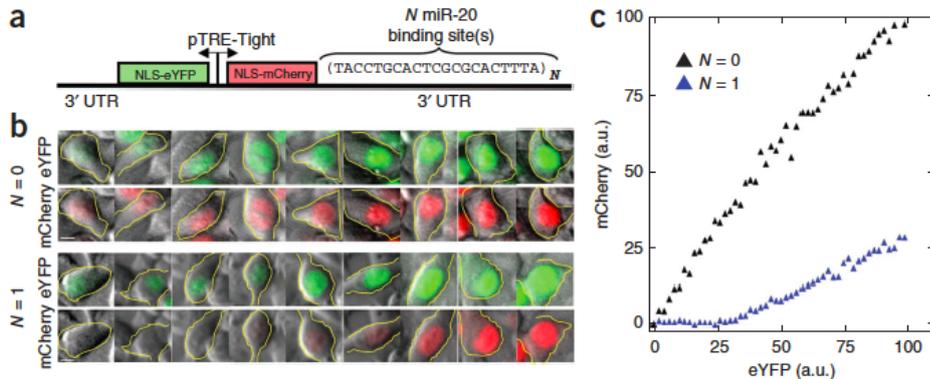


Figure 1.3: (a) Two color fluorescent reporter system that co-regulates eYFP, yellow fluorescent protein, and mCherry, red fluorescent protein, which are proxies for transcription activity in absence and in presence of regulation. (b) Sample fluorescence microscopy data. (c) Trend of mCherry as a function of eYFP for $N=0,1$ miRNA binding sites showing the appearance of a non-linear behaviour when increasing the number of miRNA binding sites.

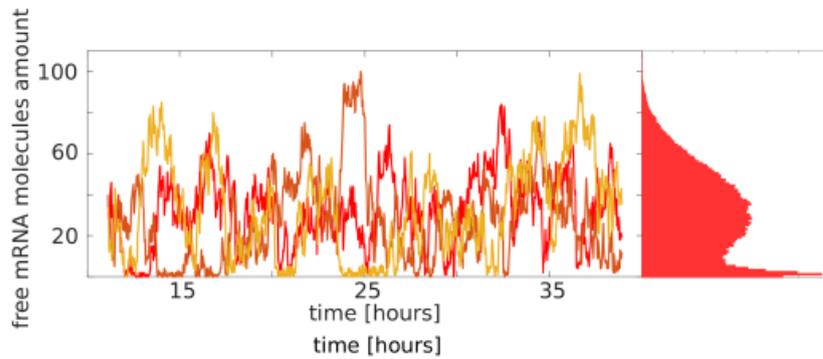
1.5 Intrinsic and extrinsic noise effects in gene regulation

In this section we want to investigate the possible sources of noise that can act on our system and their effects on it. Stochasticity is widespread in nature [18] and it is present as well in gene expression [3]. It is known that a population of identical cells exposed to the same environmental conditions can present phenotypic differences or a different level of expression from cell to cell [19]. The type of stochasticity related to biochemical reactions, which we will call "intrinsic", is strictly related to the size of the system: the larger the size of the system, the lower the fluctuations are [19] [20]. A measure for noise is given by the coefficient of variation, defined as the standard deviation to mean ratio for a certain quantity x_i :

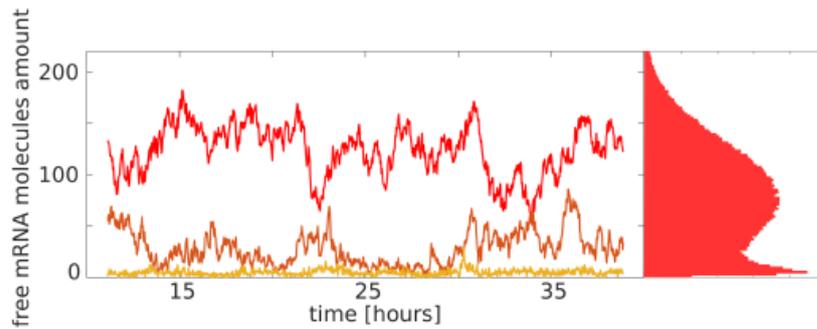
$$CV_{x_i} = \frac{\sigma_{\langle x_i \rangle}}{\langle x_i \rangle} \quad (1.1)$$

Anyway, this is not the only possible kind of noise that can affect a system. A variability in gene expression may also be induced by fluctuations on transcription rate or, more in general, by factors that are external to biochemical reactions [19]. We will refer to this type of noise as "extrinsic" noise. In the following, we will introduce two models 3.1 and 3.34 accounting for intrinsic and extrinsic noise. The model for intrinsic noise consists of a gene r that it's transcribed (in mRNA R)

and translated (into protein P) and can interact with the transcription product S of another gene s . Gene s codes for microRNA S . MicroRNA, target mRNA and protein can be degraded. In this model, the unique source of noise will come from the stochasticity of biochemical reactions and it will be taken into account by means of stochastic simulations. The model for extrinsic noise 3.34 is essentially the same, with the exception that extrinsic noise is introduced as a fluctuation on miRNA transcription rate. This stochasticity will be added to the one coming from biochemical reactions. Despite the models are similar, the results are profoundly different as we can see in Figures 1.4a and 1.4b. Intrinsic noise-induced bimodal distribution is obtained at single cell level while extrinsic noise-induced bimodal distribution is obtained at population level.



(a) Bimodal population distribution of free mRNA molecules amount given by intrinsic noise: single cells sample the two modes of the distribution.



(b) Bimodal population distribution of free mRNA molecules amount given by extrinsic noise: some cells never change state.

Figure 1.4: Intrinsic and extrinsic noise-induced bimodal population distributions. Figures adapted from [8].

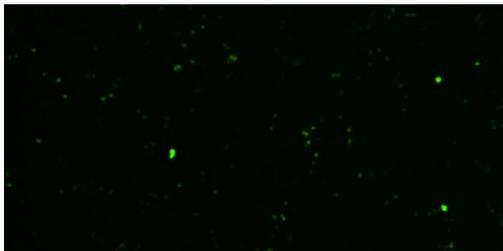
Chapter 2

Experimental work

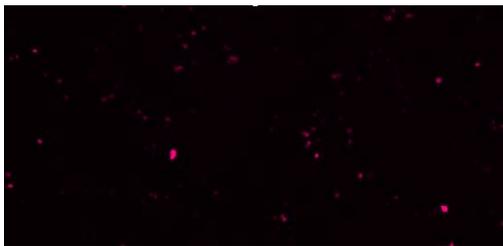
Having introduced the general problem, let us focus on the physical understanding of the system under study. The aim of the experiments is to perform single-cell tracking under different conditions of miRNA-mRNA interaction strength to see how single cells behave in time.

I took advantage of BIOMed Lab, the new interdepartmental laboratory belonging to Politecnico di Torino, to perform transfection experiments of bidirectional plasmids in epithelial human cells (HEK293). A plasmid is a circular DNA molecule separated from the chromosomal DNA that can duplicate independently from it. Although they are more common in bacteria, they can also be found in archaea and in eukaryotes [21]. They are characterized by the presence of certain genes that make them resistant to specific antibiotics [21]. In the experiment here described, tetracycline-inducible plasmids coding for two fluorescent proteins, a yellow fluorophore named eYFP and a red one named mCherry, have been used. The 3' untranslated (UTR) region of mCherry has been engineered to contain a varying number of miRNA binding sites ($N = 0,1,4,7$), while the eYFP sequence is left unchanged. mCherry and eYFP are thus respectively proxies for target expression and its constitutive activity. Thanks to this genetic device it is possible not only to distinguish between cells hosting plasmids or not but also to track gene expression in presence or in absence of miRNA regulation by measuring cell fluorescence [15]. The process by means of which these plasmids are inserted into cells is called "transfection" and it has been performed using effectene as a transfection agent. Cells have been filmed 24h after transfection, for $\approx 70h$, with the aim of a fluorescence microscope. Taking advantage of Nikon Lipsi platform [22], screenshots have been taken every twenty minutes.

As a result of these experiments, data have been extracted from time-lapse and organized in various files each one accounting for a different part of filmed cells plate and a fixed number of binding sites. Data analysis has been performed using Image Processing package belonging to Matlab in order to segment and track cells



(a) Screenshot on eYFP channel of transfected cells seen under fluorescent microscopy.



(b) Screenshot on mCherry channel of transfected cells seen under fluorescent microscopy.

and extract important information such as the fluorescence level.

As a first thing, levels of eYFP and mCherry have been measured at fixed time in order to reproduce threshold behavior [15] with the increasing in the number of binding sites. Sorting cells according to an increasing level of eYFP, it is found that for $N=0$ binding sites there is a linear relationship between mCherry and eYFP levels (see Figure 3.5d). For $N = 1, 4, 7$ binding sites, instead, a non-linearity (threshold) in mCherry level appears and it varies with time (see Figure 3.5d). Remembering previous discussion, it is expected that in proximity of the threshold there is a bimodal population distribution for mCherry expression level so a small range of eYFP has been isolated near the threshold and the distribution of mCherry levels corresponding to that values of eYFP has been plot (see Figure 2.1).

Having shown that these bimodal distributions exist (see Figure 2.1), it is now interesting to look at the trajectories of cells present in them. In order to do it, a single cell tracking has been performed. The idea behind the code is the following: once cells are segmented and labelled, each cells is "followed" over time. As cells are expected not to deviate too much from their position passing from one frame to another, one could assume that at time t a particular cell will be in the neighborhood of its position at time $t - 1$. Looking in this neighboring region, the cell is identified as the nearest to the oldest position and take the same label of the cell in the previous frame if this is still available (it has not been assigned to another

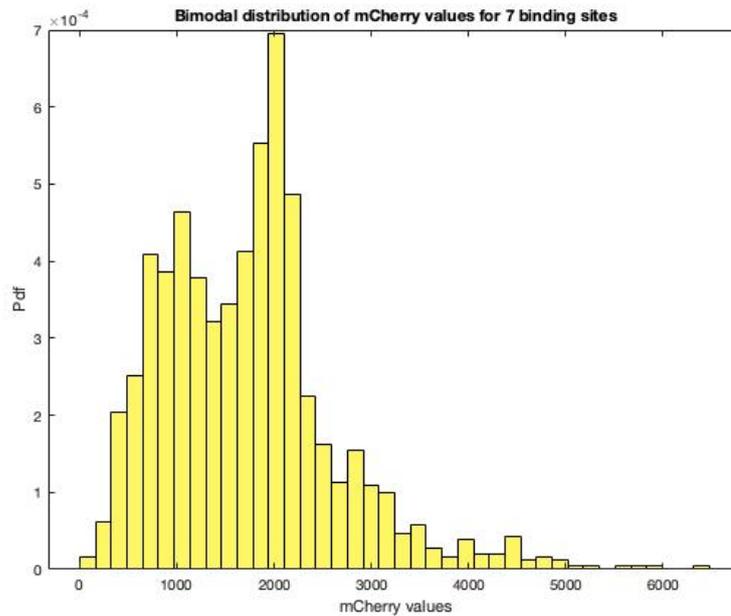


Figure 2.1: Bimodal mCherry distribution obtained from experimental data with 7 miRNA binding sites.

cell). If the label has already been assigned and the cell is in the neighborhood of a cell of the previous frame, a new label is assigned to the new cell. This situation may appear when a cell in the previous frame has divided giving rise to two cells. If we don't find a match with a cell in the previous frame, the possible scenarios are two: the first is that the cell is new, meaning that it is the first time we see its fluorescent signal. The second is that the cell has already been seen in previous frames and then it has disappeared for some frames (e.g. due to an out-of-focus): in this case we assign to it the old label. In this way, the trajectory of every single cell is reconstructed over time.

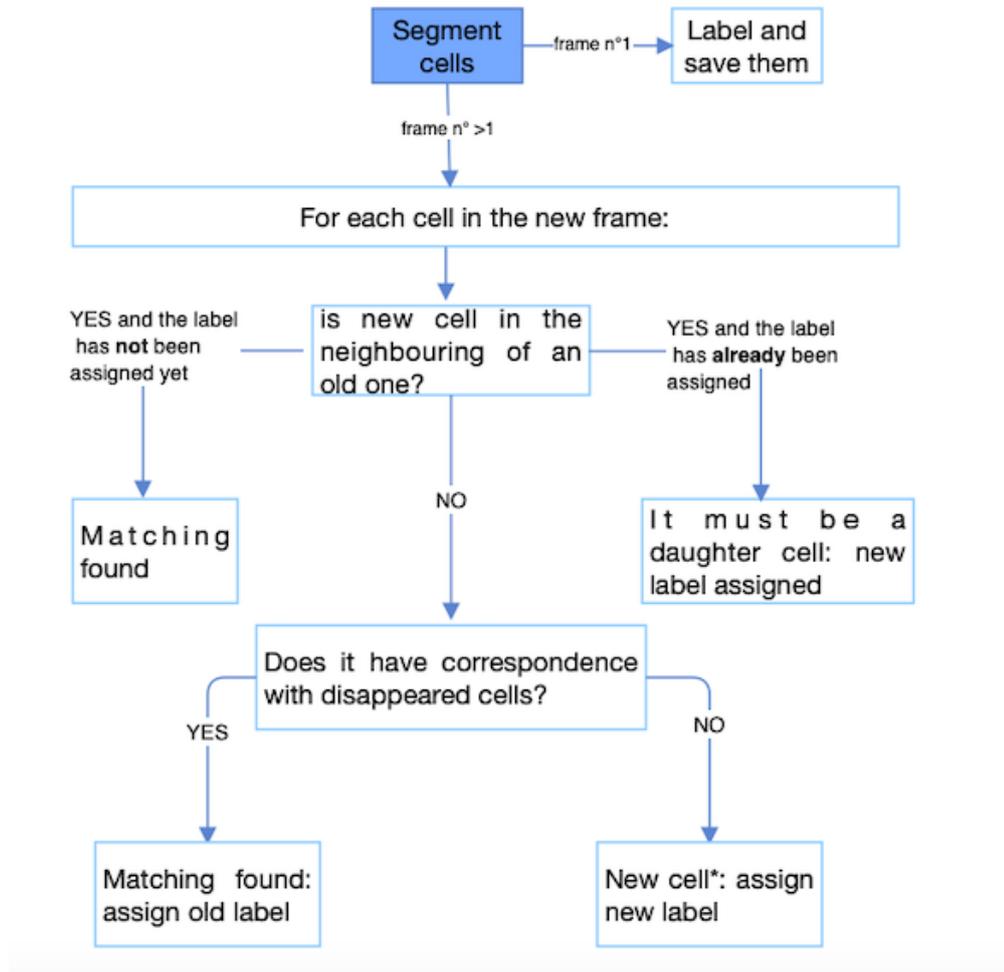


Figure 2.2: Schematic diagram of code for cells tracking.

Chapter 3

Models and methods

In this chapter, we will study the impact of intrinsic and extrinsic noise on the system under study. First, to assess the case with pure intrinsic noise, changes in the shape of the bimodal population distribution will be investigated by varying miRNA-mRNA interaction strength and target transcription rate. Then, it will be shown that a single cell trajectory summarizes the entire population, meaning that the system is ergodic. Moreover, we will introduce the concept of double potential well to explain cells permanence time distribution in the two states of intrinsic noise-induced bimodal population distribution. Then, extrinsic noise will be added to the system and its changes on bimodal population distribution with respect to the intrinsic noise case will be investigated. Permanence times of the system in the two states of extrinsic noise-induced bimodal population distributions will be studied as well.

3.1 Intrinsic noise-induced bimodality

Let's consider two distinct genes, r and s , transcribed independently at constant transcription rates k_R and k_S . Gene r codes for target mRNA R while gene s codes for microRNA S . The first RNA is translated into protein at rate k_P and can interact with a fraction α of miRNA S with an interaction strength g . Both RNAs and proteins are then degraded with rates g_R , g_S and g_P .

This model can be described in terms of the following system of ordinary differential equations (ODEs):

$$\begin{cases} \frac{dR}{dt} = k_R - g_R R - gRS \\ \frac{dS}{dt} = k_S - g_S S - \alpha gRS \\ \frac{dP}{dt} = k_P R - g_P P \end{cases} \quad (3.1)$$

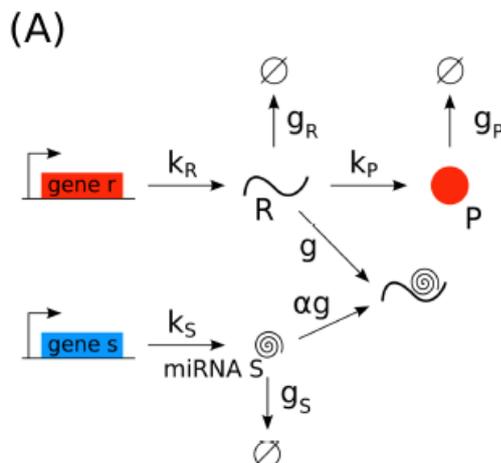


Figure 3.1: Scheme of the reactions involved in the model for intrinsic noise. Gene r codes for target mRNA R while gene s codes for miRNA S . A fraction α of the latter interact with R with an interaction strength g . Both R and S can be degraded at rate g_R and g_S . Gene R can be translated at rate k_P into protein P , which can be degraded at rate g_P . Figure adapted from [8].

The above mentioned system 3.1 must be solved to find the location of the threshold (see Figure 1.2). The system should be at the stationary state because we do not want the number of molecules to vary over long times.

However, gene expression is a stochastic process [3]: we said in previous sections that with the term "intrinsic" noise we will refer to the one induced by the stochasticity of biochemical reactions. In this framework, the variation of target and miRNA quantities leads to a bimodal distribution for target, where the two modes of the distribution account for the scenarios in which target is bound to miRNA or not [8]. Another source of noise can arise from environmental conditions: we will refer to this type of noise as "extrinsic" [8] and it will be introduced in our model as a fluctuation on miRNA transcription rate. From an experimental point of view, we are looking at cells that are in different phases of their cell cycle because they are not synchronized. It is known that miRNA changes value in a way related to its cell cycle [23] so it is as if we were looking at cells having different values of miRNA transcription rate.

In light of the stochasticity of gene expression, the most appropriate description for our system seems to be a master equation 3.2, which allows us to have information on fluctuations even in presence of few molecules. In order to write the master equation, let's write the probability distribution of observing n_R molecules of target mRNA, n_S molecules of miRNA and n_P molecules of protein at a certain time

t , given the set of parameters $\mathcal{W} = \{k_R, k_S, k_P, g, g_R, g_S, g_P, \alpha\}$. Recalling that concentration $\rho = \frac{n_{cell}}{V_{cell}}$, where V_{cell} is the typical volume of a cell, the master equation reads:

$$\begin{aligned}
 \frac{dP(\bar{n}, t)}{dt} = & k_R [P(n_R - 1, t) - P(\bar{n}, t)] + \\
 & \frac{g_R}{V_{cell}} [(n_R + 1)P(n_R + 1, t) - n_R P(\bar{n}, t)] + \\
 & k_S [P(n_S - 1, t) - P(\bar{n}, t)] + \\
 & \frac{k_P n_R}{V_{cell}} [P(n_P - 1, t) - P(\bar{n}, t)] + \\
 & \frac{g_P}{V_{cell}} [(n_P + 1)P(n_P + 1, t) - n_P P(\bar{n}, t)] + \\
 & \frac{g\alpha}{V_{cell}^2} [(n_S + 1)(n_R + 1)P(n_R + 1, n_S + 1, t) - n_S n_R P(\bar{n}, t)] + \\
 & \frac{g(1 - \alpha)n_S}{V_{cell}^2} [(n_R + 1)P(n_R + 1, t) - n_R P(\bar{n}, t)]
 \end{aligned} \tag{3.2}$$

where $\bar{n} = (n_R, n_S, n_P)$ [8]. Solving the master equation might be a complex task due to the non linear term that couples miRNA and target mRNA. A good alternative could be that of performing stochastic simulations until reactions reach a stationary state and search for bimodal distributions in the neighborhood of the threshold. In this work, Gillespie direct method algorithm has been employed to simulate the Markov processes described by the master equation 3.2. Given that \bar{n} is the vector of species counts, $a(\bar{n})$ are the propensity functions of an elementary reaction, t is the current time and τ is an infinitesimal time, direct method consists in the following steps:

1. Initialize the time $t = t_0$ and the system's state $\bar{n} = \bar{n}_0$
2. Evaluate the functions $a_j(\bar{n})$ and their sums $\sum a_j(\bar{n})$
3. Effect the next reaction by replacing $t \leftrightarrow t + \tau$ and $n \leftrightarrow n + \nu_j$
4. Record (\bar{n}, t)

3.1.1 Bimodal population distributions

The aim of this section is to investigate the changes in bimodal population distributions when the system moves away from the threshold or miRNA-target interaction strength is lowered or raised. This task is accomplished by running 100000 Gillespie simulations for a time $t_{steady} = 100000s \approx 28h$ and by taking protein values at $t = t_{steady}$.

Using parameters in table 3.1, the trend of proteins at the steady state as a function of target transcription rate shows a well-defined threshold (see Figure 3.2).

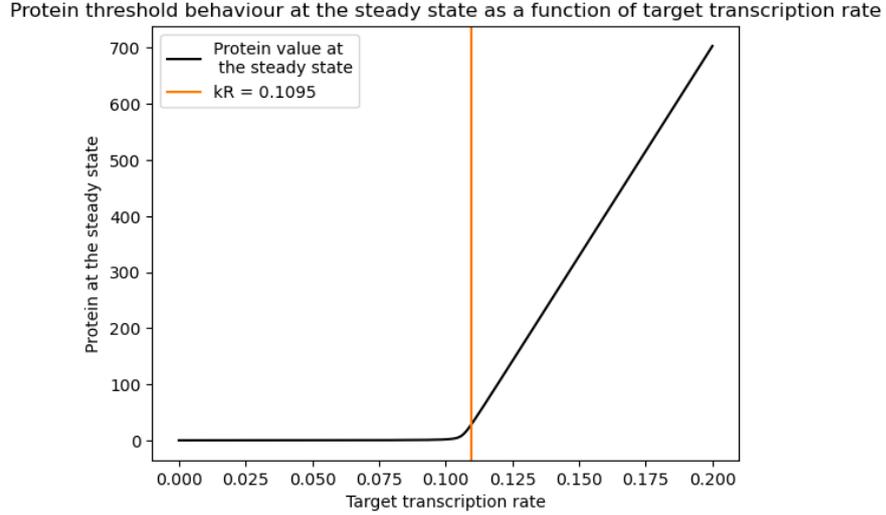


Figure 3.2: Protein trend at the steady state as a function of the transcription rate k_R . Under the threshold, target mRNA are bound in complexes with miRNAs and the concentration of free miRNA is higher than that of mRNAs (we will say that the system is repressed). Over the threshold there are more free mRNAs than miRNAs (the system is expressed). At the threshold there is equimolarity between miRNAs and mRNAs: fluctuations are crucial to determine the behaviour of the system.

Parameter	Value	Units	Description
k_r	0.1095	s^{-1}	Transcription rate of mRNA
k_s	0.085	s^{-1}	Transcription rate of miRNA
k_p	0.06	s^{-1}	Translation rate
g_r	$4 * 10^{-4}$	s^{-1}	Degradation rate of mRNA
g_s	$2 * 10^{-4}$	s^{-1}	Degradation rate of miRNA
g_p	$2 * 10^{-4}$	s^{-1}	Degradation rate of protein
g	$7 * 10^{-3}$	s^{-1}	miRNA-mRNA interaction strength
α	0.8	1	Fraction of recycled miRNA

Table 3.1: Values of parameters used to generate reference bimodal population distribution. Gillespie algorithm works with number of molecules and propensity rates. [11]

Reference bimodal population distribution in Figure 3.3 has been generated

with parameters in Table 3.1. To ensure that the distribution we have found is bimodal, a fit has been performed using the sum of an exponential and a Gaussian function with 5 parameters among which only 4 of them are independent. It is known that an unregulated gene is normally distributed, as in the case of high free target mRNA concentration. When free miRNA are much more than free targets, the distribution of this latter is exponential: the system behaves as in presence of an absorbing wall. For what concern the number of parameters, one may argue that 5 is a high number and may lead to overfitting but we are doing this fit just to extract the values of means and standard deviations, necessary to compute permanence times. The expression used for the fitting reads:

$$bim(x) = c_1 e^{-kx} + c_2 e^{-((x-\mu)^2)/2\sigma^2} \quad (3.3)$$

The parameter acting on the amplitude of the exponential function has been written in terms of other parameters by noticing that it must holds the following

$$\int_0^{\text{inf}} (c_1 e^{-kx} + c_2 e^{-((x-\mu)^2)/2\sigma^2}) dx = 1 \quad (3.4)$$

Computing the first term in the equation one gets:

$$\int_0^{\text{inf}} c_1 e^{-kx} dx = \frac{c_1}{k} \quad (3.5)$$

For the second term, it is useful to recall the following formula:

$$\int_0^{\text{inf}} e^{-x^2} dx = \frac{\sqrt{\pi}}{2} \text{erf}(z) \quad (3.6)$$

By changing variable one obtains

$$\int_0^{\text{inf}} e^{-((x-\mu)^2)/2\sigma^2} dx = \frac{\sqrt{\pi}}{2} \text{erf}\left(-\frac{\mu}{\sqrt{2}\sigma}\right) \quad (3.7)$$

and finally the value of c_1 is found:

$$c_1 = k \left[1 - c_2 \sigma \frac{\sqrt{\pi}}{2} \text{erf}\left(-\frac{\mu}{\sqrt{2}\sigma}\right) \right] \quad (3.8)$$

From the fitting, resulting parameters are $k = (0.21 \pm 0.03) \text{ of molecules}$, $c_2 = (0.0075 \pm 0.0002)$, $\mu_2 = (47 \pm 5) \text{ of molecules}$ and $\sigma_2 = (59 \pm 4) \text{ of molecules}$.

Now, we want to investigate the changes in bimodal distribution shape obtained by moving the system away from the threshold, both increasing and decreasing target transcription rate. From previous theoretical work we know that increasing target mRNA transcription rate causes a decreasing in the number of available

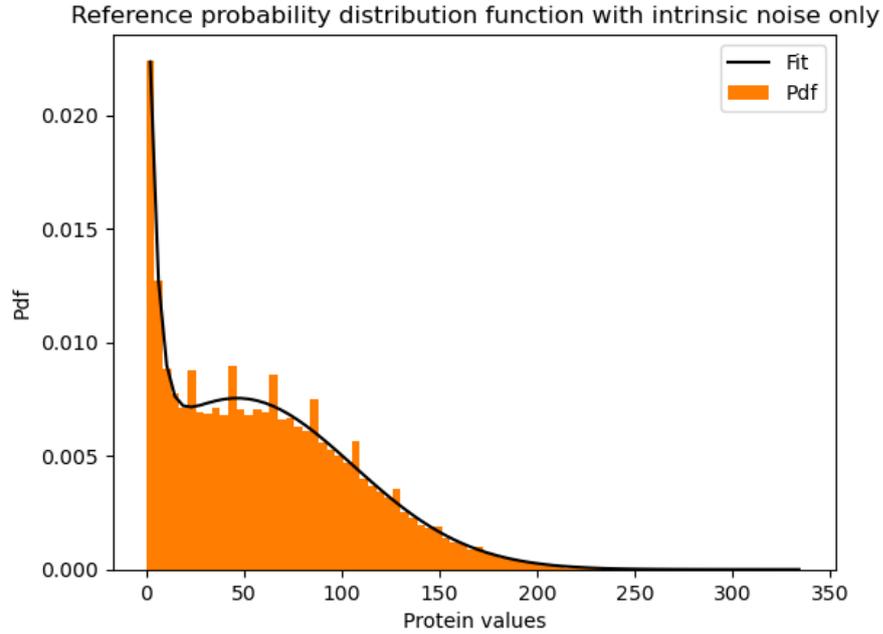


Figure 3.3: Reference bimodal population distribution obtained in case of pure intrinsic noise with parameters in 3.1. The black line is the fitting obtained with expression 3.3

miRNAs [15]: the expressed state will become more and more important leading to a Gaussian distribution in which the repressed state has disappeared. Decreasing k_R , instead, all targets are bounded by miRNAs and thus their expression is repressed [15]: the population distribution will be exponential because the first mode prevails. Figures 3.4 and 3.5 show various bimodal population distributions with different k_R .

An analogous discussion could be done for miRNA-target interaction strength. A bimodal distribution appears only in cases of high coupling between miRNA and target: this is a single-cell effect meaning that only cells with strong coupling can pass from a repressed state to an expressed one near the threshold. Moreover, the higher is the strength g , the wider is the interval of k_R for which a bimodal distribution exists [15]. In Figure 3.6, four different cases for increasing value of g have been shown. When miRNA-target interaction strength is very low as in Figure 3.20 protein values are Gaussian distributed. With the increasing of the strength g the distribution becomes bimodal and the height of the leftmost barrier increases.

k_R (s^{-1})	k (# of molecules)	c (number)	μ_2 (# of molecules)	σ_2 (# of molecules)
0.107	0.016 ± 0.01	0.0075 ± 0.0003	34 ± 8	54 ± 5
0.1095	0.22 ± 0.03	0.0075 ± 0.0002	47 ± 5	59 ± 4
0.112	0.11 ± 0.03	0.0073 ± 0.0002	72 ± 3	56 ± 3
0.115	0.16 ± 0.07	0.0072 ± 0.0001	95 ± 1	57 ± 1

Table 3.2: Values of the parameters estimated from the fit of bimodal population distributions with relative errors. The fit has not been performed for transcription rate $k = 104s^{-1}$ and $k = 124s^{-1}$ in that these are not bimodal population distribution so we cannot compute permanence times in the two states for them.

g (s^{-1})	k (# of molecules)	c (number)	μ_1 (# of molecules)	σ_2 (# of molecules)
$3.5 \cdot 10^{-3}$	0.021 ± 0.006	0.004 ± 0.001	79 ± 7	40 ± 9
$7 \cdot 10^{-3}$	0.22 ± 0.03	0.0075 ± 0.0002	47 ± 5	59 ± 4
$3.5 \cdot 10^{-2}$	0.56 ± 0.05	0.0071 ± 0.0002	52 ± 3	55 ± 2
$7 \cdot 10^{-2}$	0.6 ± 0.1	0.0070 ± 0.0003	55 ± 4	51 ± 3

Table 3.3: Values of the parameters estimated from the fit of bimodal population distributions with relative errors. The fit has not been performed for a miRNA-target interaction strength $g = 7 \cdot 10^{-4}s^{-1}$ because its population distribution is not bimodal.

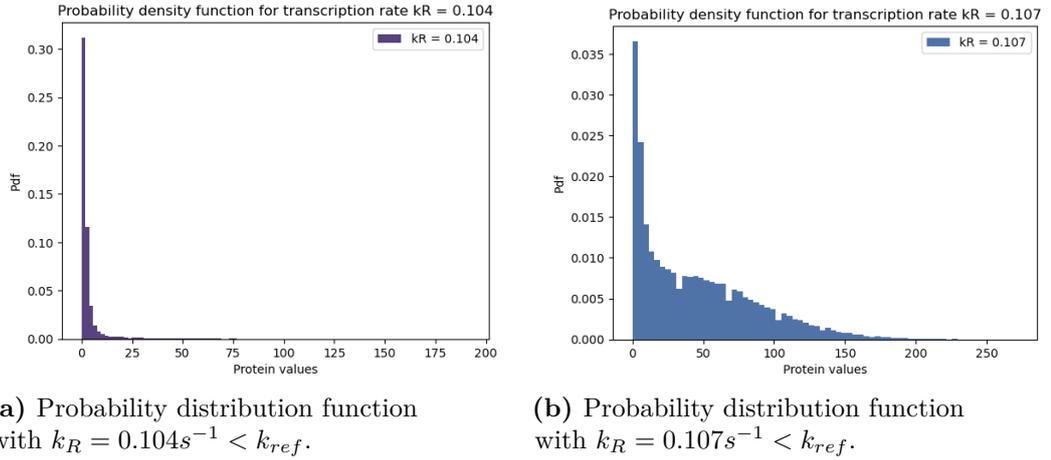


Figure 3.4: Probability distribution functions for target transcription rate k_R under threshold.

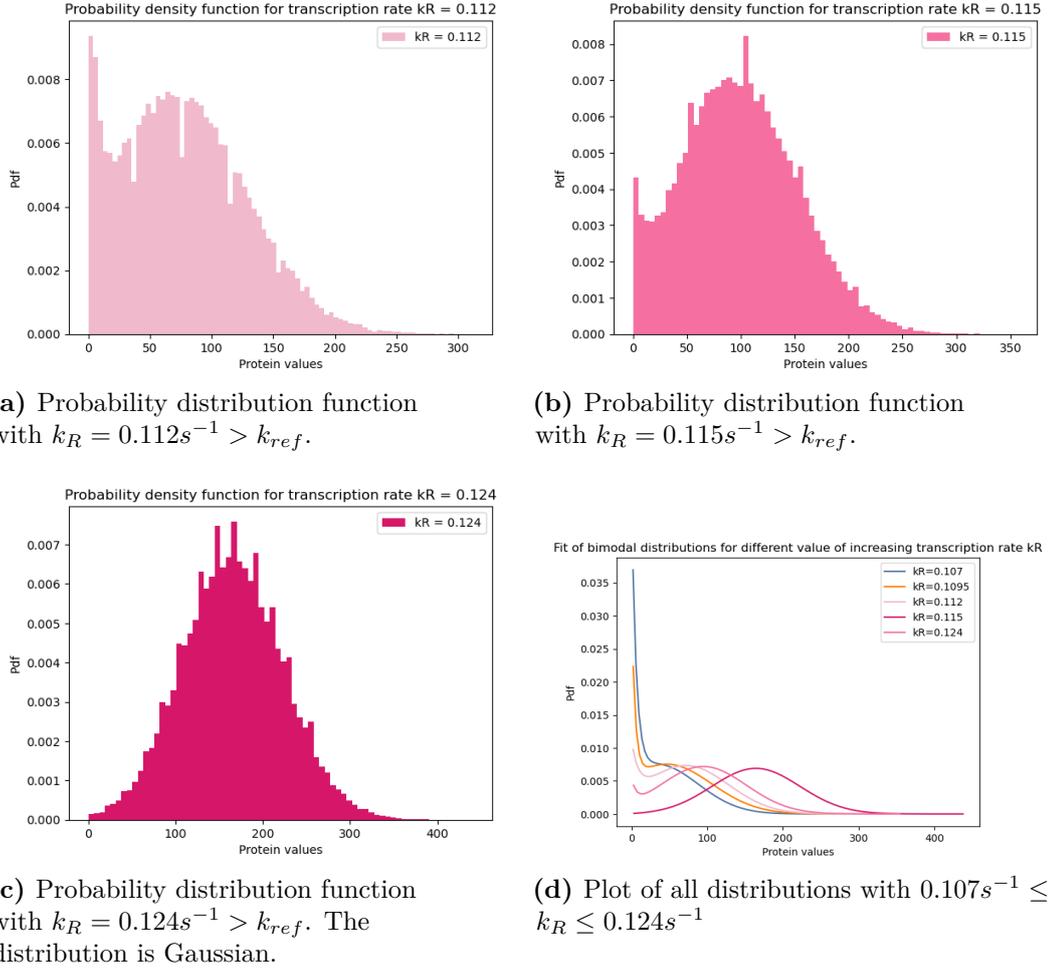
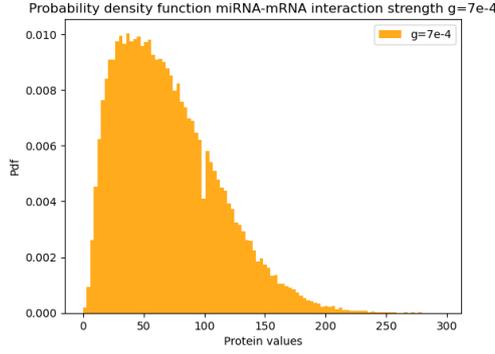


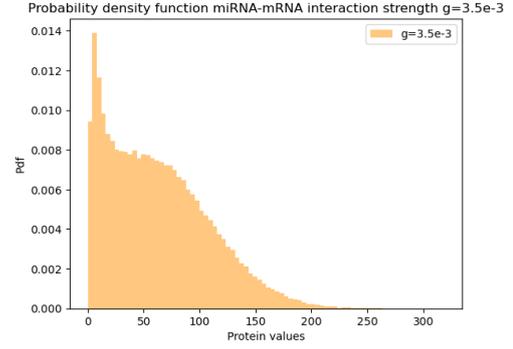
Figure 3.5: Different bimodal distribution for increasing transcription rate k_R . The higher target transcription rate is, the higher the number of free targets in the distribution and the more important the expressed state becomes.

3.1.2 Single-cell trajectory

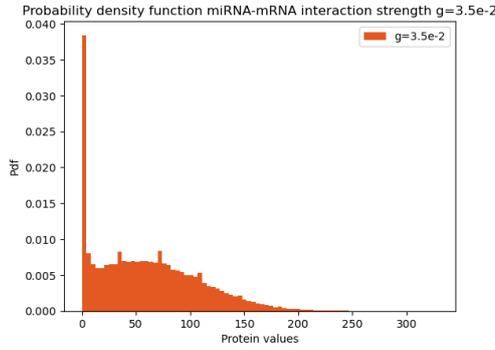
In the previous section, different bimodal population distributions generated by intrinsic noise have been shown. The idea behind their existence is that single cells composing the population jump from a repressed state to an expressed one, sampling the two modes of the distribution. The question we want to address now is whether we can actually observe a bimodal population distribution by looking at one single cell for a very long time. We are wondering if the system is really ergodic.



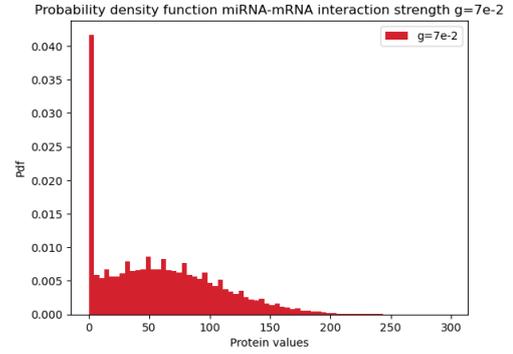
(a) Probability distribution function with miRNA-target interaction strength $g = 7 \cdot 10^{-4} s^{-1}$. The coupling is not strong enough to give rise to a bimodal population distribution.



(b) Probability distribution function with miRNA-target interaction strength $g = 3.5 \cdot 10^{-3} s^{-1}$.



(c) Probability distribution function with miRNA-target interaction strength $g = 3.5 \cdot 10^{-2} s^{-1}$.



(d) Probability distribution function with miRNA-target interaction strength $g = 7 \cdot 10^{-2} s^{-1}$.

Figure 3.6: Different bimodal population distribution for increasing miRNA-target interaction strength g . When this interaction strength is low, as in Figure 3.6b, the bimodal distribution can't be seen. Increasing the coupling between miRNA and target the distribution becomes bimodal and the expressed state begins to flatten.

In order to answer this question, a single Gillespie simulation has been run for a time $t = 30000000s$, keeping the same parameters used to simulate the reference bimodal distribution. As the protein value has been saved every time a reaction occurs, the vector of protein values is not equally spaced. To solve this issue, the trajectory 3.9a has been reconstructed (starting from the steady state $t_{steady} = 100000s$) by sampling it at temporal steps $\delta t = 0.1 * avg_t$, where avg_t is the average time between a reaction and the subsequent one. In order to obtain independent samples, we

need an estimate of the auto-correlation time of this new trajectory. This time has been estimated to be $t_{autocorr} = 340000s$ 3.9b. Sampling the trajectory with a step $t = t_{autocorr}$, the obtained data reproduce a bimodal population distribution. In order to verify if the obtained single cell distribution is compatible with the population distribution in Figure 3.3, an Approximate Two Sample Kolmogorov-Smirnov statistical test has been performed 3.7. This test is used to compare two different datasets (our single-cell and population distributions). The null hypothesis h_0 entails that the two datasets have been extracted from the same distribution [24]. As expected, it comes out that the two sets of data have been extracted from the same distribution with 95% of confidence and so a single cell trajectory is able to recapitulate the behaviour of the entire population.

```

Approximate two sample Kolmogorov-Smirnov test
-----
Population details:
  parameter of interest:  Supremum of CDF differences
  value under h_0:       0.0
  point estimate:         0.0273634

Test summary:
  outcome with 95% confidence: fail to reject h_0
  two-sided p-value:      0.0624

Details:
  number of observations: [2370,100000]
  KS-statistic:           1.316611038741642
    
```

Figure 3.7: Result of Kolmogorov-Smirnov statistical test performed on bimodal population distribution and single cell bimodal distribution. With 95% of confidence the null hypothesis, entailing that the compared datasets have been extracted from the same distribution, is verified.

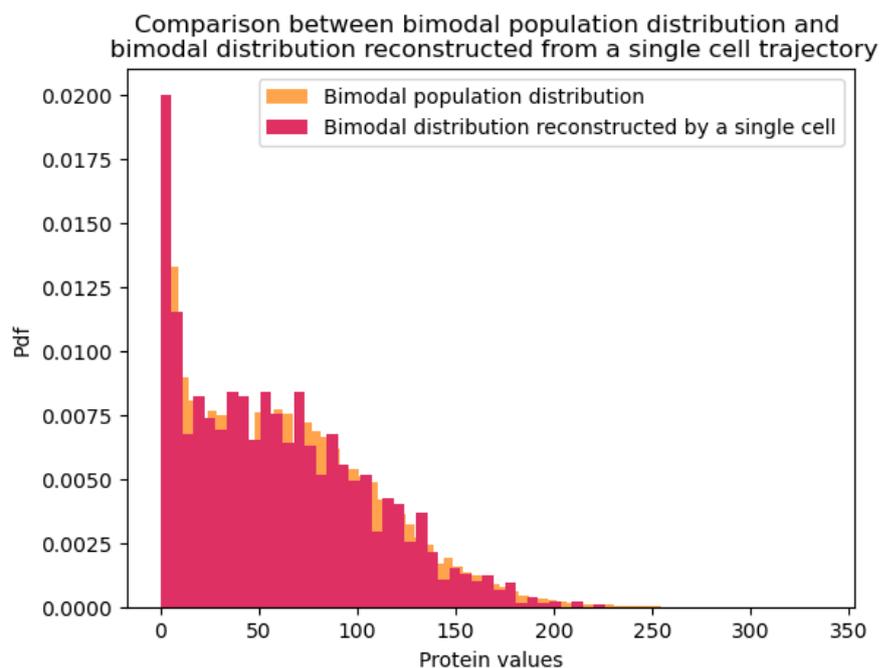
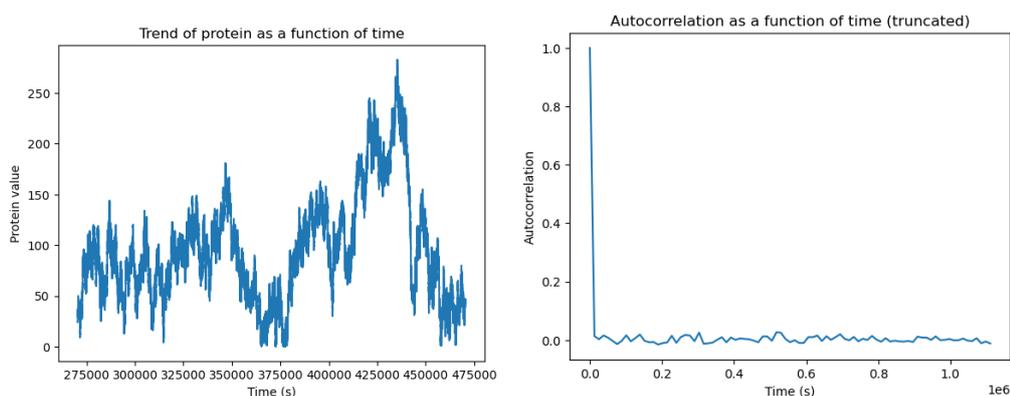


Figure 3.8: Comparison between bimodal population distribution and bimodal single cell distribution. Performing a Kolmogorov-Smirnov statistical test it has been proved that a single cell trajectory summarize the behaviour of the entire population.



(a) Reconstructed trajectory of a single cell simulated for a time $t = 30000000s$. (b) Auto-correlation computed on the reconstructed trajectory as a function of real time.

Figure 3.9: Trajectory of a single cell and its auto-correlation function.

3.1.3 Permanence times

We have just proven that single cells in presence of pure intrinsic noise jump from one state to the other. We want to describe this situation as a dynamic process between two states A and B in which particles jump from A to B with rate k_{AB} and from B to A with rate k_{BA} . Calling $p_A(t)$ the probability that our system is in state A at time t and $p_B(t)$ the probability that it's in state B at time t , we can write the following master equation:

$$\frac{dp_A(t)}{dt} = k_{BA}p_B(t) - k_{AB}p_A(t) \quad (3.9)$$

$$\frac{dp_B(t)}{dt} = k_{AB}p_A(t) - k_{BA}p_B(t) \quad (3.10)$$

This equation can be easily solved and its solution is of the form:

$$p_A(t) = p_A^{eq} + \delta p_A \cdot e^{-kt} \quad (3.11)$$

$$p_B(t) = p_B^{eq} + \delta p_B \cdot e^{-kt} \quad (3.12)$$

with the condition that $k = k_{AB} + k_{BA}$. This solution shows that the probability to be in a state decreases exponentially in time. In physical terms, the two states can be associated to the minima of a double well potential function.

Having seen how bimodal population distribution changes as a function of parameters g and k_r , permanence times of the population in the expressed and repressed state can be studied for various cases.

First of all, there is the need to define what is the expressed state and what is the repressed state. In this work, two different definitions have been used: the first is obtained considering the states as $\mu_i \pm b_i \cdot \sigma_i$ ($i=1,2$), where b_i is the considered percentage of σ_i , and the second one is to define as repressed all protein values $x_{repressed} \leq \mu_1 + b_1 \cdot \sigma_1$ and as expressed that values $x \geq \mu_2 - b_2 \cdot \sigma_2$. The function that computes these times has been defined in the following way: all values belonging to the repressed state have been set to the value 1, all those belonging to the expressed state have been set to the value 2 and those that don't belong neither to the first nor to the second, that are transition times, have been set to 3.

Let's begin by exploring permanence times of the reference probability distribution in Figure 3.3 with the first definition. At the beginning, the first permanence times have been studied with the idea to compare them with permanence times extracted from the single trajectory. These times are exponentially distributed both for the repressed state and for the expressed one (see Figure 3.10). The same is true for the distributions of all times (see Figure 3.11).

With the second definition, instead, the distributions of Figures 3.12 and 3.13 have been obtained. The distribution of first permanence times in the expressed state shows a peak on long times. A possible explanation for this phenomenon is the

following: with this definition we are considering values of protein expression that are in the tails of the distribution, so values that are very unlikely. We think that these peaks will disappear by increasing observation time.

Now we may wonder how permanence times of cells in the two states of the bimodal distribution change with changing miRNA-mRNA interaction strength g and target transcription rate k_R . Let's begin with the case of bimodal population distribution with $k_R = 0.107s^{-1}$, $k_R = 0.112s^{-1}$ and $k_R = 0.115s^{-1}$. The analysis has been performed using the definition $x_{repressed} \leq \mu_1 + b_i \cdot \sigma_1$ and $x_{expressed} \geq \mu_2 - b_i \cdot \sigma_1$ and the results are in Figures 3.10, 3.11, 3.12 and 3.13.

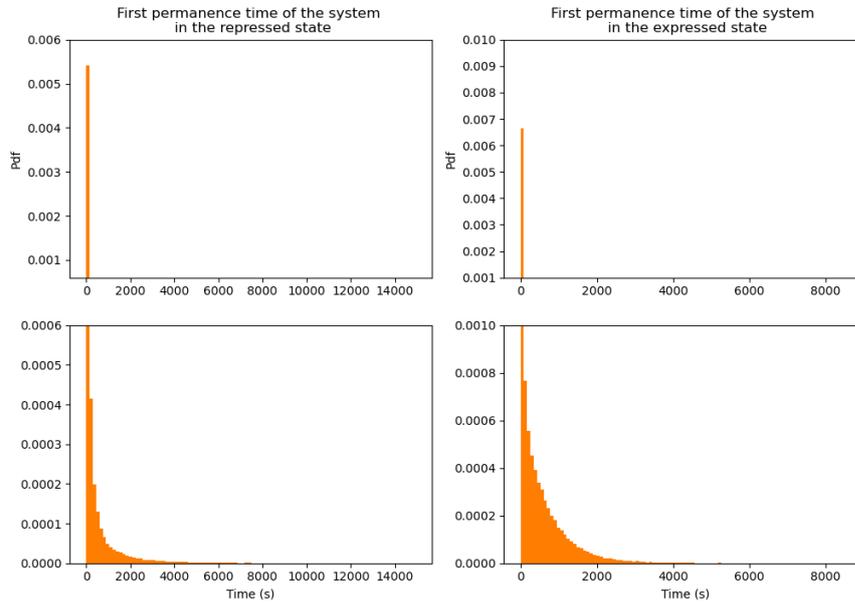


Figure 3.10: First permanence times of the system with $k_R = 0.1095s^{-1}$ for repressed and expressed state in case of pure intrinsic noise with states defined as $\mu_i \pm b_i \cdot \sigma_i$.

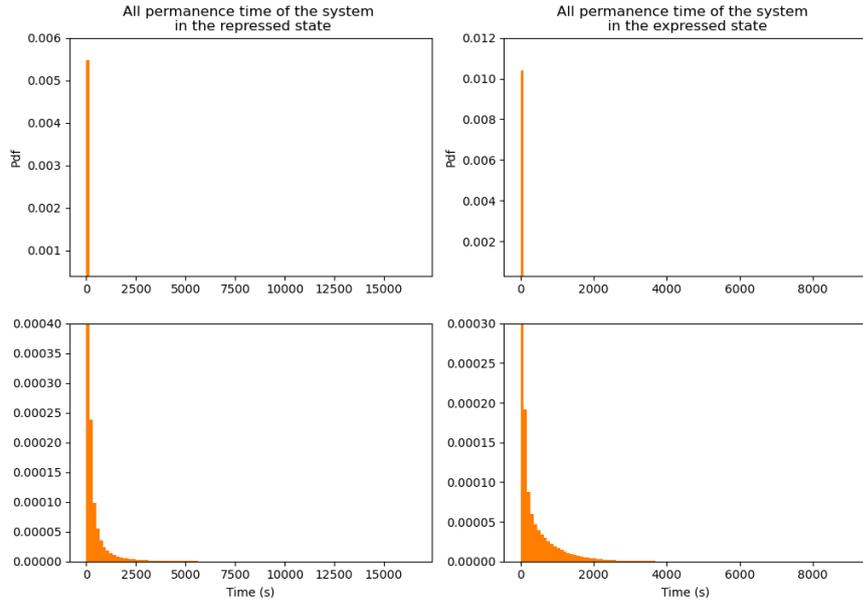


Figure 3.11: All permanence times of the system with $k_R = 0.1095s^{-1}$ for repressed and expressed state in case of pure intrinsic noise with states defined as $\mu_i \pm b_i \cdot \sigma_i$.

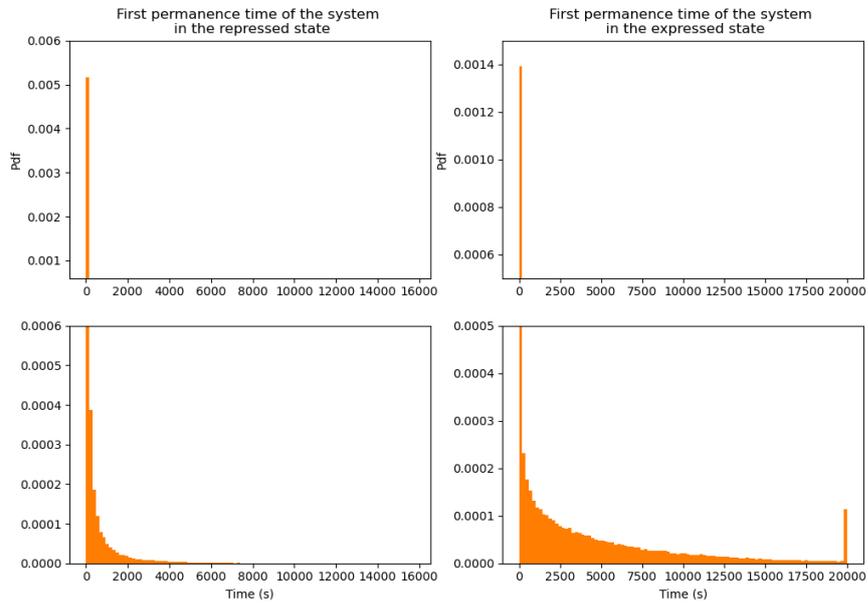


Figure 3.12: First permanence times of the system with $k_R = 0.1095s^{-1}$ for repressed and expressed state in case of pure intrinsic noise with states defined as $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

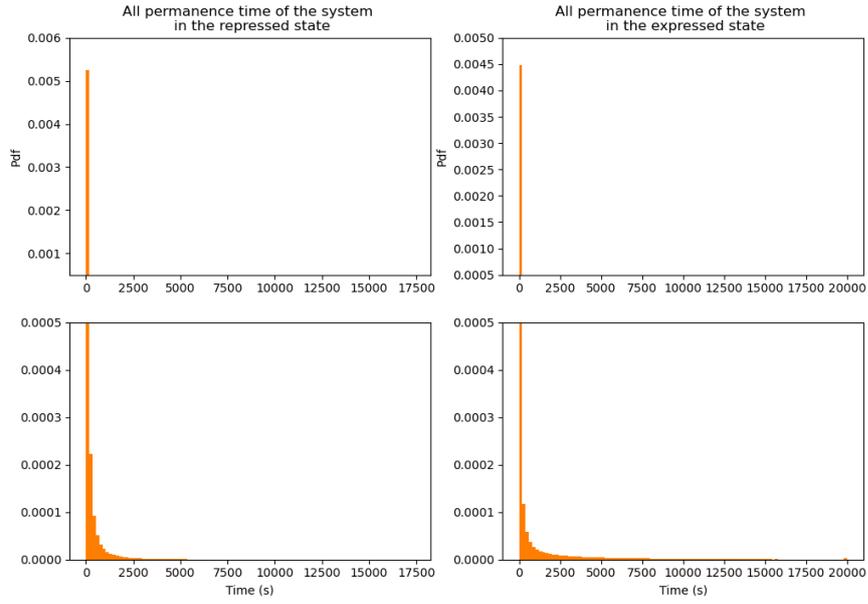


Figure 3.13: All permanence times of the system with $k_R = 0.1095s^{-1}$ for repressed and expressed state in case of pure intrinsic noise with states defined as $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

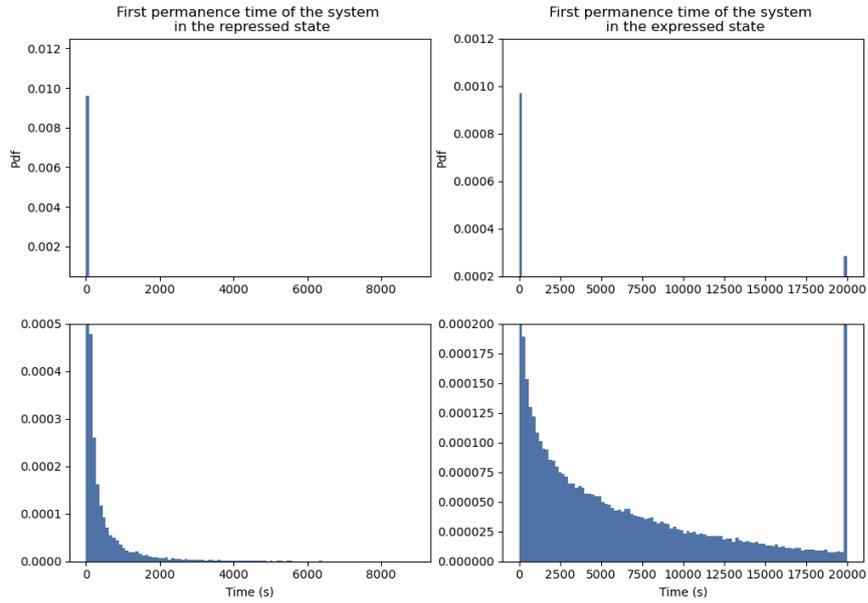


Figure 3.14: First permanence times of the system with $k_R = 0.107s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

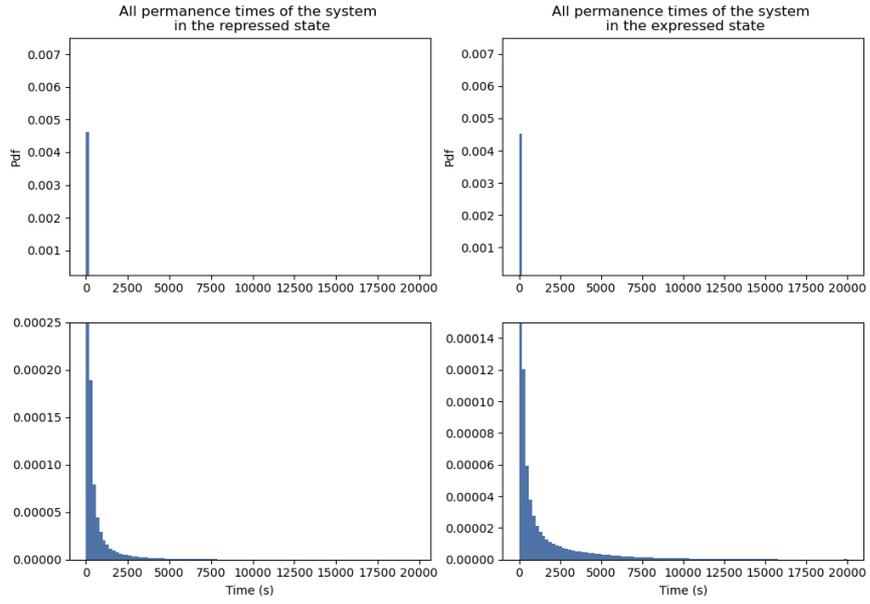


Figure 3.15: All permanence times of the system with $k_R = 0.107s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

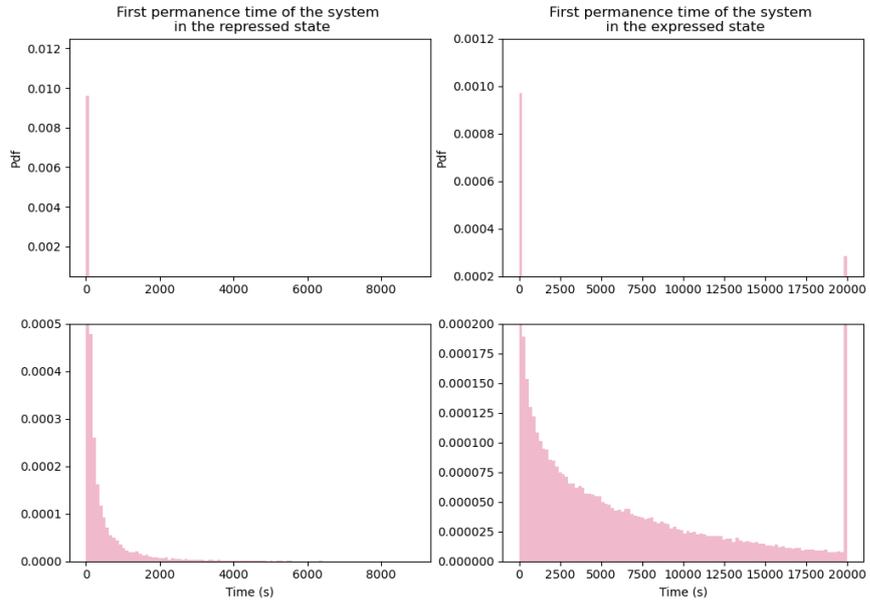


Figure 3.16: First permanence times of the system with $k_R = 0.112s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

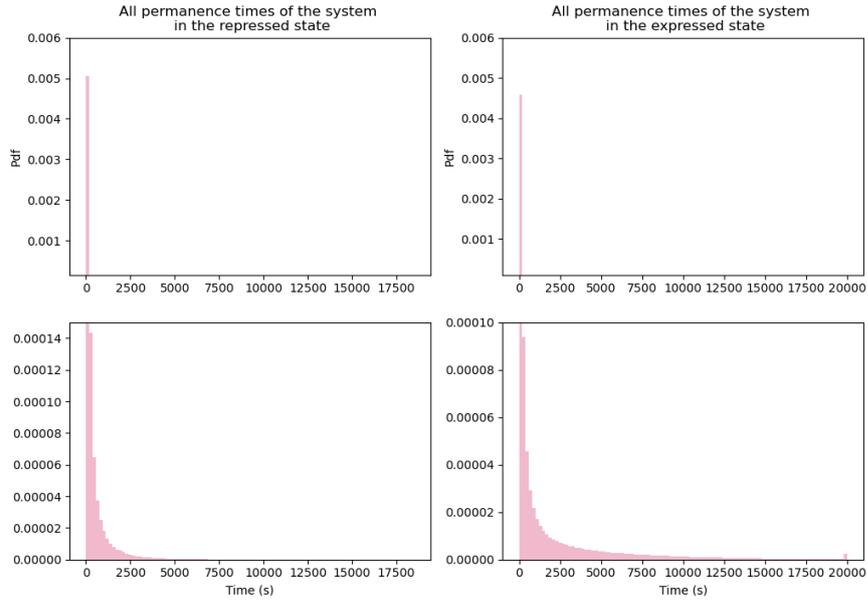


Figure 3.17: All permanence times of the system with $k_R = 0.112s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

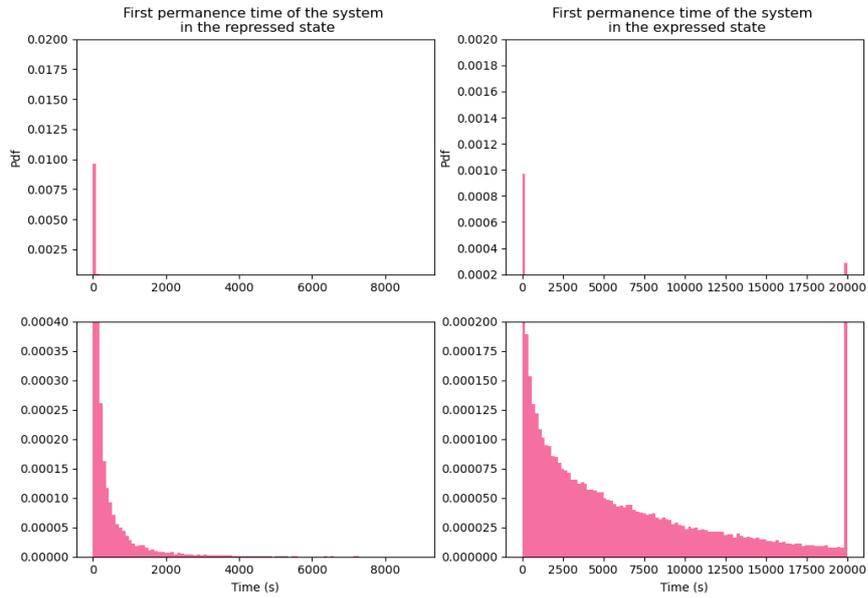


Figure 3.18: First permanence times of the system with $k_R = 0.115s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

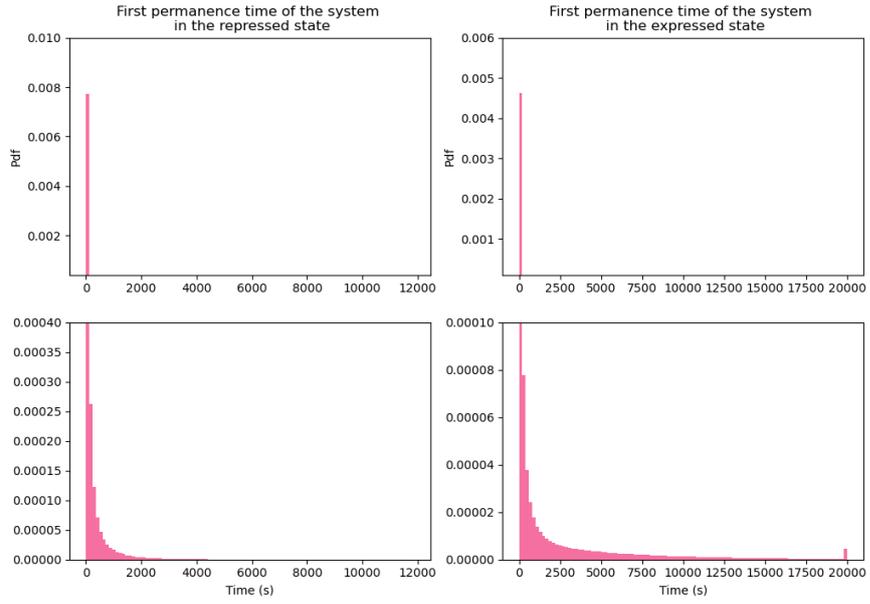


Figure 3.19: All permanence times of the system with $k_R = 0.115s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

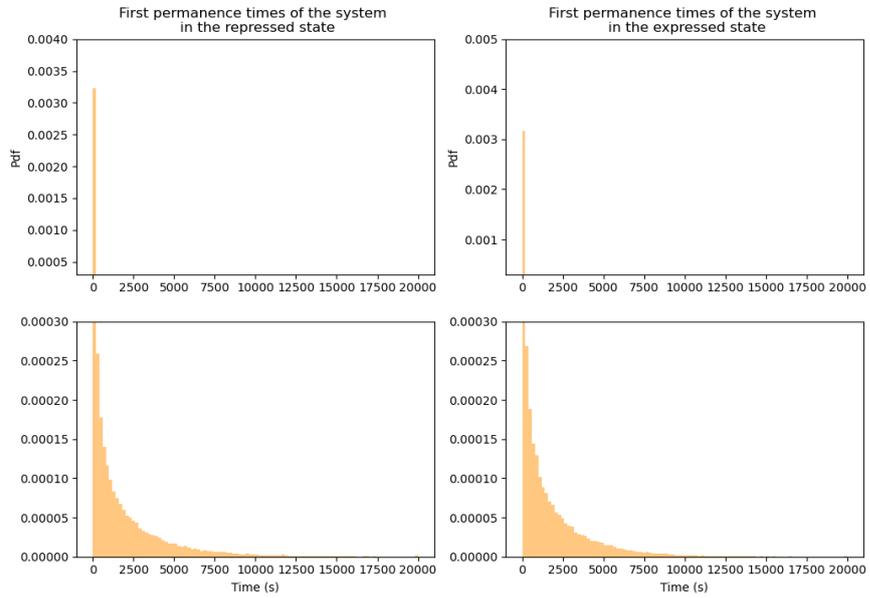


Figure 3.20: First permanence times of the system with $g = 3.5 * 10^{-3}s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

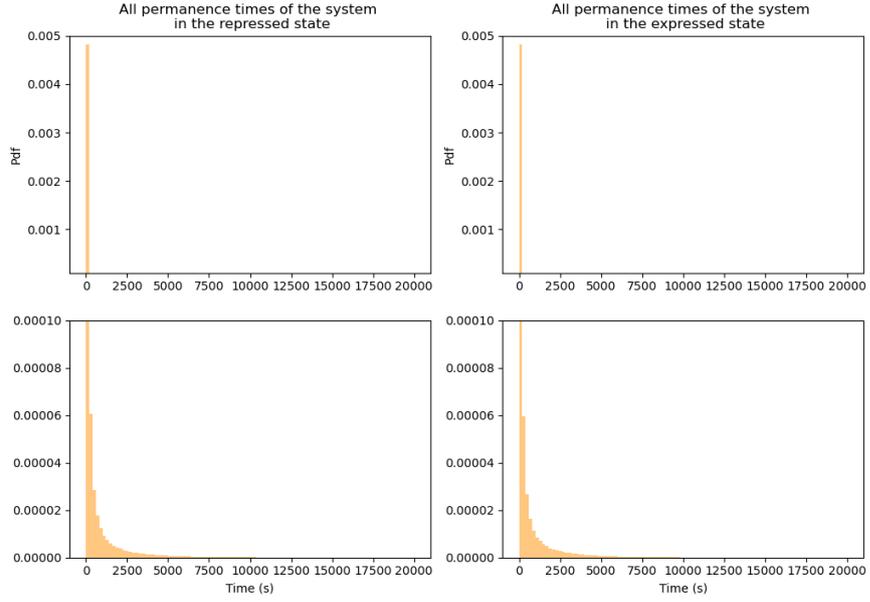


Figure 3.21: All permanence times of the system with $g = 3.5 \cdot 10^{-3}s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

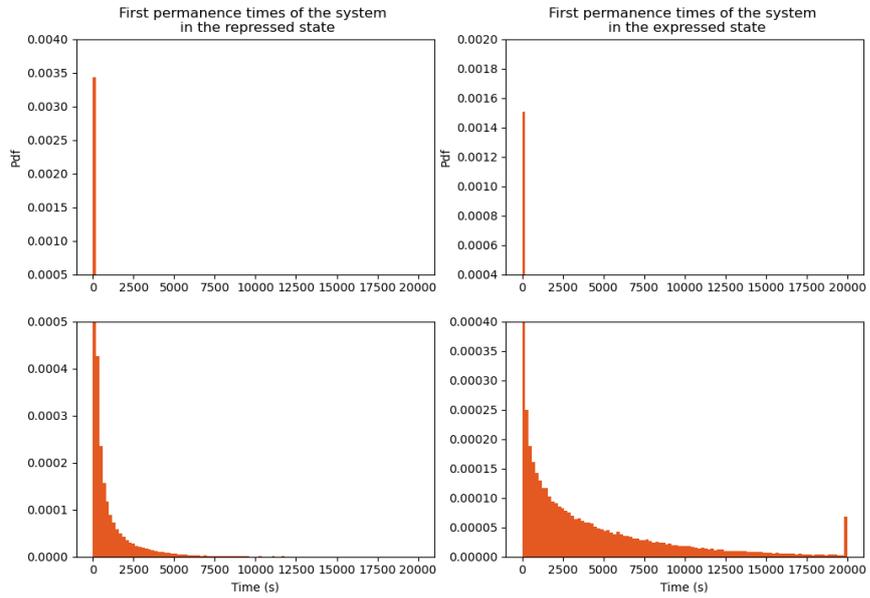


Figure 3.22: First permanence times of the system with $g = 3.5 \cdot 10^{-2}s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

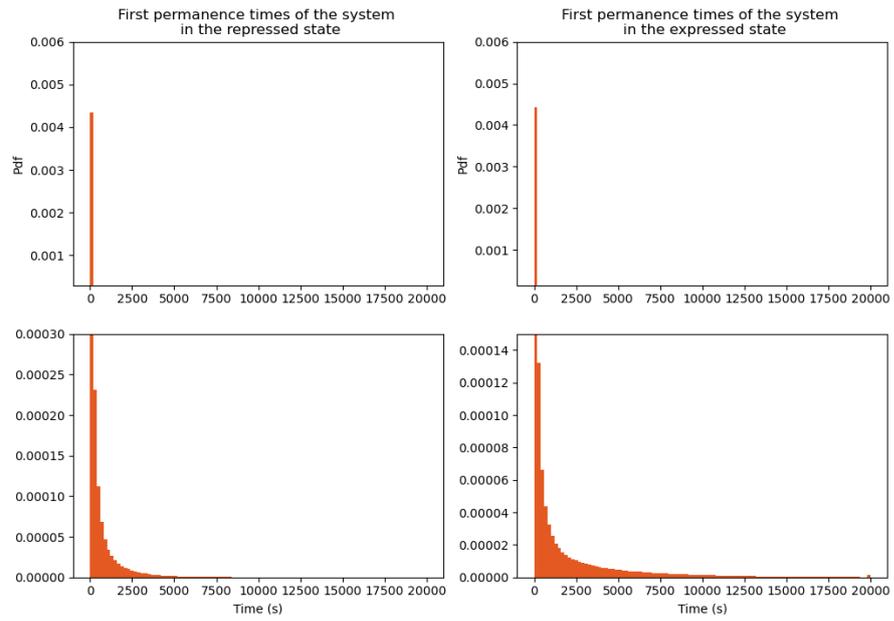


Figure 3.23: All permanence times of the system with $g = 3.5 * 10^{-2}s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

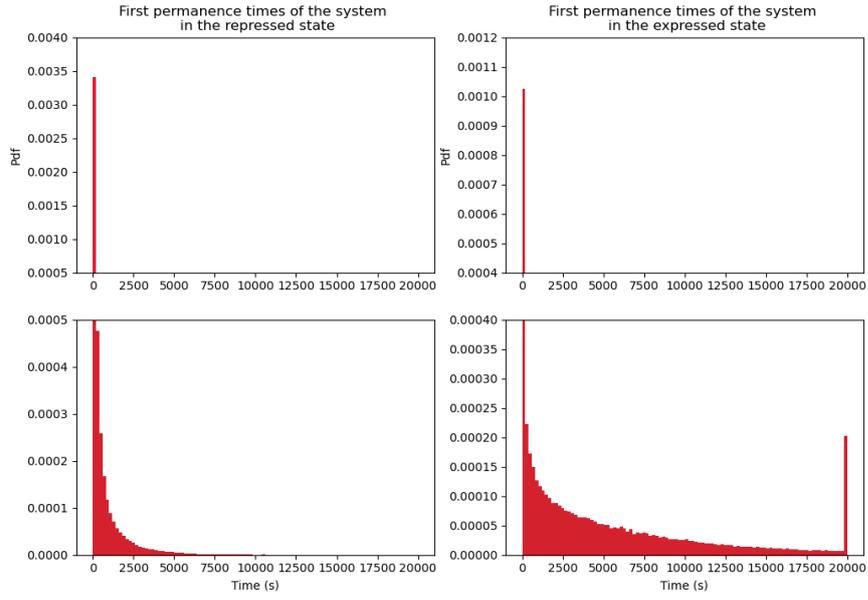


Figure 3.24: First permanence times of the system with $g = 7 * 10^{-2}s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

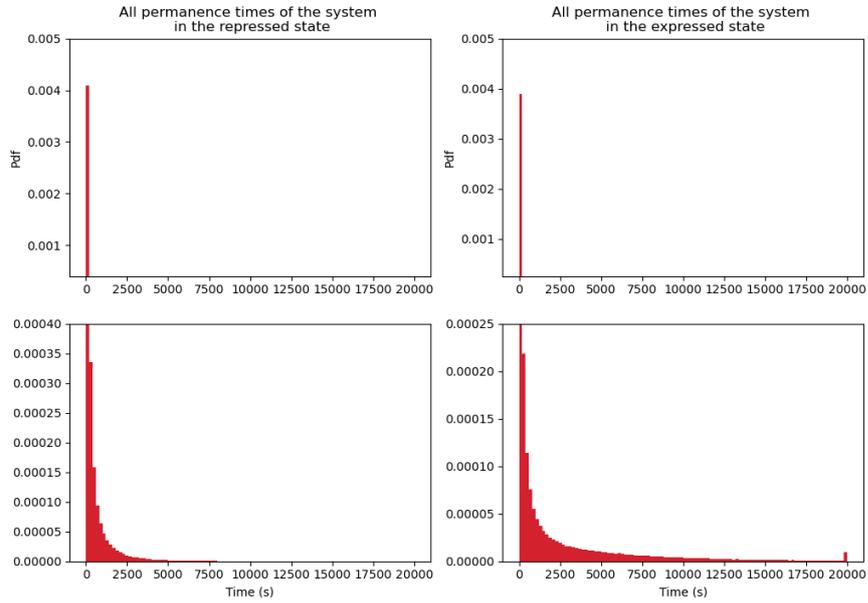


Figure 3.25: All permanence times of the system with $g = 7 * 10^{-2}s^{-1}$ in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

Figure 3.26: Caption

The same procedure has been followed for the distributions with $g = 3.5 \cdot 10^{-3} s^{-1}$, $g = 3.5 \cdot 10^{-2} s^{-1}$ and $g = 7 \cdot 10^{-2} s^{-1}$. Results can be found in Figures 3.20, 3.21, 3.22, 3.23, 3.24 and 3.25. These results are qualitative but we would like to have a quantitative estimate of decay rates of cell population permanence times in the repressed and expressed states of the system. All these distributions with repressed and expressed state defined as $x_{repressed} < \mu_1 + b_1 \cdot \sigma_1$ and $x_{expressed} > \mu_2 + b_2 \cdot \sigma_2$ have been fit with a single normalized exponential function of the following form:

$$exp(t) = const \cdot \exp(-kt) \quad (3.13)$$

with the constraint that

$$\int_0^{\infty} const \cdot \exp(-kt) = 1 \quad (3.14)$$

implying that $const = k$. With this expression the results are those shown in Table 3.4. Fit of the distributions are shown in Figures 3.27, 3.28, 3.29 and 3.30. Being these distribution really flattened on the y axis, we restricted x range from 1s to 100s to see the trend in a clearer way.

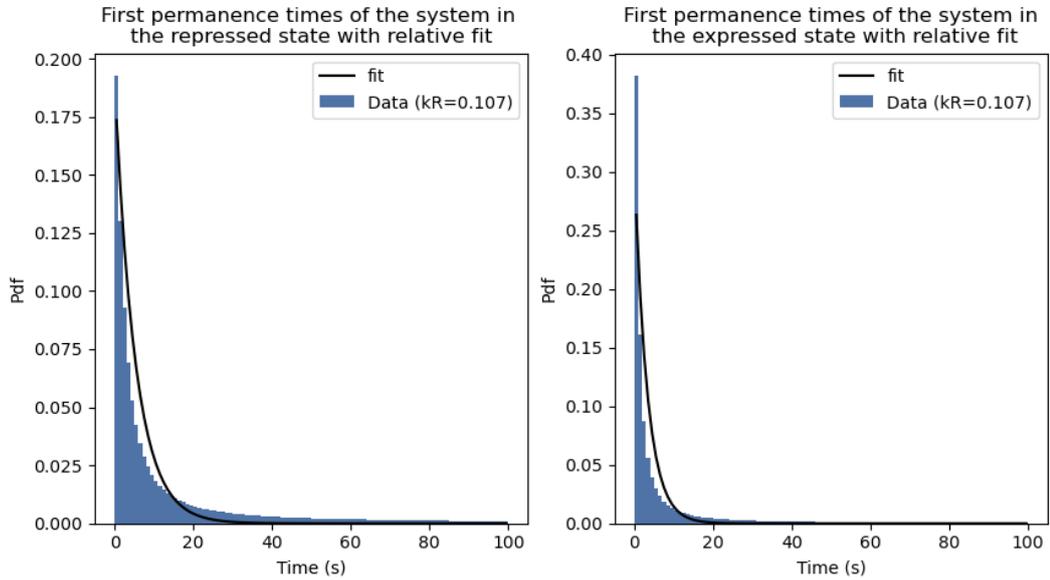


Figure 3.27: Permanence times of cell population with transcription rate $k_R = 0.107 s^{-1}$ in repressed and expressed states with relative fit. Time range at which we look at has been restricted to 100s.

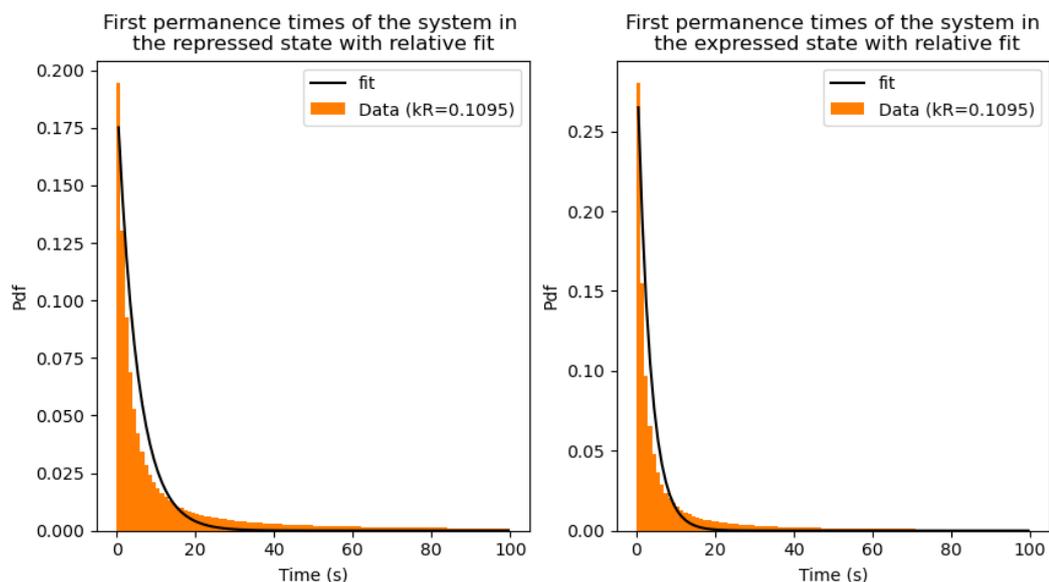


Figure 3.28: Permanence times of cell population with transcription rate $k_R = 0.1095s^{-1}$ in repressed and expressed states with relative fit. Time range at which we look at has been restricted to 100s.

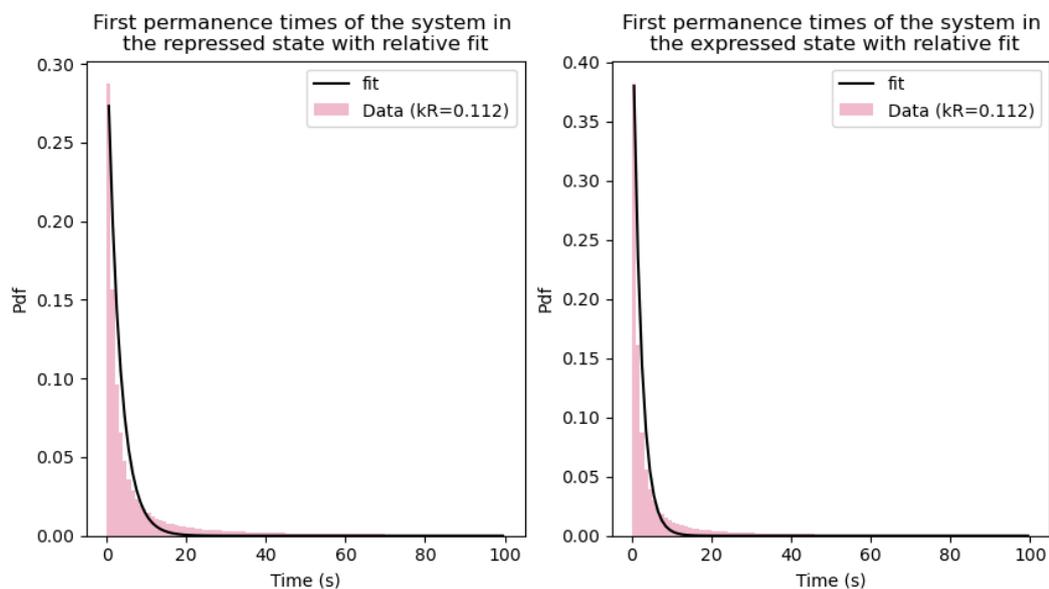


Figure 3.29: Permanence times of cell population with transcription rate $k_R = 0.112s^{-1}$ in repressed and expressed states with relative fit. Time range at which we look at has been restricted to 100s.

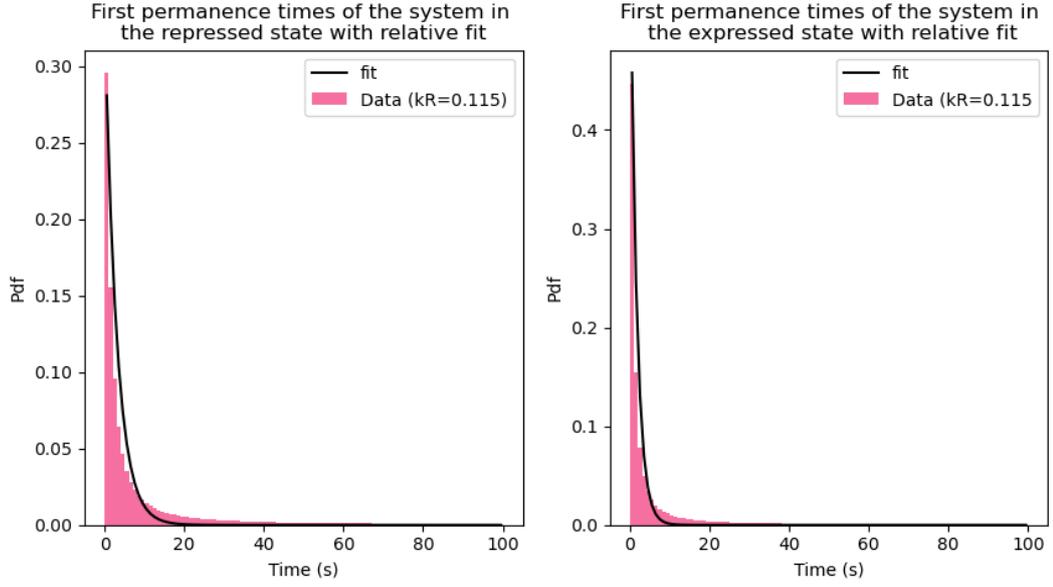


Figure 3.30: Permanence times of cell population with transcription rate $k_R = 0.115s^{-1}$ in repressed and expressed states with relative fit. Time range at which we look at has been restricted to 100s.

The decay rates for repressed and expressed states in Figures 3.27, 3.28, 3.29 and 3.30 are reported in Table 3.4.

$k_R (s^{-1})$	$k_1 (s^{-1})$	$k_2 (s^{-1})$
0.107	0.191 ± 0.012	0.307 ± 0.019
0.1095	0.193 ± 0.012	0.309 ± 0.019
0.112	0.32 ± 0.02	0.48 ± 0.03
0.115	0.33 ± 0.02	0.63 ± 0.04

Table 3.4: Values of decay rates, i.e. the inverse of permanence times of cell population distribution in the repressed (k_1) and expressed (k_2) states, for different target transcription rates.

g (s^{-1})	k_1 (s^{-1})	k_2 (s^{-1})
$3.5 \cdot 10^{-3}$	0.77 ± 0.05	0.88 ± 0.05
$7 \cdot 10^{-3}$	0.193 ± 0.012	0.309 ± 0.019
$3.5 \cdot 10^{-2}$	0.17 ± 0.01	0.35 ± 0.02
$7 \cdot 10^{-2}$	0.09 ± 0.01	0.35 ± 0.02

Table 3.5: Values of decay rates, i.e. the inverse of permanence times of cell population distribution in the repressed (k_1) and expressed (k_2) states, for different values of miRNA-mRNA interaction strength.

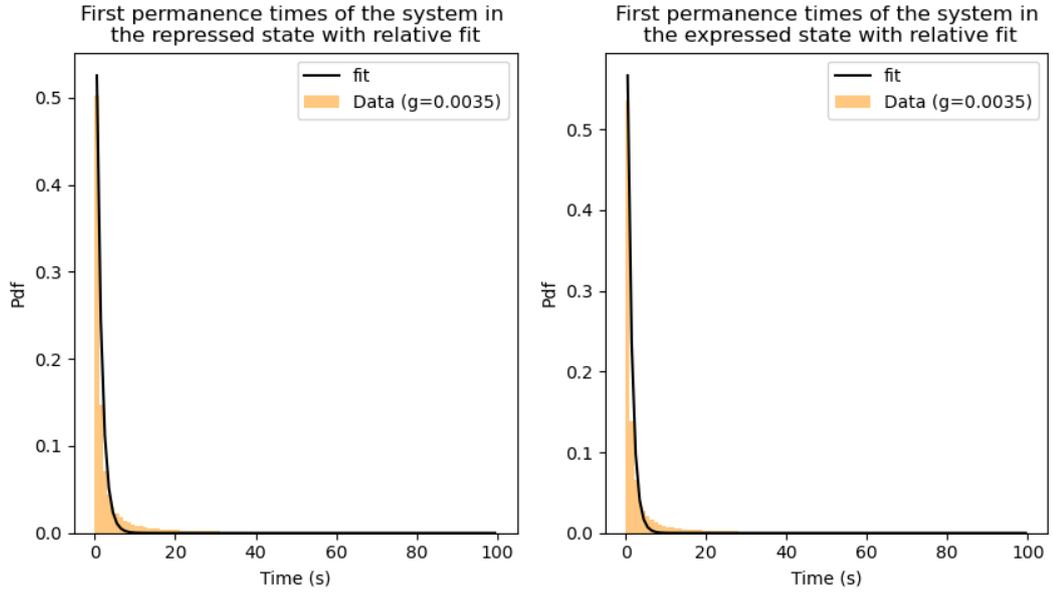


Figure 3.31: Permanence times of cell population with transcription rate $g = 0.0035s^{-1}$ in repressed and expressed states with relative fit. Time range at which we look at has been restricted to 100s.

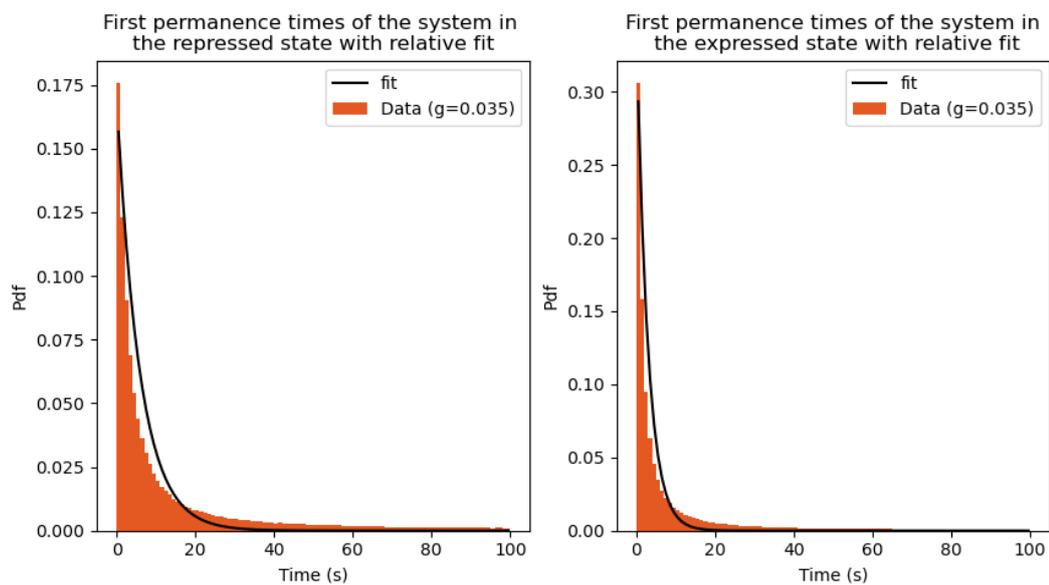


Figure 3.32: Permanence times of cell population with transcription rate $g = 0.035s^{-1}$ in repressed and expressed states with relative fit. Time range at which we look at has been restricted to 100s.

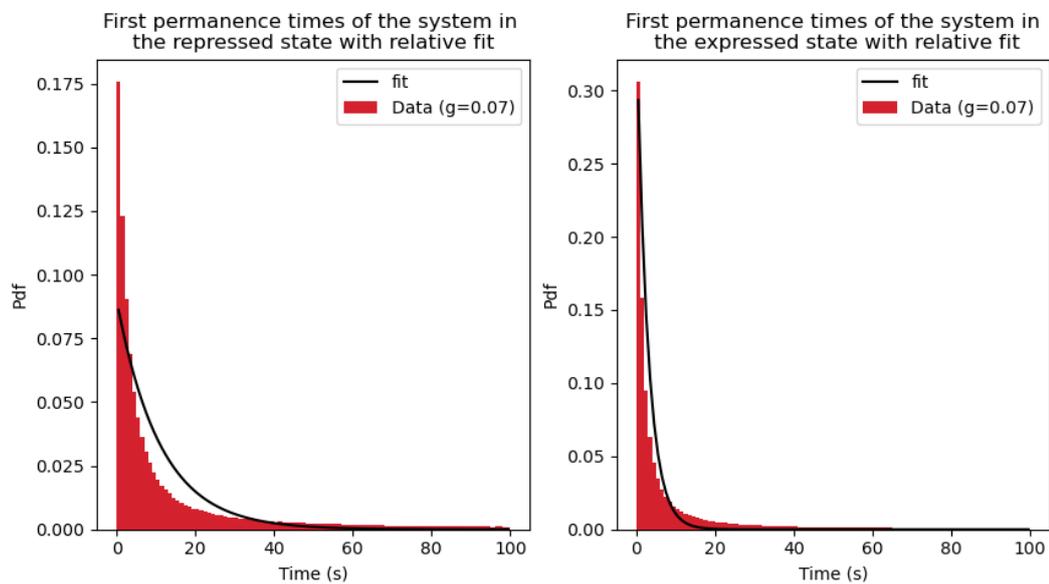


Figure 3.33: Permanence times of cell population with transcription rate $g = 0.07s^{-1}$ in repressed and expressed states with relative fit. Time range at which we look at has been restricted to 100s.

All we can say is that increasing target transcription rates lead to an increasing of decay rates of cell population in both states of the system, meaning that permanence times of cells in the states will be shorter. When increasing miRNA-target interaction strength, instead, decay rates of cell population decrease and the system will pass more time in that states.

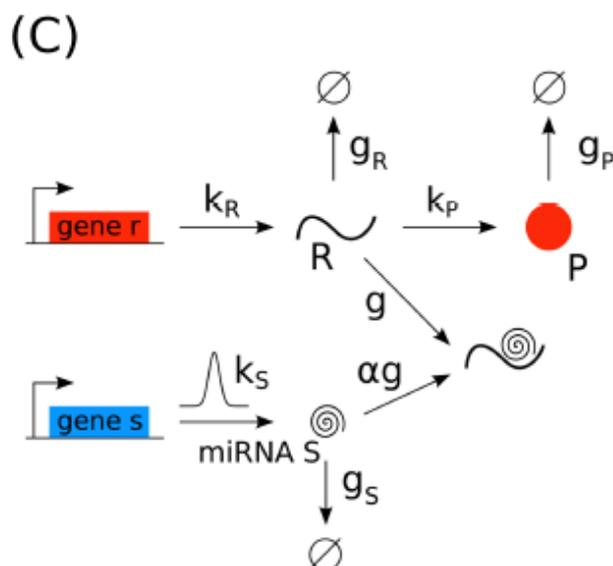


Figure 3.34: Model of the system in presence of extrinsic noise. Gene r is transcribed at constant rate k_R into target mRNA R , which can be translated into protein P at rate k_P or be degraded with rate g_R or still can interact with a fraction α of microRNA S . This latter is transcribed by gene s at a fluctuating transcription rate, extracted from a Gaussian distribution. MiRNA can then be degraded at rate g_S and protein can be degraded as well at rate g_P [8]. Figure adapted from [8].

3.2 Extrinsic noise-induced bimodality

The model which accounts for the extrinsic noise is pretty similar to the model for intrinsic noise. A gene r codes for target mRNA R , which is translated into protein and can interact with miRNA S , transcribed by gene s independently, with an interaction strength g . The difference with respect to the previous model is that miRNA transcription rate is not constant anymore but it is rather drawn from a Gaussian distribution peaked around a mean value μ with a certain standard deviation σ . Then target, protein and miRNA can be degraded at constant rates

g_R , g_P and g_S .

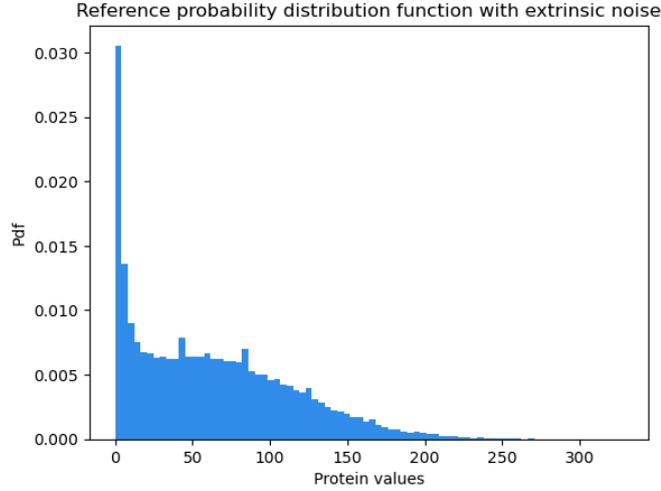


Figure 3.35: Extrinsic noise-induced bimodal population distribution generated with the same parameter of Table 3.1. Extrinsic noise has been taken into account by allowing miRNA transcription rate k_S fluctuations from cell to cell. The value of k_S has been drawn by a Gaussian distribution with mean $\mu = 0.085s^{-1}$ and standard deviation $\sigma = 5 \cdot 10^{-3}s^{-1}$.

At this point, to investigate the effect of extrinsic noise on the system, let's take into account the bimodal population distribution of Figure 3.3: the transcription rate for microRNA will be drawn from a Gaussian distribution with mean $\mu = 0.085s^{-1}$, which is the same used in Fig. 3.3, and a standard deviation $\sigma = 5 \cdot 10^{-3}s^{-1}$. The resulting distribution is displayed in Figure 3.35. As we can see from the Figure 3.35, the effect of extrinsic noise on distribution 3.3 is to slowly flatten the expressed state and to raise the first peak. This effect is due to the fact that we are sampling values both under and over threshold. However, the two distributions do not differ too much because the standard deviation σ of the Gaussian distribution from which miRNA transcription rates are sampled is not high. Then, we can try to increase it to $\sigma = 1 \cdot 10^{-2}s^{-1}$ to clearly see the effect (see Figure 3.36).

Although in this case the distribution does not undergo radical changes, there might be cases in which extrinsic noise leads to a completely different distribution. Let's take into account the distribution which has an interaction strength $g = 7 \cdot 10^{-4}s^{-1}$, see Figure 3.20 - it is not bimodal - and try to add an extrinsic noise to it. The Gaussian distribution from which miRNA transcription rate is extracted will be peaked on the same value of k_S used for intrinsic case and the standard deviation will be $\sigma = 1 \cdot 10^{-2}s^{-1}$. The shape of the distribution has passed from

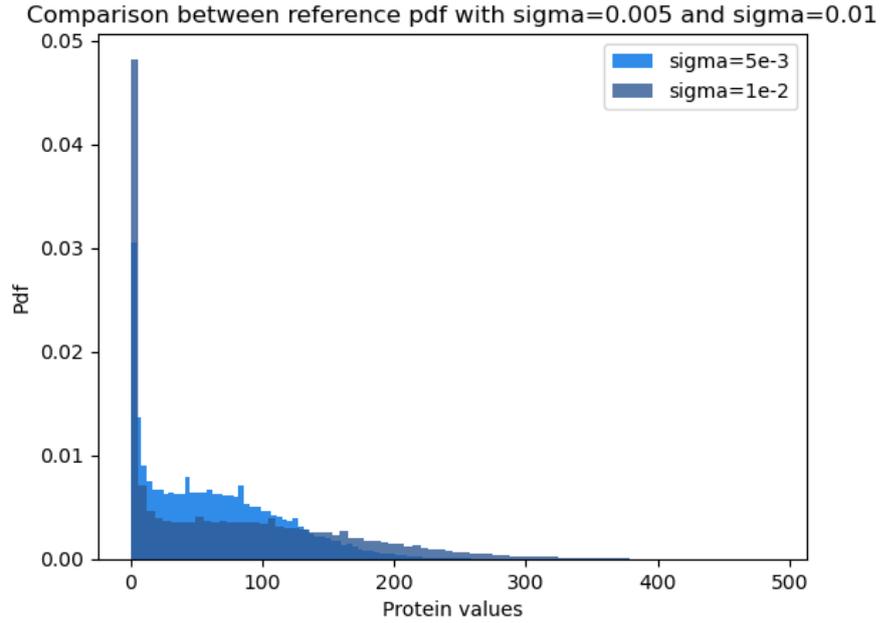
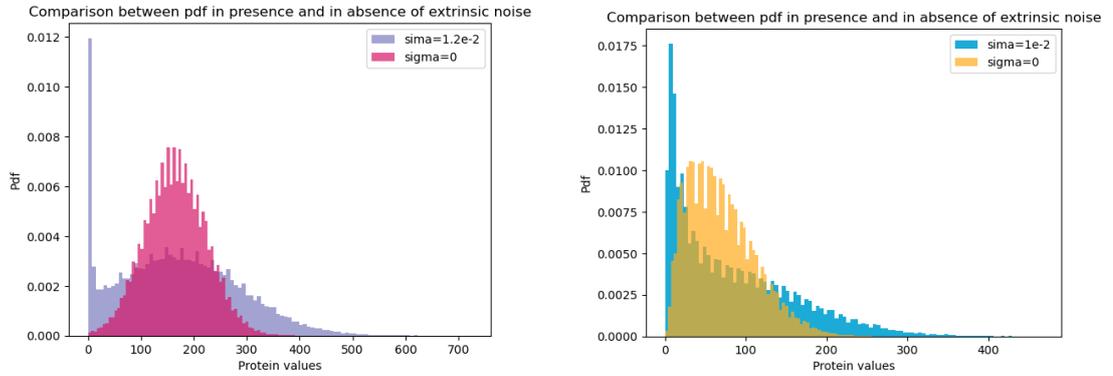


Figure 3.36: Comparison between two extrinsic noise-induced bimodal population distributions. They have been generated with the parameters in Table 3.1 but the standard deviations σ_i ($i=1,2$) of the distribution from which miRNA transcription rate has been drawn are different. Increasing σ corresponds to increasing the extrinsic noise.



(a) Comparison between Gaussian population for intrinsic noise with $g = 7 \cdot 10^{-4} s^{-1}$ and corresponding extrinsic noise counterpart, which generates a bimodal distribution.

(b) Comparison between Gaussian population for intrinsic noise with $k_R = 0.124 s^{-1}$ and corresponding extrinsic noise counterpart, which generates a bimodal distribution.

Figure 3.37: Extrinsic noise can induce bimodality even when intrinsic noise cannot.

Gaussian to exponential. Another insightful example can be that of considering Fig. 3.5c, which is a Gaussian distribution with target transcription rate heavily over the threshold. Let's add extrinsic noise to this distribution. We will draw miRNA transcription rate from a Gaussian distribution peaked around the mean values $\mu = 0.085s^{-1}$ with a standard deviation $\sigma = 1.2 * 10^{-2}s^{-1}$.

In order to have a physical insight of what is happening, we could think of simulating the system (see Figure 3.1) represented by ODEs 3.1 with the lowest and the highest miRNA transcription rates of the above mentioned Gaussian distribution. As we can see from Figure 3.43, when miRNA transcription rate is lower than the reference one, $k_S = 0.085s^{-1}$, the system moves to the expressed state because there will be more free targets than miRNAs, that are almost all bound in complexes with their target mRNAs. On the contrary, when miRNA transcription rate is higher we are moving toward the threshold with an equimolarity situation between miRNAs and targets.

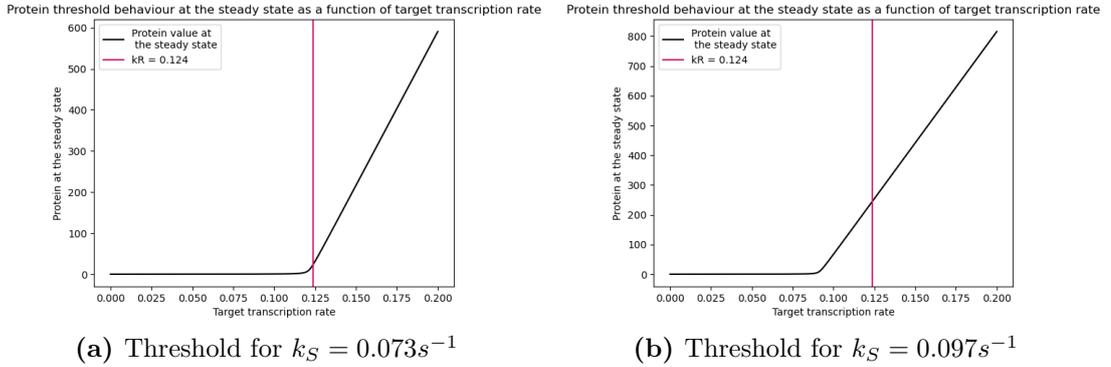


Figure 3.38: Moving miRNA transcription rate k_S allows to draw values in a large interval. On the left, threshold's location is represented for the highest possible k_S extracted, that is mean value plus a standard deviation. On the right, threshold's location is represented for the lowest possible k_S extracted, that is mean value minus a standard deviation.

3.2.1 Permanence times

It has been shown that adding extrinsic noise to the system can lead to substantial changes in the population distribution under study: we expect that also permanence times of the system in its two states will change. In fact, while previously the population distribution was made of cells jumping in and out of the two states, now there are some cells that never change state. We want to study how the presence of these cells impacts on permanence times.

As before, the first thing to do is to fit the bimodal distribution and to extract the

two means and standard deviations. These parameters will be used to compute permanence times with the definitions $\mu_1 \pm b_1 \cdot \sigma_1$ for the repressed state and $\mu_2 \pm \sigma_2$ for the expressed one. The resulting cells permanence times are displayed in Figure 3.39 and Figure 3.40.

Let's take into account the first cells permanence times, remembering that they have been introduced to make a comparison with all permanence times. The distribution of first cells permanence times in the expressed state is heavily non exponential. As we expect, it shows a high peak on large times due to those cells that are stucked in one of the two states. The distribution of times for the repressed state is still exponential due to the effect of fluctuations: being the repressed state really narrow, fluctuations may lead cells to jump in and out of the state, distorting permanence times.

Now, looking at the distributions of all cells permanence times, one can see that they reproduce the same behavior of Figures 3.39 and 3.40: the repressed state still has an exponential cells permanence times distribution due to the reduced width we are considering as "repressed state". For the expressed state, there is still a peak on long times. This fact marks a distinction between cells permanence times computed on intrinsic vs. extrinsic noise-induced bimodal distributions.

Another definition for repressed and expressed states could be that of considering as repressed all the values on the left of $\mu_1 + b_1 \cdot \sigma_1$ and as expressed the values on the right of $\mu_2 - b_2 \cdot \sigma_2$. With this definition, the distributions for the repressed state are still exponential both for the first and for all permanence times: as we said before, the form of bimodal distribution from which extracted means and standard deviations, forces us to consider a narrow range for the "repressed state". For the expressed state's cells permanence times, instead, a peak on long times appears both for the distribution of all and first cells permanence times.

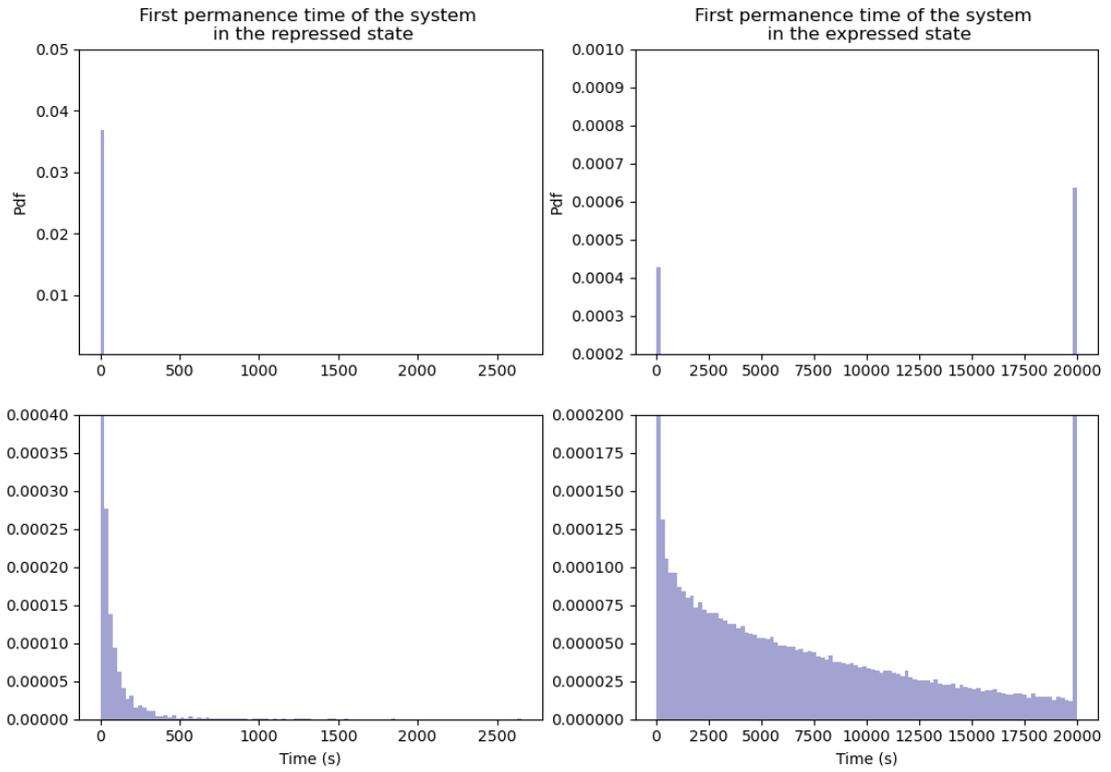


Figure 3.39: First permanence times of the system in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

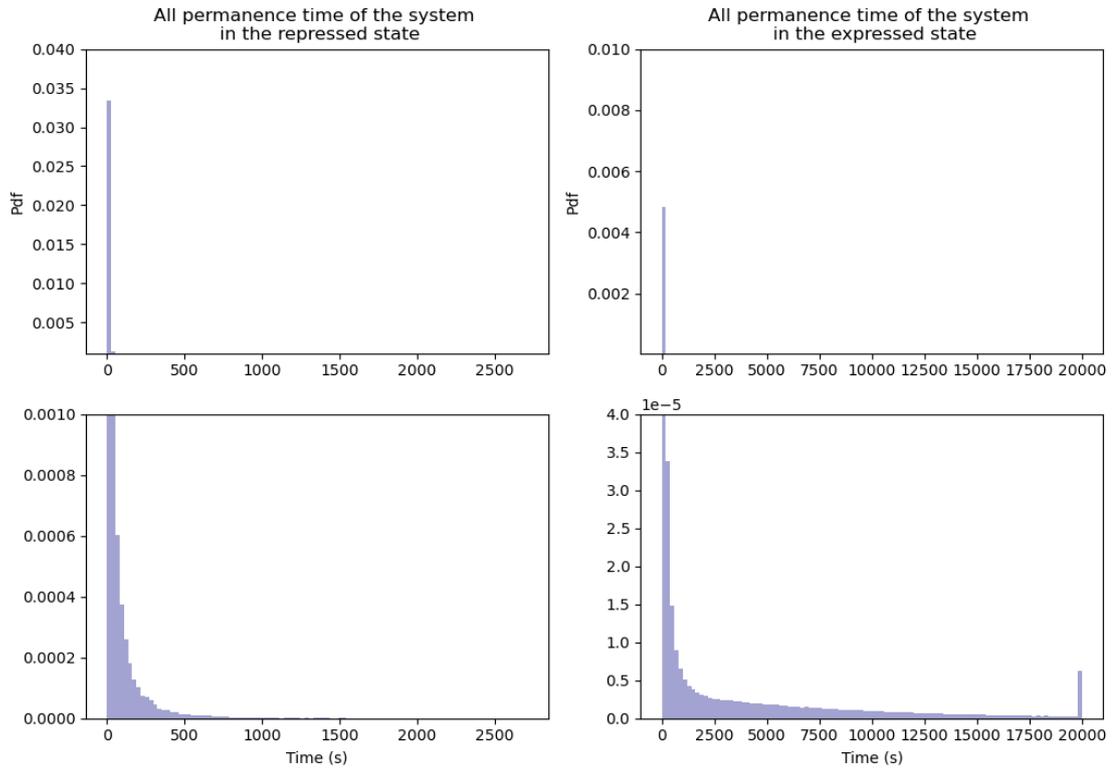


Figure 3.40: All permanence times of the system in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

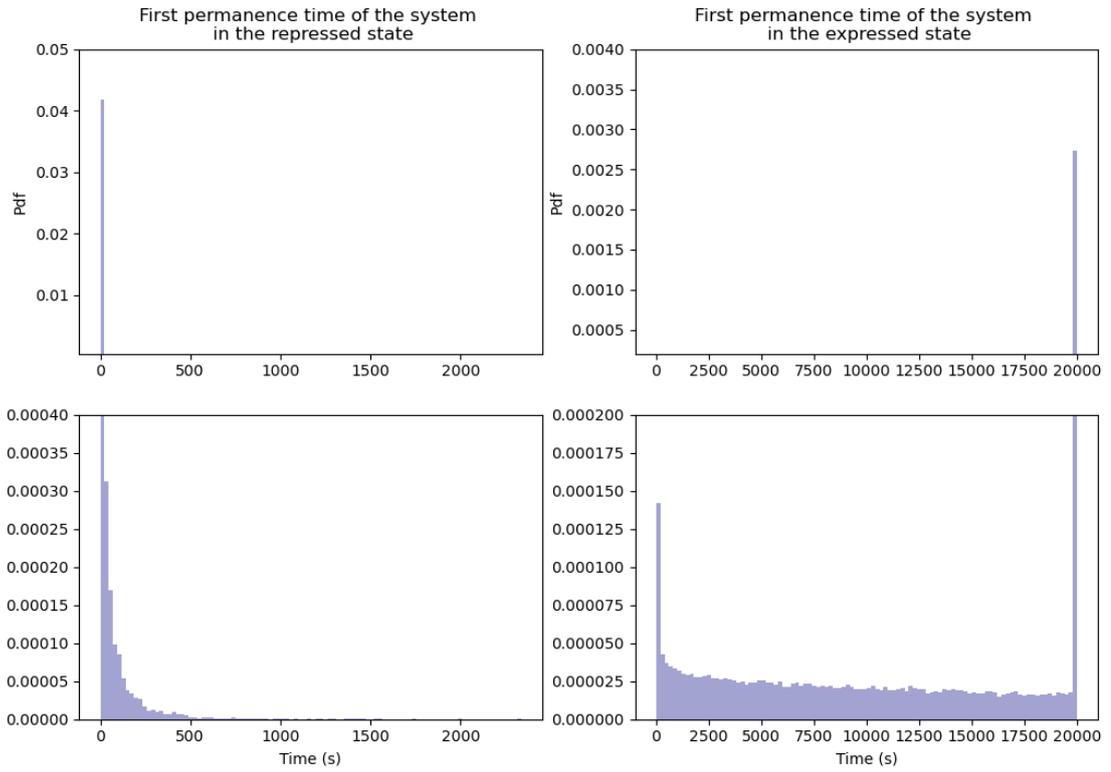


Figure 3.41: First permanence times of the system in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

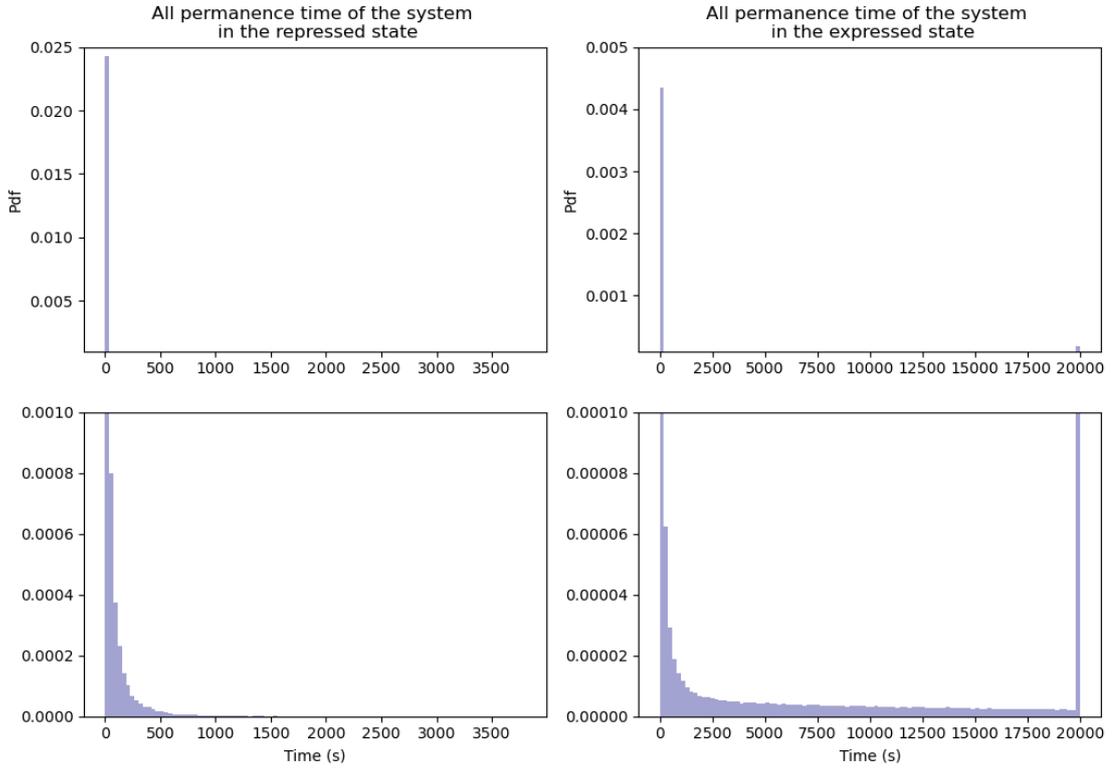


Figure 3.42: All permanence times of the system in the repressed and the expressed state computed using the definition $x_{repress} \leq \mu_1 + b_1 \cdot \sigma_1$ and $x_{express} \geq \mu_2 - b_2 \cdot \sigma_2$.

3.3 Experimental data

3.3.1 Bimodal distributions

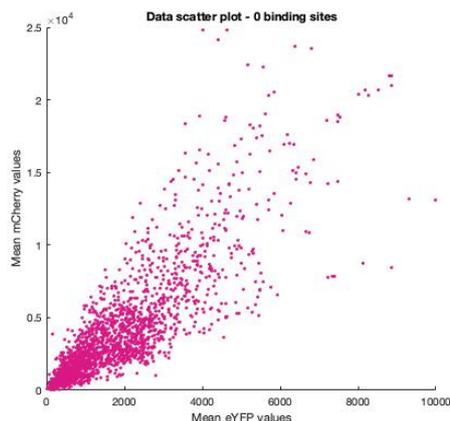
Let's recall what has been found theoretically: intrinsic noise-induced bimodal distributions are found only near the threshold and the coupling between miRNA and target must be sufficiently high [8]. Furthermore, increasing miRNA-mRNA interaction strength g the threshold becomes sharper and the range in which the bimodal distribution is allowed becomes wider. The opposite happens when decreasing g . Upon increasing target transcription rate k_r the population distribution concentrates in the expressed state becoming Gaussian while lowering it the distribution is exponentially decreasing due to the low concentration of free targets. Adding an extrinsic noise contributes to widen the region in which bimodal distribution can be found [8]. Now, we want to investigate experimental data with the idea to distinguish what is the principal source of noise by looking at

permanence times of cells in both repressed and expressed states of the system. As a first thing, we should identify bimodal distributions. We recall from previous sections that we are monitoring the system with a two-color fluorescent reporter system which gives us information about protein levels in presence and in absence of regulation. The constitutive gene is represented by eYFP, helpful to monitor transcriptional activity in single cells, while mCherry is the target protein. In performing time-lapse microscopy, screenshots have been taken every twenty minutes both on the eYFP channel and on mCherry one. We will work with data going from frame 1, corresponding to $\approx 24h$ after transfection, to frame 170, corresponding to $\approx 57h$: after frame $n \circ 170$ a net increase in the level of fluorescence has been detected both on eYFP and on mCherry. This effect might be related to the stresses cells have been exposed to, among which we can mention transfection and light excitation [25]. In addition, we should consider that after having been transfected cells live in the same culture medium for all the duration of the experiment. It is very likely that at a certain time cells have no longer nutrients. Extracted data have been arranged in two matrices, one for eYFP values and one for mCherry ones, where each row represents a specific cell and each column refers to a certain frame, thus to a particular time. As the system should be investigated at the "stationary state", i.e. after it has stabilized, let's take frame $n^\circ 100$ as a reference, which corresponds to a time $t = 206400s \approx 57h$. At this time, we look at the scatter plots of mCherry as a function of eYFP to locate the expected threshold [15]. Data have been binned according to their eYFP values and corresponding mCherry values for varying number of miRNA binding sites have been plot to spot the different trends. Scatter plots are shown in Figure 3.43. In order to see clearly the non linearity appearance when increasing miRNA binding sites, fit of scatter plots have been plot on the same graph 3.44. It is clear that only in absence of miRNA binding site mCherry is in a linear relation with eYFP. Now we consider three regions for eYFP values, which are respectively:

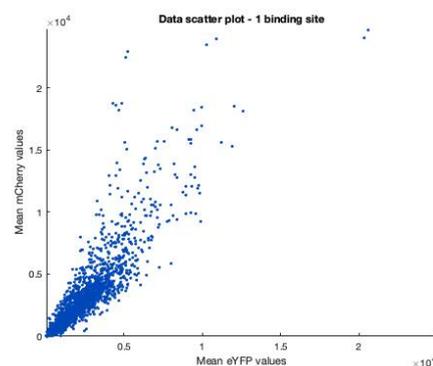
- $range_1$: under threshold with $(0 < eYFP < 1200)$ a.u.;
- $range_2$: at the threshold with $(1200 < eYFP < 2500)$ a.u.;
- $range_3$: over threshold with $(eYFP > 2500)$ a.u.

Let's investigate the aspect of mCherry abundance distributions corresponding to those eYFP regions. Distributions for all binding sites can be seen in Figure 3.45, where in each plot three distributions corresponding to the three regions have been shown. By looking at mCherry distributions for $range_1$ one can notice that they are all exponential. For $range_3$, instead, all mCherry values are Gaussian distributed. The most insightful case is obtain in $range_2$, at the threshold, where mCherry distributions for different binding sites behave differently. mCherry distributions with 0 binding sites (see Figure 3.45a) and 1 binding site (see Figure 3.45b) are

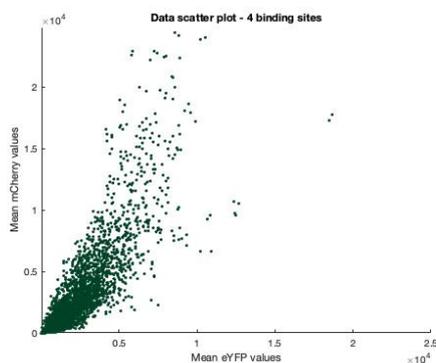
distributed according to a Gaussian distribution. In presence of 4 miRNA binding sites (see Figure 3.45c), mCherry distribution presents a peak in correspondence of the expressed state and a flat region in correspondence of the repressed state. The case of 7 binding sites (see Figure 3.45d) is the most interesting one in that a bimodal distribution is clearly visible: the expressed state has a larger value than the repressed one, namely there are much more free targets than free miRNAs.



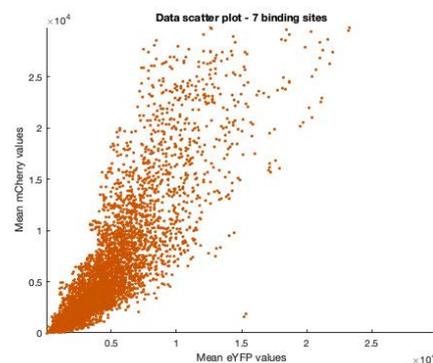
(a) Scatter plot of mean mCherry values versus mean eYFP value in absence of miRNA binding sites.



(b) Scatter plot of mean mCherry values versus mean eYFP value in presence of one miRNA binding site.



(c) Scatter plot of mean mCherry values versus mean eYFP value in presence of four miRNA binding sites.



(d) Scatter plot of mean mCherry values versus mean eYFP value in presence of seven miRNA binding sites.

Figure 3.43: Scatter plots of mean mCherry values versus mean eYFP values for varying number of binding sites. In going from 0 to 7 binding sites, miRNA-mRNA interaction strength is increasing resulting in a non-linearity.

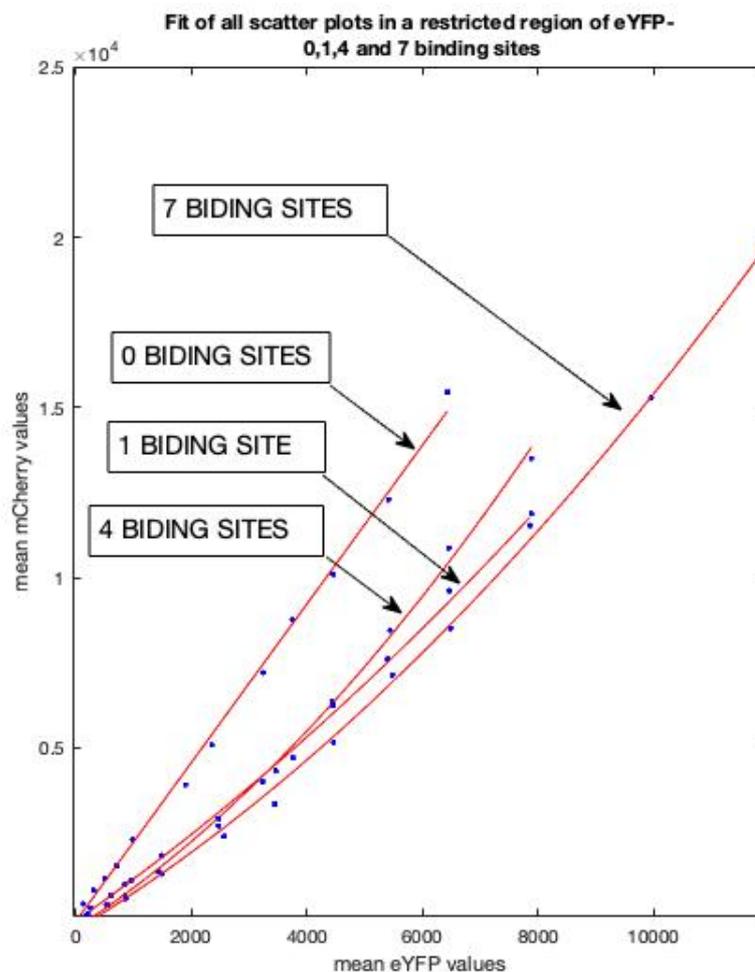
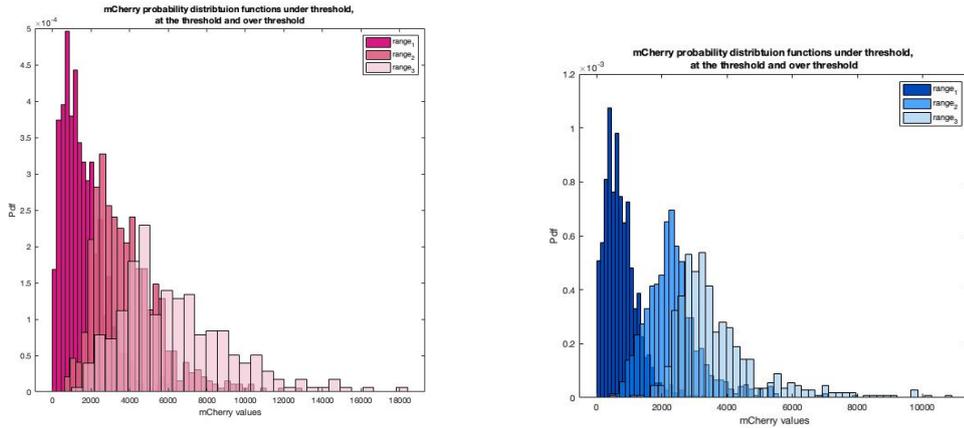


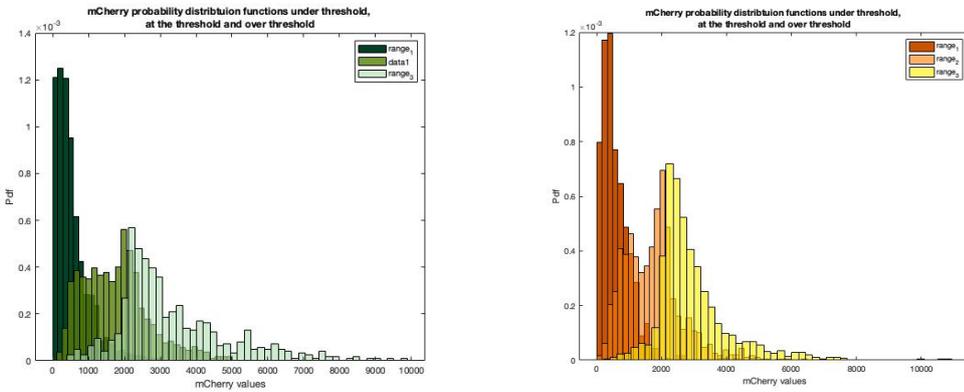
Figure 3.44: Figure showing the non linear behavior of mCherry as a function of eYFP for increasing number of binding sites.

3.3.2 Permanence times

In the previous section, a bimodal distribution for mCherry has been found in $range_2$ of eYFP for 7 binding sites while an emerging mCherry bimodal distributions is appearing in the case of 4 binding sites. We would like to estimate permanence times of the cells in the two states. In order to compute permanence times, we have to fit 7-binding sites bimodal distribution to extract the information about the means and the standard deviations. The choice to fit 7 binding sites bimodal distribution is a choice of convenience: bimodal distribution is much more well defined with respect to the 4 binding sites case. Means and standard deviation



(a) mCherry distribution in absence of miRNA binding sites. (b) mCherry distribution in presence of one miRNA binding sites.



(c) mCherry distribution in presence of four miRNA binding sites. (d) mCherry distribution in presence of seven miRNA binding sites

Figure 3.45: mCherry distributions under, over and at the threshold for the cases of 0,1,4 and 7 binding sites. The higher the number of binding sites, the stronger the interaction strength between miRNA and target. A bimodal distribution can be seen in 7 binding sites picture.

values will be useful to define which is the repressed state and which is the expressed one. Having determined parameters characterizing the repressed and the expressed states we can proceed to compute permanence times.

Let's begin to consider mCherry probability distribution function for 7 binding sites in $range_2$. By looking at the form of the distribution, it seems reasonable to fit it with a mixture of two Gaussian distributions. The result of the fitting is shown in Figure 3.46. As for theoretical simulations, the fit has the only scope of finding the values of means and standard deviations. It has been found that $\mu_1 = (941 \pm 130)$

a.u., $\sigma_1 = (435 \pm 129)$ a.u., $\mu_2 = (2017 \pm 82)$ a.u. and $\sigma_2 = (328 \pm 71)$ a.u.. With the results of this fit, we define as repressed the state formed by all values $x_1 < \mu_1 + \sigma_1$ and as expressed the state formed by all values $x_2 > \mu_2 - 1.5 \cdot \sigma_2$. This definition for the two states will be applied both to 4-binding sites distribution and to 7-binding sites one. We want the states to be as large as possible in order to consider the highest number of cells but we do not want fluctuations to falsify our results (there must be a certain separation between the two state).

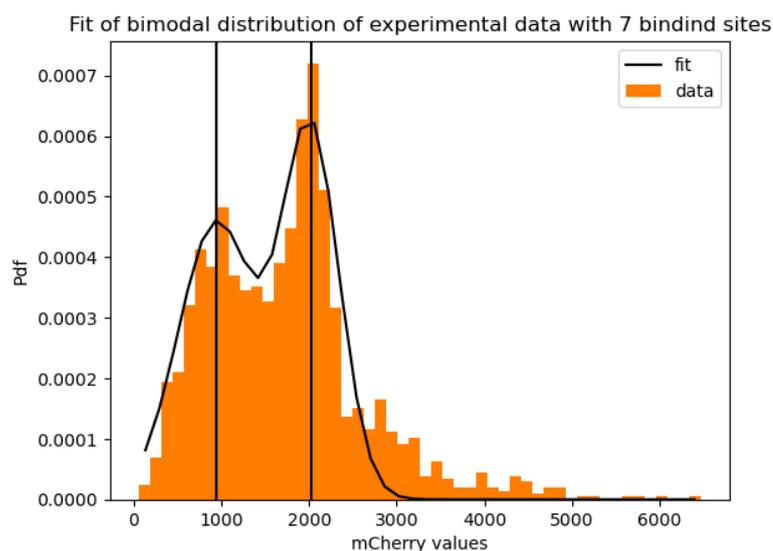


Figure 3.46: Fit of 7 binding sites bimodal distribution using the combination of two Gaussian distributions.

In Figure 3.47 we can see mCherry probability distribution functions for the cases of 4 and 7 binding sites, where the leftmost vertical lines are the right limit of the repressed state while the rightmost one are the left limit of the expressed state.

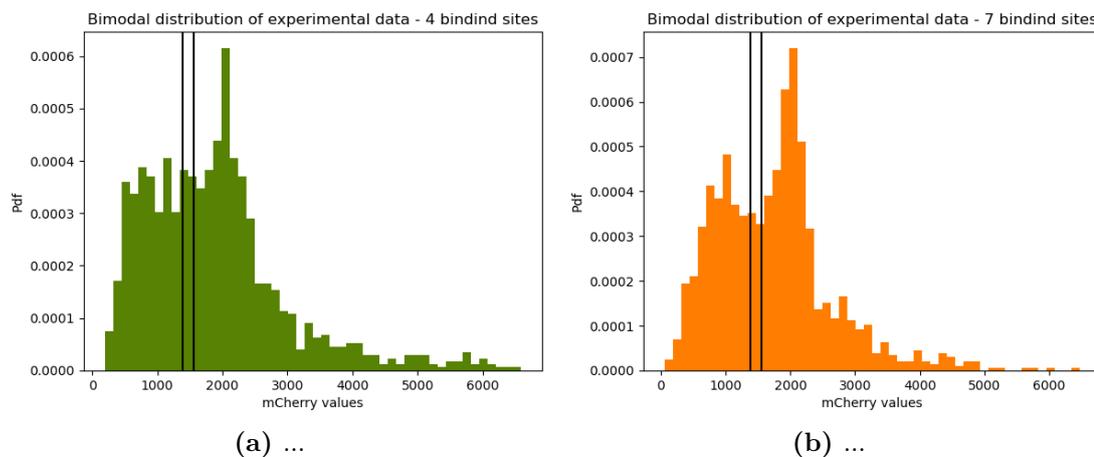


Figure 3.47: mCherry probability distribution functions for the cases of 4 (green distribution) and 7 (orange distribution) binding sites. Vertical lines represent the end and the beginning of repressed and expressed states respectively.

Now that we have both distributions and definitions of repressed and expressed states, we have to look at mCherry values (over time) of all those cells having eYFP in $range_2$ (see Figure 3.48). On these entire trajectories, from frame $n^{\circ}1$ to frame $n^{\circ}170$, the same function used in Section 3.1.3 will be applied. Again, the choice to consider the entire trajectory is a convenience choice: we need as much cells permanence times as possible to have a distribution. One may argue that the system we are looking at should be at the "steady state". We can't say for sure that the system is at the "steady state" but by looking at cells trajectories after 24h from transfection, we can see that they are quite stable. In the future experiments will be repeated and it might be interesting to start filming cells at earlier times and look for a "transient". The resulting permanence times are shown in Figure 3.49 and Figure 3.50.

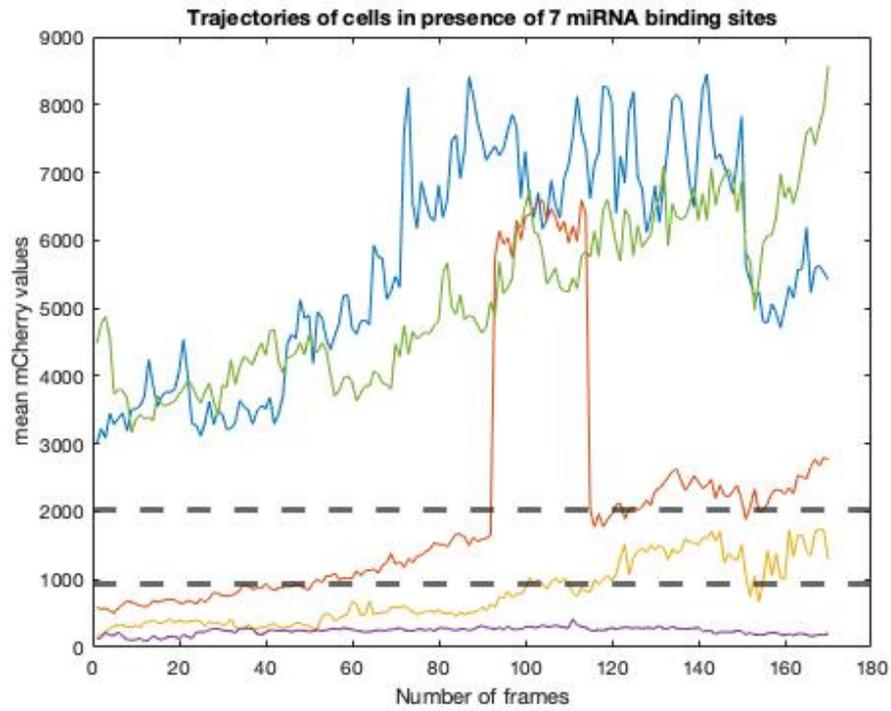


Figure 3.48: Cells trajectories in presence of 7 miRNA binding sites. The dotted lines represent the two means extracted from the fit of the bimodal distribution in Figure 3.46

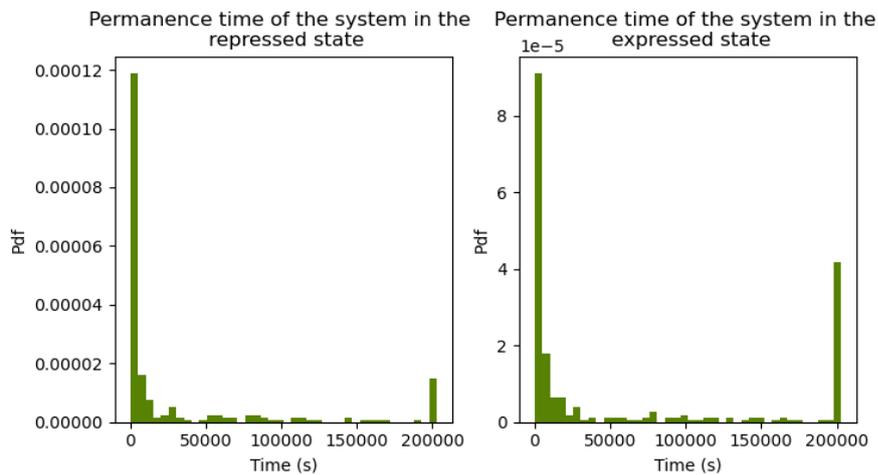


Figure 3.49: Cells permanence times in case of 4 miRNA binding sites. Both repressed and expressed states show peaks on long times.

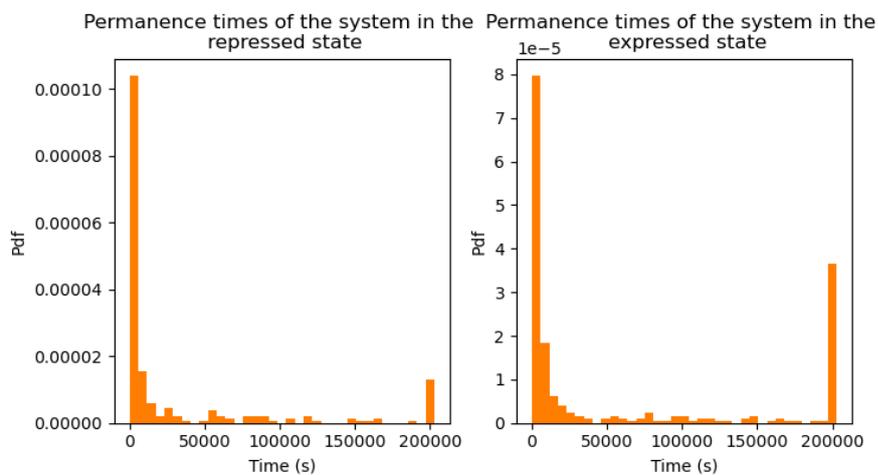


Figure 3.50: Cells permanence times in case of 7 miRNA binding sites. Both repressed and expressed states show peaks on long times.

These distributions show peak on long times, suggesting that extrinsic noise prevails with respect to the intrinsic one. In fact, we are looking at a system with a consistent number of cells and we said that intrinsic noise decreases with increasing system size [3]. Moreover, from a comparison with theoretical results, we can appreciate the same behavior of theoretical and experimental population distribution.

Chapter 4

Discussion and conclusions

MicroRNAs exert important regulatory functions on their target genes [13] and their aberrant expression may lead to several diseases [26]. In previous chapters, we have studied two models for miRNA-driven inhibition [8] taking into account the presence of intrinsic and extrinsic noise in the system. We have seen that intrinsic noise, related to fluctuation of biochemical reactions [19], is able to induce a bimodality in cell population expression only if interaction between miRNA and target is strong enough [8]. Extrinsic noise, instead, is related to environmental fluctuations [19]: here, it has been introduced as a noise on miRNA transcription rate. On the contrary of intrinsic noise, extrinsic noise is capable of inducing a bimodal population distribution for a wider range of parameters. The kind of bimodality induced by these two different noises is extremely different: intrinsic noise-induced bimodality is a single cell effect while extrinsic noise-induced bimodality is a population effect [8]. In this work we have seen that a single-cell trajectory is able to recapitulate the behavior of the entire population thanks to the ergodicity of the system under study. The key idea of this work was to distinguish intrinsic noise-induced bimodality from extrinsic noise-induced bimodality by looking at cells permanence times in the repressed and expressed states of the system. We have found out that cells belonging to a bimodal population induced by intrinsic noise have permanence times in the two systems distributed according to a decreasing exponential. This is due to the analogy of this bimodal with a double potential well. Cells belonging to an extrinsic noise-induced bimodal population distribution are characterized by the appearance of peaks on long cells permanence times in the two states of the system. In fact, in bimodal population distributions given by extrinsic noise we are looking at cells with heterogeneous values of miRNA transcription rate k_S . MiRNA transcription rate k_S is involved in the location of the threshold: in fact, protein expression values depends on the concentration of miRNAs and targets. Having different values of k_S in our cell population may imply the existence of some cells that never exit the repressed or the expressed

state. Thus, when computing cells permanence times in the two states, these cells will have a permanence time related to the state they are in which is equal to the observation time.

Starting from this thesis work, we could ask ourselves other interesting questions: what does it happen suddenly after transfection? Do our cells show a transient? Which are the cell cycle phases for miRNA in the repressed and expressed state? These questions could be a starting point for new research works on this fascinating and complex class that is microRNAs.

Bibliography

- [1] Erwin Schrödinger. *What is Life? The Physical Aspect of the Living Cell*. Cambridge University Press, 1944 (cit. on p. 1).
- [2] Fazli Wahid, Adeeb Shehzad, Taous Khan, and You Young Kim. «MicroRNAs: Synthesis, mechanism, function, and recent clinical trials». In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1803.11 (2010), pp. 1231–1243. ISSN: 0167-4889. DOI: <https://doi.org/10.1016/j.bbamcr.2010.06.013>. URL: <https://www.sciencedirect.com/science/article/pii/S0167488910001837> (cit. on p. 1).
- [3] Vahid Shahrezaei and Peter S. Swain. «Analytical distributions for stochastic gene expression». In: *PNAS* (2008). DOI: 10.1073/pnas.0803850105 (cit. on pp. 1, 6, 13, 54).
- [4] Nasser Bashkeel, Theodore J. Perkins, Mads Kærn, and Jonathan M. Lee. «Human gene expression variability and its dependence on methylation and aging». In: *BMC Genomics* 20.1 (Dec. 2019), p. 941. ISSN: 1471-2164. DOI: 10.1186/s12864-019-6308-7. URL: <https://doi.org/10.1186/s12864-019-6308-7> (cit. on p. 1).
- [5] Arjun Raj, Scott A. Rifkin, Erik Andersen, and Alexander van Oudenaarden. «Variability in gene expression underlies incomplete penetrance». In: *Nature* 463.7283 (Feb. 2010), pp. 913–918. DOI: 10.1038/nature08781. URL: <https://doi.org/10.1038/nature08781> (cit. on p. 1).
- [6] Jian Liu, Jean-Marie François, and Jean-Pascal Capp. «Gene Expression Noise Produces Cell-to-Cell Heterogeneity in Eukaryotic Homologous Recombination Rate». In: *Frontiers in Genetics* 10 (2019), p. 475. ISSN: 1664-8021. DOI: 10.3389/fgene.2019.00475. URL: <https://www.frontiersin.org/article/10.3389/fgene.2019.00475> (cit. on p. 1).
- [7] S. Garg and P. A. Sharp. «Single-cell variability guided by microRNAs». In: *Science* 352.6292 (June 2016), pp. 1390–1391. DOI: 10.1126/science.aag1097. URL: <http://dx.doi.org/10.1126/science.aag1097> (cit. on p. 1).

- [8] Silvia Grigolon Marco Del Giudice Stefano Bo and Carla Bosia. «On the role of extrinsic noise in microRNA mediated bimodal gene expression». In: *PNAS* (2018). DOI: <https://doi.org/10.1073/pnas.1008965107> (cit. on pp. 1, 5, 7, 13, 14, 38, 46, 55).
- [9] Marcin Tabaka Anna Ochab-Marcinek. «Bimodal gene expression in non cooperative regulatory systems». In: *PLoS Computational Biology* (2010). DOI: <https://doi.org/10.1371/journal.pcbi.1006063> (cit. on p. 1).
- [10] Ralph Leo Johan Meuwissen Kemal Ugur Tüfekci Meryem Gülfem Öner and Sermin Genç. «The Role of MicroRNAs in Human Diseases». In: *MicroRNA Biology and Computational Analysis Volume 1107* (2014). DOI: 10.1007/978-1-62703-748-8_3 (cit. on pp. 1, 3).
- [11] Daniel T Gillespie. «A general method for numerically simulating the stochastic time evolution of coupled chemical reactions». In: *Journal of Computational Physics* 22.4 (1976), pp. 403–434. ISSN: 0021-9991. DOI: [https://doi.org/10.1016/0021-9991\(76\)90041-3](https://doi.org/10.1016/0021-9991(76)90041-3). URL: <https://www.sciencedirect.com/science/article/pii/0021999176900413> (cit. on pp. 1, 15).
- [12] M. Bhaskaran and M. Mohan. «MicroRNAs: History, Biogenesis, and Their Evolving Role in Animal Development and Disease». In: *Vet Pathol* (2013). DOI: 10.1177/0300985813502820 (cit. on p. 2).
- [13] Zayed Y O'Brien J Hayder H and Peng C. «Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation.» In: *Front. Endocrinol.* 9:402. (2018). DOI: 10.3389/fendo.2018.00402 (cit. on pp. 2–4, 55).
- [14] Chanhyun Park Ahmed Maher Abdelfattah and Michael Y. Choi. «Update on non-canonical microRNAs». In: *Biomol Concepts* (2015). DOI: 10.1515/bmc-2014-0012 (cit. on p. 3).
- [15] Shankar Mukherji, Margaret S Ebert, Grace X Y Zheng, John S Tsang, Phillip A Sharp, and Alexander van Oudenaarden. «MicroRNAs can generate thresholds in target gene expression». In: *nature genetics* (2011). DOI: 10.1038/ng.905 (cit. on pp. 4, 5, 8, 9, 17, 47).
- [16] Nicolas E. Buchler and Matthieu Louis. «Molecular Titration and Ultrasensitivity in Regulatory Networks». In: *Journal of Molecular Biology* 384.5 (2008), pp. 1106–1119. ISSN: 0022-2836. DOI: <https://doi.org/10.1016/j.jmb.2008.09.079>. URL: <https://www.sciencedirect.com/science/article/pii/S0022283608012485> (cit. on p. 4).

- [17] A. Townshend. «TITRIMETRY | Overview». In: *Encyclopedia of Analytical Science (Second Edition)*. Ed. by Paul Worsfold, Alan Townshend, and Colin Poole. Second Edition. Oxford: Elsevier, 2005, pp. 105–113. ISBN: 978-0-12-369397-6. DOI: <https://doi.org/10.1016/B0-12-369397-7/00625-7>. URL: <https://www.sciencedirect.com/science/article/pii/B0123693977006257> (cit. on p. 5).
- [18] Wei Wu. «MicroRNA, Noise, and Gene Expression Regulation». In: *MicroRNA and Cancer*. Springer New York, Oct. 2017, pp. 91–96. DOI: 10.1007/978-1-4939-7435-1_7. URL: https://doi.org/10.1007/978-1-4939-7435-1_7 (cit. on p. 6).
- [19] William J. Blake Mads Kaern Timothy C. Elston and James J. Collins. «STOCHASTICITY IN GENE EXPRESSION: FROM THEORIES TO PHENOTYPES». In: *Nature Publishing Group* (2005). DOI: 10.1038/nrg1615 (cit. on pp. 6, 55).
- [20] Thomas B. Kepler and Timothy C. Elston. «Stochasticity in Transcriptional Regulation: Origins, Consequences, and Mathematical Representations». In: *Biophysical Journal* 81.6 (2001), pp. 3116–3136. ISSN: 0006-3495. DOI: [https://doi.org/10.1016/S0006-3495\(01\)75949-8](https://doi.org/10.1016/S0006-3495(01)75949-8). URL: <https://www.sciencedirect.com/science/article/pii/S0006349501759498> (cit. on p. 6).
- [21] Zoe K. Sanchez Masaki Shintani and Kazuhide Kimbara. «Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy». In: *Frontiers in Microbiology* (2015). DOI: <https://doi.org/10.3389/fmicb.2015.00242> (cit. on p. 8).
- [22] *LIPSI | Cell Screening | Products | Nikon Europe B.V.* URL: https://www.microscope.healthcare.nikon.com/en_EU/products/cell-screening/lipsi (cit. on p. 8).
- [23] Yingyi Zhang, Ting Han, Guo Wei, and Yajie Wang. «Inhibition of microRNA-17/20a suppresses cell proliferation in gastric cancer by modulating UBE2C expression». In: *Oncol Rep* 33.5 (May 2015), pp. 2529–2536. DOI: 10.3892/or.2015.3835. URL: <https://doi.org/10.3892/or.2015.3835> (cit. on p. 13).
- [24] Ramesh S.V. Teegavarapu. *Methods for Analysis of Trends and Changes in Hydroclimatological Time-Series*. Elsevier, 2019 (cit. on p. 21).
- [25] Jack Davis. *Avoiding Cell Death in Fluorescence Microscopy*. Feb. 2021. URL: <https://www.azolifesciences.com/article/Avoiding-Cell-Death-in-Fluorescence-Microscopy.aspx> (cit. on p. 47).

- [26] Ali M. Ardekani and Mozhgan Moslemi Naeini. «The Role of MicroRNAs in Human Diseases». eng. In: *Avicenna journal of medical biotechnology* 2.4 (Oct. 2010). PMC3558168[pmcid], pp. 161–179. ISSN: 2008-2835. URL: <https://pubmed.ncbi.nlm.nih.gov/23407304> (cit. on p. 55).