ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Physics (30^{th}cycle)

Molecular sequestration as a way to control variability and induce crosstalk

from theory to experiment

By

Marco Del Giudice

******

Supervisor(s):
Dr. Carla Bosia, Supervisor

Doctoral Examination Committee:
Prof. Roberto Mulet , Referee, University of Havana
Dr. Matteo Osella, Referee, University of Torino
Prof. Michele Caselle, University of Torino
Prof. Arianna Montorsi, Politecnico di Torino
Prof. Luca Dall'Asta, Politecnico di Torino

Politecnico di Torino
2018

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Marco Del Giudice
2018

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

To my best friend

Acknowledgements

I would like to thank Carla Bosia for her constant helpfulness and incitement, she is the best supervisor a PhD student can ask. Thanks to Stefano Bo and Silvia Grigolon for having shared my research efforts with patience and essential advice. Thanks to Chiara Enrico Bena for having shared many hours of work, making them lighter. Thanks to Matteo Osella, Roberto Mulet, Andrea De Martino, Carlo Campa, Andrea Pagnani, Riccardo Zecchina, Francesco Sgro, Emilia Turco.

Finally, grateful thanks to my mother and my father for their support, advice and trust in me. Thanks to Giulia for being with me in the best and worst moments of my life, you always keep bringing me fun, happiness and love.

Abstract

The mechanism of molecular sequestration is involved in many biological processes, ranging from growth factors signalling to transcriptional and post-transcriptional regulation. This kind of dynamics involves two types of molecular species, targets and sequesters, that bind to form a complex. In the framework of mass-action law, key features of this kind of systems appear to be threshold-like profiles of the amounts of free molecules as a function of the parameters determining their possible maximum abundance. In biology, stochastic fluctuations, i.e. noise, play an undisputed role at the molecular level. Such noise can be usually divided into an intrinsic component, due to the probabilistic nature of biochemical processes, and an extrinsic one, related to the coupling with the variability of the environment in which reactions take place. Several studies highlighted the relevance of the variability induced by extrinsic fluctuations in shaping cell decision making and differentiation in molecular networks. Bimodal distributions of gene expression levels are a common feature of this kind of processes. Indeed, the two modes of the distribution often indicate the presence of two different physiological states of the system. In this thesis, we investigate the consequences of the introduction of a source of extrinsic noise onto a system governed by molecular sequestration, focussing on the appearance of bimodal distributions. To do that, we first study a minimal stochastic model of molecular sequestration and introduce extrinsic noise through a fluctuating parameter. In this framework, we analytically show how bimodal distributions can appear and characterize them as a result of noise filtering mediated by the threshold response. We then investigate the behavior of the correlations between targets of the same sequester and show how extrinsic fluctuations can induce a positive correlation that counterbalances the negative one due to competition. Given these results, we move to investigate the appearance of bimodal phenotypes in the context of microRNA (miRNA)-mediated gene regulation. MiRNAs are small noncoding RNA molecules that downregulate the expression of their target mRNAs.

The interaction between miRNAs and targets is based on molecular sequestration and threshold-like responses are a known feature of this system. Recent studies suggested that miRNAs can be involved in the appearance of bimodal expression distributions of their target genes. To investigate this phenomenon, we characterize the system through an analytic and numerical approach and introduce extrinsic noise as a fluctuating miRNA transcription rate. We observe that bimodal distributions of target expression can appear for a wide range of parameters in presence of extrinsic noise. Furthermore, we show how the bimodal shape of the distribution can be tuned by the interplay between different target mRNAs competing for a common miRNA. In conclusion of this thesis we present some synthetic-biology experiments that are aimed at studying the role of extrinsic fluctuations in the appearance of bimodal distributions in the context of miRNA-mediated regulation.

Contents

List of Figures

List of Tables

Chapter 1

State of the art and thesis outline

Different kinds of regulatory mechanisms and related network motifs are involved in gene expression. Transcriptional regulation is among the best characterized: one or more proteins, called transcription factors, bind to the promoter of a target gene and enhance or impede its transcription into RNA. The combinatorial and cooperative nature of this regulation gives rise to an enormous and heterogeneous network of interactions between genes. In recent years, a completely new layer of gene regulation, coupled to the transcriptional one, was discovered and characterized [11–15]. This kind of regulation occurs at the post-transcriptional level and involves interactions among transcripts. Post-transcriptional regulation is mediated by 22 nucleotides long non-coding RNA molecules called microRNAs (miRNAs) [16–19]. MiRNAs recognize and bind to their target RNAs thanks to the complementarity between their sequences. The miRNA molecule is loaded on a protein complex called RNA-induced silencing complex (RISC). When the miRNA binds to the target RNA, RISC promotes its degradation or reduces its translation into protein, leading to a downregulation of the expression of the corresponding gene. For this reason, miRNAs are considered as post-transcriptional down-regulators of gene expression.

Several studies [4, 7, 8, 20–23] suggest that miRNA-target interaction is based on a mechanism of molecular sequestration. The idea is that, when a miRNA binds to its target, they form a complex which prevents further interaction with other molecules until dissociation. This mechanism of "titrative" interaction has already been well studied in other contexts. Examples are protein ubiquitination [24], growth factors signalling [25, 26], transcription factors sequestration [27, 28, 1],

the interaction between RNA polymerase and its sigma factors in bacteria [29] and bacterial persistence [30]. Titration has the property to induce threshold-like responses in the mean amount of molecules as a function of some key parameter as, for example, their production rate. Indeed, if the amount of sequester (e.g. the miRNA) is larger than the amount of target (e.g. the mRNA), most of the target molecules are bound to a sequester and their free amount is close to zero (the repressed regime). On the contrary, when the amount of target is larger than the amount of sequester, almost all the sequester molecules are bound, but some target molecules are always free (unrepressed regime). In this case, the amount of free targets increases linearly with their total amount. The combination of these repressed and unrepressed regimes originates the threshold behavior, with the threshold located approximately at the equimolarity point between target and sequester. In proximity to the threshold, a small change in the total amount of target can lead to a big fold change in its free amount. This property is called ultrasensitivity [1].

The regulation operated by miRNAs is combinatorial: one species of miRNA can regulate several different targets and one target is regulated by many different miRNAs [31–33]. The underlying interaction network is based on the competition between targets for binding to the same pool of miRNAs and such competition is a source of crosstalk among them. In such a network, the overexpression of one of the targets, achieved for example by increasing its transcription rate, induces a derepression of the other targets that share the same pool of miRNAs and leads to a net increase of their free molecule share. This happens because the transcripts of the overexpressed target act as a "sponge" for the miRNAs, sequestering them and consequently reducing the level of repression on other targets [34, 35]. The importance of this crosstalk, also called Competing Endogenous RNA (ceRNA) effect [36–38], was demonstrated in different physiological and pathological conditions [38–40] and represents a key element that has to be taken into account in order to understand the whole interaction network among genes, including the ones originating non-coding transcripts.

Previous works [20, 7] focused on the ceRNA effect from a theoretical point of view. Mathematical modeling, based on a stochastic approach, has been useful both in describing crosstalk among targets and in predicting the coupling of their fluctuations. The model, that takes into account the processes of miRNA and target transcription, degradation and interaction, is associated to a master equation whose solutions can be found at the steady state through analytic approximations. These analytic results were

experimentally validated by exploiting a synthetic system composed of two different genes having binding sites for the same miRNA [8]. The two genes coded for different fluorescent proteins and were inserted in mammalian cells that endogenously express their common miRNA. Measurements were conducted using flow cytometry, a technique that allows to measure fluorescence in single cells obtaining then a proxy for the targets' expression. The results, besides confirming the model's predictions, revealed regions of optimal crosstalk where the relative fluctuations of targets can be potentially synchronized. Furthermore, these experiments were also a source of new interesting interrogatives, related to the appearance of distinct phenotypes in the distribution of target expression in case of strong miRNA repression.

A key feature of molecular networks is their intrinsic stochasticity [41, 42]. This is due to the inherent randomness of chemical reactions which are governed by thermal fluctuations. This intrinsic noise is negligible for macroscopic systems, but it plays a crucial role in biological systems at the cell and sub-cell level, where the number of molecules involved in the reactions is small. These systems can be influenced by other sources of noise, different from the intrinsic one [43–48]. These sources of extrinsic noise can be fluctuations in the external environment which affect in a different way identical systems. At the cell level, extrinsic noise is composed of all the fluctuations that affect the space in which biochemical reactions occur. Variations, across identical cells, of the amount of cellular components and molecular machineries, for example ribosomes, can be considered as extrinsic noise, as well as gradients of signaling molecules in the external environment. The presence of extrinsic noise, together with the intrinsic one, results in phenotypic variability across a population of identical cells. This phenotypic variability is sometimes organized in bimodal distributions, which highlight the existence of two distinct phenotypes. Bimodal expression distributions are often encountered in processes of cell decision-making and cell differentiation, but they can also point out the presence of an ill phenotype, together with the healthy one [49–54].

As mentioned above, miRNA-mediated gene regulation can give rise to bimodal distributions. Given the strong involvement of miRNAs in processes of cell differentiation and in the preservation of homeostasis, their ability to produce bimodal distributions can be a crucial element of their regulatory role. A core part of this thesis is dedicated to the quantitative understanding of how miRNAs can give rise to bimodal distributions and to the role that can be played in this context by some extrinsic noise.

In chapter 2 of this thesis, we briefly review the mechanism of molecular sequestration and use it to explain the process of miRNA-target interaction. We then present some properties of this kind of regulation, focusing in particular on the role of miRNAs as "micromanagers of gene expression" and noise buffers. Consequently, we introduce the ceRNA effect and describe a mathematical model that reproduces the phenomenon and allows to study the stochastic properties of the system. Finally, we describe in detail the experimental setup used to validate the theoretical predictions of models of miRNA regulation and ceRNA effect.

Chapter 3 is dedicated to the study of how miRNA regulation can give rise to bimodal distributions, in relation to the presence of an extrinsic source of noise. In this chapter, we first introduce a minimal model of molecular sequestration and show analytically that the addition of extrinsic noise can lead to bimodal distributions. We then study a more complex model describing in more detail the process of miRNA-mediated gene regulation. Through analytic and numerical approaches, we characterize the presence of bimodal distributions in this system, highlighting the difference between the single-cell and the population levels.

In chapter 4, we extend the analysis of the previous chapter to a system of two targets competing for the same sequester. Following the outline of chapter 2, we first extend the minimal model with the addition of one target species, consequently, we study a model of miRNA-mediated gene regulation with two genes competing for the same miRNA. Both the models are studied in presence and absence of a Gaussian source of extrinsic noise. Thanks to the analytic control on the minimal model, we characterize in detail the behavior of the correlations induced by the competition between the targets. Finally, we use the model of miRNA-target interaction to obtain some insights on how bimodal distributions can be tuned by the crosstalk between the targets.

In chapter 5, we analyze some slightly modified variations of the minimal model of molecular sequestration, in order to build the bridge towards the model of miRNA-mediated gene regulation. The addition of some elements that make the minimal model more similar to the complete one leads to non-trivial results, like the disappearance of the threshold behavior.

Finally, chapter 6 is dedicated to the description of our experimental activity. We are currently following two lines of research aimed at studying different aspects of miRNA-mediated gene regulation.

The aim of the first experiment is the in vitro exploration of the role played by the extrinsic noise in the rise of bimodal distributions of miRNA targets. It is based on the study of the expression distributions of a synthetic target of an endogenous miRNA. The heterogeneity of the cells with respect to cell cycle is assumed to be an extrinsic source of noise and single-cell measurements are performed at the flow cytometer.

With the second experiment, we want to investigate the behavior of the correlation between two miRNA targets beyond the steady state. In order to investigate the time behavior of the system, we follow an experimental approach based on fluorescence time-lapse microscopy. With this technique, we study temporal correlations between two synthetic miRNA targets in single cells. Furthermore, we plan to use this experimental setup to study the response of the system to an external perturbation, like the sudden switch on or off of one of the two synthetic targets.

The work described in this thesis led to two theoretical papers and was partially included in a book chapter on Computational Biology of non-coding RNA in Biomedicine:

- **D. G. M.**, Bo S., Grigolon S., Bosia C., On the role of extrinsic noise in microRNA-mediated bimodal gene expression, Plos Computational Biology, 14:1-26, DOI 10.1371/journal.pcbi.1006063, 2018 [10]
- **D. G. M.**, Bosia C., Grigolon S., Bo S., Stochastic sequestration dynamics: a minimal model with extrinsic noise for bimodal distributions and competitors correlation, Scientific Reports, 8(1):10387, DOI 10.1038/s41598-018-28647-9, 2018 [9]
- Martirosyan A., **D. G. M.**, Enrico Bena C., Pagnani A., Bosia C., De Martino A., Kinetic modelling of competition and depletion of shared miRNAs by competing endogenous RNAs, Methods in Molecular Biology, under revision [55]

An additional paper on the experiments described in chapter 6 is in preparation.

The research activity conducted during this Ph.D. involved also the collaboration to other projects not described in this thesis. These collaborations led to the following publications:

- Campa C.C., Margaria J.P., Derle A., **D. G. M.**, De Santis M.C., Gozzelino L., Copperi F., Bosia C., Hirsch E., Rab11 activity and PtdIns(3)P turnover removes recycling cargo from endosomes, Nature Chemical Biology, DOI 10.1038/s41589-018-0086-4, 2018 [56]

- Enrico Bena C., **D. G. M.**, Gueudré T., Miotto M., Turco E., De Martino A., Bosia C., Inoculum-density dependent growth reveals inherent cooperative effects and stochasticity in cancer cell cultures, arXiv:1710.10978 [57]

Chapter 2

Overview on molecular sequestration and microRNA-mediated gene regulation

In this chapter we will introduce the properties of the process of molecular sequestration and we will relate it to post-transcriptional gene regulation mediated by microRNAs. The main feature of molecular sequestration is the possibility to obtain threshold responses and ultrasensitivity in absence of molecular cooperativity. This property is found also in gene networks regulated by microRNAs. Further studies in this context showed that this mechanism of interaction plays an essential role in gene regulation, by providing stability to specific phenotypes, buffering expression noise and inducing crosstalk between target genes.

2.1 Main properties of molecular sequestration

The process of molecular sequestration involves two molecular species: a titrating molecule (titrant) and a target. The titrant binds, with a given rate k_{on}, to the target molecule and forms a complex (heterodimer) which sequester the target from the environment, often inhibiting its activity. The complex can in turn dissociate with rate k_{off}, releasing both the titrant and the target again in the environment, where they are free to interact. This mechanism of interaction is common to proteins. Imagine for example a transcription factor (TF) whose activity is kept under control

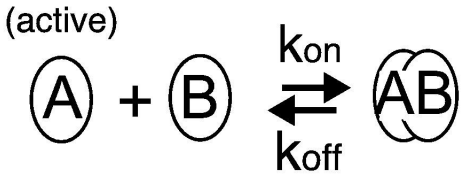

Fig. 2.1 Scheme of molecular sequestration for protein interaction. Adapted from [1].

by another protein. Such protein sequesters the TF by binding it in an inactive complex, see figure 2.1. This simple interaction network was analyzed in detail by Buchler and Louis in [1] with the help of mathematical modeling, based on a deterministic approach. In the following, we will summarize the most important results of this analysis.

Following the scheme of figure 2.1, we can easily define the dissociation constant of the network:

$$K_d = \frac{k_{off}}{k_{on}} = \frac{[A][B]}{[AB]},$$

where $[A]$, $[B]$ and $[AB]$ are the concentrations of species A, B and complex AB respectively. Additional equations that hold for this system are the conservation laws which state that the total concentration of A and B is constant ($[A_T] = [A] + [AB]$ and $[B_T] = [B] + [AB]$). The resulting system of equations can be solved, obtaining the equilibrium solution for the concentration of free A.

When $[B_T] > K_d$, the presence of the complex AB is favored against the presence of free A. In this condition, two regimes can be discriminated, depending on the total concentration of A present in the system. If $[A_T] < [B_T]$, almost all the molecules of the target, A, are bound to molecules of the titrant, B, and their free amount is close to zero. In particular, in this regime $[A] \approx [A_T] \cdot K_d/[B_T]$, and the system is more repressed for high values of the binding rate k_{on}. If $[A_T] > [B_T]$, the titrant is saturated by the target A, whose free concentration grows linearly with $[A_T]$: $[A] \approx [A_T] - [B_T]$. These two regimes give rise to a threshold-like response of $[A]$ as a function of $[A_T]$, with the threshold located in $[A_T] = [B_T]$. The threshold gets steeper as the dissociation constant K_d decreases, meaning that the binding rate k_{on} gets higher with respect to the unbinding rate k_{off}. This behavior makes the system ultrasensitive in proximity to the threshold. Indeed, in this region, a small

fold change in the total amount of A can result in a huge fold change in the amount of free A. The intensity of the response increases as $K_d/[B_T]$ decreases.

The typical context in which ultrasensitivity in molecular reactions can be found is the formation of protein homo-oligomers: protein complexes composed of few identical molecules. In these reactions, the affinity for binding can increase with the number of monomers already bound in a complex. This phenomenon, called molecular cooperativity [58–61], induces nonlinearities between the total concentration of the input monomer and the concentration of the output oligomer. These nonlinearities lead to ultrasensitivity.

Molecular sequestration represents a way to build up systems with high ultrasentitivity without cooperative binding. This mechanism of interaction is found in different contexts, ranging from protein-protein interaction to microRNA-mediated gene regulation. This last will be the subject of the next sections.

2.2 Molecular sequestration in microRNA-mediated gene regulation

In the following sections we will review some experimental and theoretical results regarding the process of microRNA-mediated gene regulation, focussing on its effects on phenotypic stabilization, expression noise buffering and crosstalk among transcripts.

2.2.1 Overview about microRNAs

MicroRNAs (miRNAs) are small non-coding RNA molecules that actively downregulate gene expression at the post-transcriptional level [3, 16–19, 62, 63].

MiRNAs were first identified as small temporal RNAs (stRNA) regulating timing of developmental events in Caenorhabditis Elegans [11–15]. "*The extension of miRNA-mediated regulation from being a curiosity in C. elegans to a potentially general mechanism for gene regulation began with the work of the Tuschl, Bartel and Ambros groups* [13–15]*, who collectively identified more than 100 novel miRNAs by cloning and sequencing endogenous small RNAs of 21–25 bp long from worms, flies and mammals.*" [64].

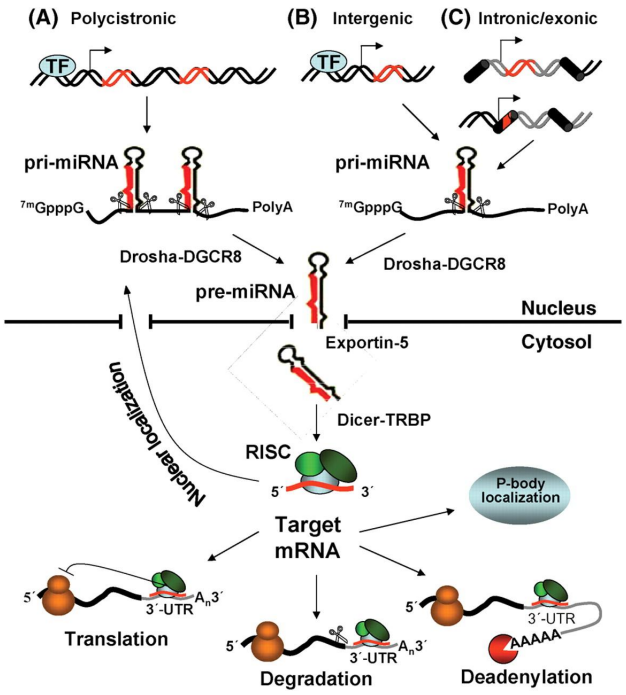

Fig. 2.2 MiRNA biogenesis. Adapted from [2].

Up to now more than 25000 miRNAs have been identified in animals, plants and viruses, about 2500 in human cells [65, 66]. MiRNAs, well conserved in animals and plants, play a role in a wide range of regulatory mechanisms involved in organogenesis and differentiation [67–69], apoptosis [70, 71] and cell proliferation [72–74] and their dysregulation can lead to important diseases, including cancer [65, 75–78].

In the following, we will briefly review miRNA biogenesis in animals and describe their mechanism of interaction with target mRNAs, see figure 2.2.

Genes coding for miRNAs can be found as autonomous units in intergenic regions of the genome [79–81], in introns (i.e. non-coding regions inside a gene that are removed after splicing) of both protein coding and non-coding genes [63, 82–85] and even in exons (i.e. regions coding for a portion of the mature RNA) of non-coding genes [82, 86–88]. MiRNAs transcribed antisense from exons of protein-coding genes exist, but a detailed characterization is still missing [89–91]. MiRNA genes can also form clusters regulated by a common promoter, creating a polycistronic transcription unit [92, 82]. The primary product of the transcription of a miRNA gene is a long transcript called pri-miRNA which contains a 5' cap structure and a

3' poly(A) tail. Pri-miRNAs are processed in the nucleus by the enzyme Drosha, originating $\sim$100 nucleotides long miRNA precursors with a hairpin structure, called pre-miRNAs. These precursors are transferred to the cytoplasm by the protein Exportin-5, where they undergo further processing by the Dicer enzyme. This enzyme cuts the hairpin loop of the pre-miRNA, leading to a double stranded RNA (dsRNA) duplex with imperfect match between the sequences. The duplex is then incorporated into a protein complex called RNA-induced silencing complex (RISC). Within the RISC, only one of the two strands of the miRNA is loaded, the other being released and degraded.

Once loaded in the RISC, the mirRNA is ready to exert its repressive action on its target mRNA, whose sequence is recognized through Watson–Crick base pairing. The complementarity of the two sequences is usually almost perfect in plants. Conversely, in animals, miRNAs recognize their targets through the nearly perfect pairing with a smaller region, about 6 nucleotides long, contained in the 5' end of the miRNA, the so called seed region [31]. Usually, the sequence of the target complementary to the seed region is located in the 3'-UnTranslated Region (3'UTR), but evidences of miRNAs binding the 5'UTR [93] and the coding sequence have also been found [65, 94, 95].

In plants, due to the high complementarity between the sequences of miRNA and target, the target is cleaved by the RISC. Differently, in animals, the incomplete base pairing usually does not lead to the target cleavage, but promotes translational repression and mRNA destabilization [3]. The sequestration of target molecules, possibly combined to an enhanced degradation, is then at the basis of miRNA-mediated gene regulation.

Computational models predict that more than the 60% of human genes is targeted by miRNAs [31, 96]. Moreover, miRNAs bind also to other non-coding transcripts, like long non-coding RNAs and pseudogenes [39, 97]. The action of the miRNAs is combinatorial: one miRNA can regulate several different targets and one target is regulated by many different miRNAs [31–33]. Altogether, these properties underline the fact that miRNAs are key players in gene regulation.

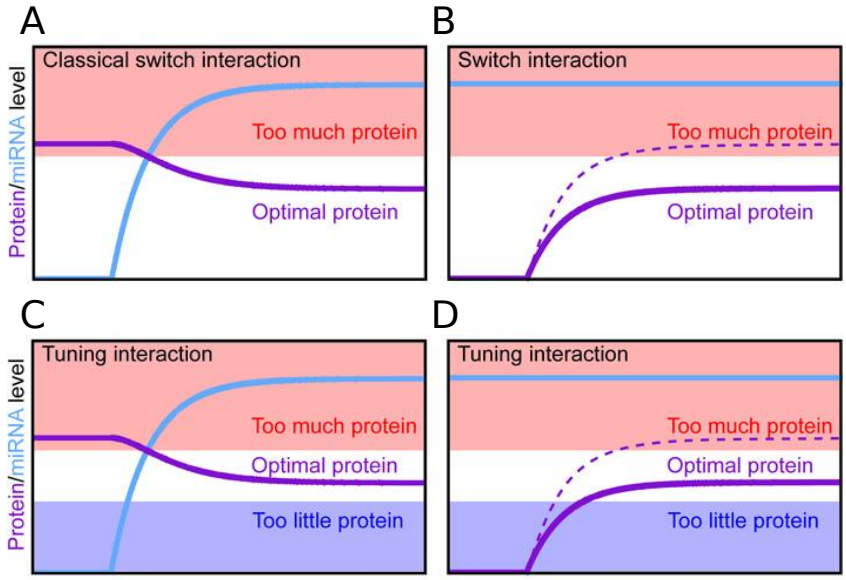

Fig. 2.3 MiRNA action as a switch or a fine tuner. Adapted from [3].

2.2.2 Theory: noise

The main role of miRNAs has been debated for long. In this section we introduce the most important interpretations and theoretical results on the action of miRNAs.

An important interpretation of the role played by miRNAs was introduced by the Bartel's group [98, 3]. The idea is that miRNAs can be considered "micromanagers of gene expression" [98] that fine tune the protein output of a gene, to maintain physiological expression levels. In this context, miRNAs can help to stabilize the protein expression in different cell phenotypes and can act as a switch in response to environmental or developmental stimuli. Indeed, cells belonging to different tissues tend to present anticorrelated expressions of miRNAs and relative target mRNAs [99–102]. This means that miRNAs can act as reinforcers and stabilizers of transcriptional patterns, sharpening differences in gene expression among different tissues or different developmental stages.

In figure 2.3, adapted from [3], the role of the miRNA as a switch or a fine tuner is described. The left panels represent the response of the target protein level to the sudden expression of the miRNA. The right panels represent the opposite situation, in which the transcription of the target is switched on in a cell in which the miRNA is already present. As can be seen, the miRNA can act as a switch between two

phenotypes (panel A), it can keep the protein amount below a threshold level that separates two phenotypes (panel B) and it can bring or keep the amount of proteins in an optimal region (panels C and D).

In [4] the miRNA-target interaction mechanism was investigated through both single-cell experiments exploiting a synthetic miRNA target and deterministic mathematical modeling. A result of this analysis was that the regulation mediated by the miRNA induces nonlinearities between the constitutive expression of the target and the amount of the target protein. In particular, below a certain threshold level of constitutive expression, the protein production is highly repressed, while, above the threshold, the protein amount increases linearly with the target expression level.

The threshold-like response observed in these data is compatible with a mechanism of molecular titration [1] and can be accurately described through mathematical modeling. In the model of Mukherji and coworkers, miRNAs that bind to target mRNAs impede their translation into proteins and enhance their degradation. The steepness of the threshold transition is related to the value of the binding rate: a higher value of the binding rate, which means a stronger miRNA-target affinity, corresponds to a steeper threshold, see figure 2.4 (B,D). The level of target expression corresponding to the threshold point, is instead set by the amount of miRNA present in the cell. As the number of miRNAs increases, the threshold point is shifted towards higher values of target expression, see figure 2.4 (C,E).

In the repressed region below threshold, the target molecules are significantly outnumbered by the miRNAs, almost no mRNAs are free to be translated and the protein level is very low. On the contrary, above threshold, free mRNAs are present and their amount increases linearly with the target expression level. As a consequence, proteins are translated and also their amount increases linearly with the target expression level. These results support the idea that miRNAs can act both as a switch, when the system is in the repressed regime, and as a fine tuner of the target expression, when the system is in the linear regime.

Open questions still remain about the effective role of miRNAs in cell physiology. The point that rises these questions is the fact that miRNAs are highly conserved and regulates thousands of genes, but their repressive effect on targets is usually weak [103, 104]. Moreover, knockouts experiments, in which a miRNA gene is silenced, very rarely produce phenotypic changes. In order to investigate these questions, it was hypothesized that miRNAs can have a role in controlling the noise level of

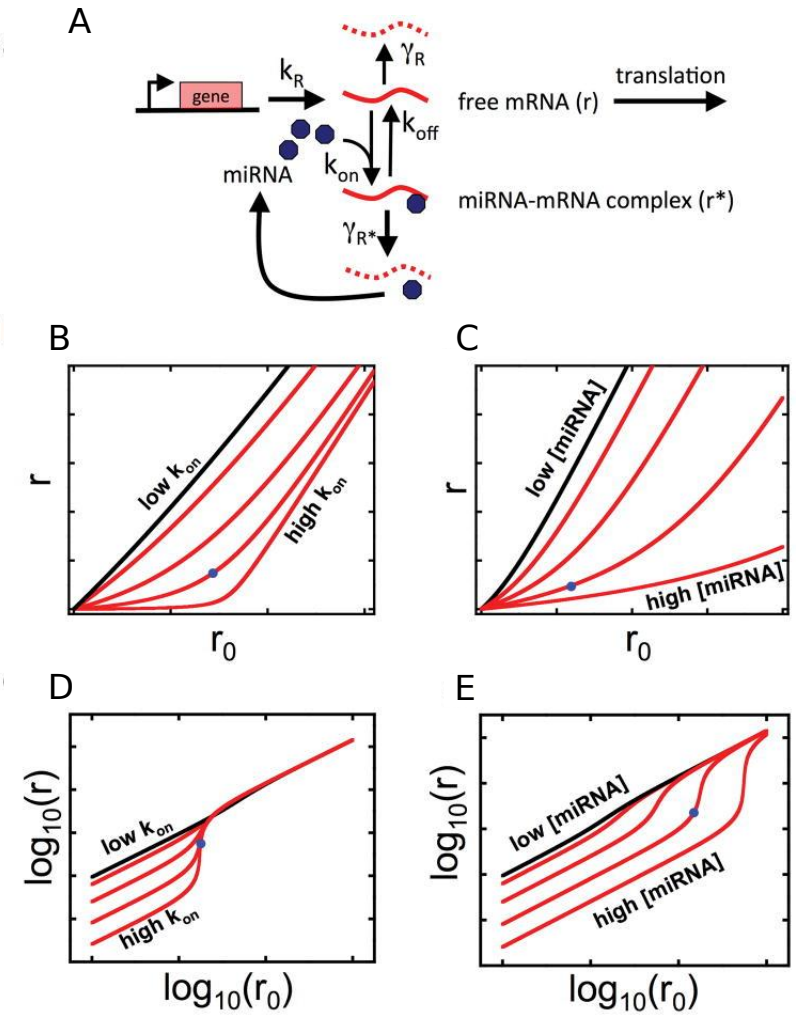

Fig. 2.4 (A) Model of miRNA-target interaction studied in [4]. The total concentration of miRNA is constant, the target mRNA is transcribed and degrades, both spontaneously and as a consequence of the interaction with the miRNA. (B-E) Steady state solutions of the model describing the concentration of free mRNA as a function of the transcriptional activity (concentration of mRNA in absence of miRNA regulation. (B,D) Solutions for different values of the binding rate. (C,E) Solutions for different concentrations of miRNA. Adapted from [4].

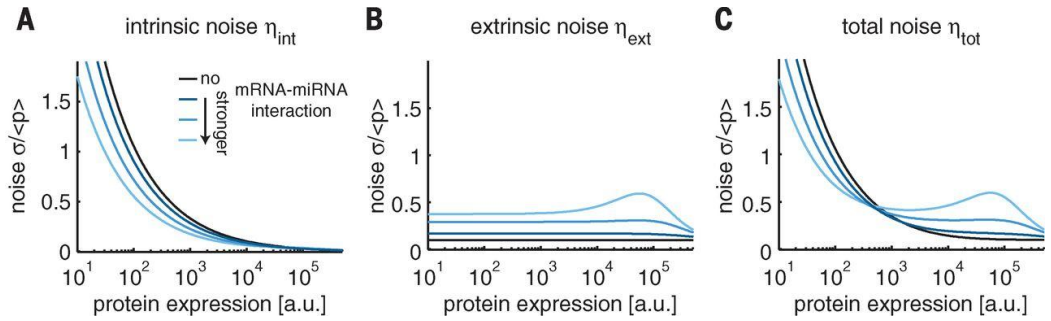

Fig. 2.5 Predictions of the model described in [5] for the noise of a miRNA-regulated gene. Regulation reduces intrinsic noise at low expression levels and increases extrinsic noise at high expression levels. Adapted from [5].

protein expression. The intrinsic component of the expression noise is mainly set by the variability of the transcription process and by low mRNA copy numbers [43].

Theoretical works have investigated the ability of post-transcriptional control to attenuate intrinsic noise [105]. Moreover, it was speculated that miRNAs can reduce the intrinsic component of expression noise with respect to the unregulated case. This is due to the combination of their repressive action, that enhances mRNA turnover, and the increased target transcription which is needed to achieve the same protein level of the unregulated case [99]. Nonetheless, the interaction with the miRNA is expected to constitute an additional source of noise for the system. This problem was recently addressed by Schmiedel and coworkers in [5]. In this work, the authors developed a mathematical model that accounted for both the intrinsic noise reduction and the additional extrinsic fluctuations introduced by the miRNA interaction. In addition to that, experiments with synthetic targets of an endogenous mRNA were performed in order to validate predictions of the model. As a result, both the model predictions and the experimental evidences pointed out a decrease of the total noise at low target expression levels and an increase at high expression levels, with respect to the unregulated case, see figure 2.5. Intrinsic noise reduction dominates at low expression, while extrinsic fluctuations due to the coupling with the miRNA dominate at high expression. Moreover, the extrinsic contribution due to the coupling with miRNA fluctuations increases as the miRNA-target interaction gets stronger.

These conclusions suggest an interpretation to the fact that miRNAs target preferentially genes that are expressed at low levels: in this case the noise reduction would be significant. Also the combinatorial action of different miRNAs weakly

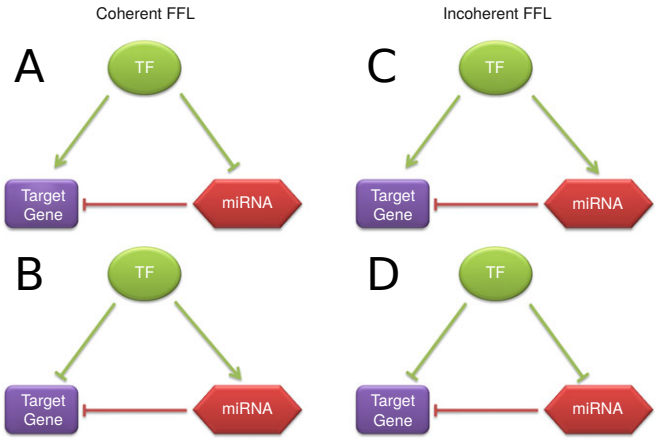

Fig. 2.6 MiRNA-mediadet Feed-Forward Loops (FFLs). (A) Positive coherent FFL. (B) Negative coherent FFL. (C,D) Incoherent FFLs. Adapted from [6].

repressing the same target might be understood in this framework. Indeed, the resulting target repression can be significant and noise can be reduced, due to the fact that the extrinsic contribution is small.

Another evidence of the role played by miRNAs as noise buffers derives from the study of specific network motifs called miRNA-mediated feedforward loops (FFLs). These motifs consist of a transcription factor that simultaneously regulates the expression of a gene and the expression of a miRNA that targets the regulated gene [6]. Different computational approaches [106, 107] showed that this kind of circuits is overrepresented in the network representing the interactions among transcription factors, miRNAs and target genes. The fact that miRNA-mediated FFLs are conserved is a hint of their important role in gene regulation.

The topology of these motifs can be divided into two categories: coherent FFL and incoherent FFL, see figure 2.6. In the coherent FFL the two pathways of regulation originating from the transcription factor are consistent, meaning that they both lead to the activation or repression of the target gene. On the contrary, in the incoherent FFL, one pathway leads to positive regulation of the target and the other one to negative regulation. The overall effect of the coherent FFL can be either positive, with the transcription factor activating the gene and repressing the miRNA (fig. 2.6 A), or negative, with the transcription factor repressing the gene and activating the miRNA (fig. 2.6 B). Also the incoherent FFL can be divided in two variants, with the transcription factor activating or repressing both the gene and the

miRNA, see figure 2.6 C and D respectively. On the one hand, a negative coherent FFL can be interpreted as a failsafe control, with the miRNA helping the transcription factor at preventing the gene expression. On the other hand, an incoherent FFL can be seen as a fine tuner of the gene expression [108].

In addition to these observations, the ability of the miRNA-mediated incoherent FFL at controlling the variability of gene expression was also investigated. In [108] the authors showed with stochastic modeling and numerical simulations that the incoherent FFL is able to buffer fluctuations originating from the transcription factor, conferring robustness to the expression of the target. This result supports the idea of the incoherent miRNA-mediated FFL as a fine tuner of gene expression: this motif can both tune the expression level and reduce its fluctuations. The fine-tuning function of many miRNAs can then represent a way to approach the issue of the weak repression exerted by miRNAs on most of their targets.

2.2.3 Overview about ceRNAs

As described in previous sections, the action of miRNAs is combinatorial. Indeed, one miRNA can regulate several different targets and one target can be regulated by many different miRNAs. Moreover, miRNAs can target both protein coding and non-coding transcripts. As a result, a potentially enormous miRNA-target interaction network exists. This fact led to the idea that the miRNA-target interaction network could be based on the competition between targets for binding to the same pool of miRNAs and that this competition could be a source of crosstalk between them. Indeed, in principle, the overexpression of one of the targets of a common miRNA, for example by increasing its transcription rate, could induce a derepression of the other targets via a sponge effect, leading to an increase of their free molecule share (i.e. the amount of mRNA molecules not bound to miRNAs). This is due to the fact that, when a miRNA molecule binds to a target mRNA, it is sequestered from the environment and it cannot repress other target molecules. The sponge effect consist in the fact that, when one target is overexpressed, its high number of mRNAs capture most of the miRNAs, leaving other targets less repressed. MiRNAs constitute then a limiting resource for which target transcripts compete.

When one of the target genes is upregulated, the miRNA molecules will preferentially bind to it, due to the increased amount of its transcripts. The miRNA repression

on the mRNAs of the competing genes will be then reduced and their free amount will increase. The crosstalk between targets of a common pool of miRNAs induced by this phenomenon was named competing endogenous RNA (ceRNA) effect [37] and its importance was experimentally demonstrated in different physiological and pathological conditions [38–40]. Particularly illustrative is the case of the tumor suppressor gene PTEN and its pseudogene PTENP1 [39]. Indeed, it was shown that the transcript of the pseudogene [109, 110] PTENP1, thanks to the high similarity of their sequences, exerts a regulatory function on PTEN by binding to a pool of common miRNAs. Furthermore, a dysregulation of PTENP1 seems to be involved in human cancer.

These results suggest that the crosstalk between targets of a common pool of miRNAs can represent a key element that has to be taken into account in order to understand the whole interaction network between protein coding and non-coding genes.

2.2.4 Theory: competition and crosstalk

Several studies [7, 20, 21] theoretically investigated the ceRNA effect by exploiting mathematical modeling. In [7] the authors followed a solution approach that took into account the stochastic nature of the system. In particular, a general network of M miRNAs and N target mRNAs (ceRNAs) was investigated by studying the master equation describing the dynamics of the system. The processes included in the master equation were the transcription and spontaneous degradation of miRNAs and ceRNAs and the titrative interaction between them, parameterized through an effective association rate, which constitutes an additional degradation channel. A supplementary parameter of the model is the so called catalyticity, which represents the fraction of the miRNAs that are degraded during the interaction with the target mRNA (which is always degraded). The value of this parameter, ranging from 0, completely catalytic interaction, to 1, completely stoichiometric interaction, has been discussed for long and there is not unanimous agreement on it [111–116].

The master equation resulted to be too difficult to solve analytically, even at the steady state. Indeed, the hierarchy of the moments obtained from the moment generating function was not closed. In order to go around this problem, a Gaussian approximation was used. In this way, the joint probability distribution describing the

system at the steady state was approximated with a multivariate Gaussian distribution. Then, the fact that for this kind of distribution each moment can be written as a function of the first two was used to close the hierarchy and solve the master equation. The approximated steady-state probability distribution was used, together with exact numerical simulations, to investigate the behavior of mean values, fluctuations and correlations of the system.

As a first result, the threshold-like behavior of the means (both of ceRNAs and miRNAs), due to the titrative nature of the miRNA-target interaction, was recovered. Particular attention was payed to the fact that this behavior can be seen as a function of the transcription rate of one of the ceRNAs, but also as a function of the transcription rate of one of the miRNAs, see figure 2.7. By looking at the behavior of the means, the crosstalk induced by the competition was also shown. In a network composed of multiple ceRNAs competing for binding to common miRNAs, the threshold level is located at the equimolarity point between ceRNAs and miRNAs. As a consequence, a variation in the expression of one of the targets can propagate to the other competitors and move the entire system above or below threshold. CeRNAs can then crosstalk due to the fact that they bind to the same pool of miRNAs, but, in a mirror-like way, also miRNAs can crosstalk by binding to the same pool of ceRNAs. From now on, we will focus on the behavior of the targets, even though all the results hold also from the miRNA perspective.

In addition to the results described above, the approximated probability distribution allowed to analytically compute noise indexes, i.e. Fano factor and coefficient of variation, and correlations, in terms of the Pearson correlation coefficient. The interesting result was that both noise and correlation presented a maximum in proximity to the threshold, where the number of free molecules is small and the system is ultrasensitive, see figure 2.8. These observations seem to indicate that in proximity to the threshold relative fluctuations are high, but also highly correlated. This phenomenon could be crucial for systems in which the relative amount of the protein output of different targets should be taken under control. For example, this could be the case of two target genes coding for proteins that have to interact making a dimer.

The results described above were shown to be robust with respect to the value of the catalyticity parameter. Indeed, even in case of purely catalytic interaction, in which all miRNAs are recycled after binding to the mRNA, the threshold behavior

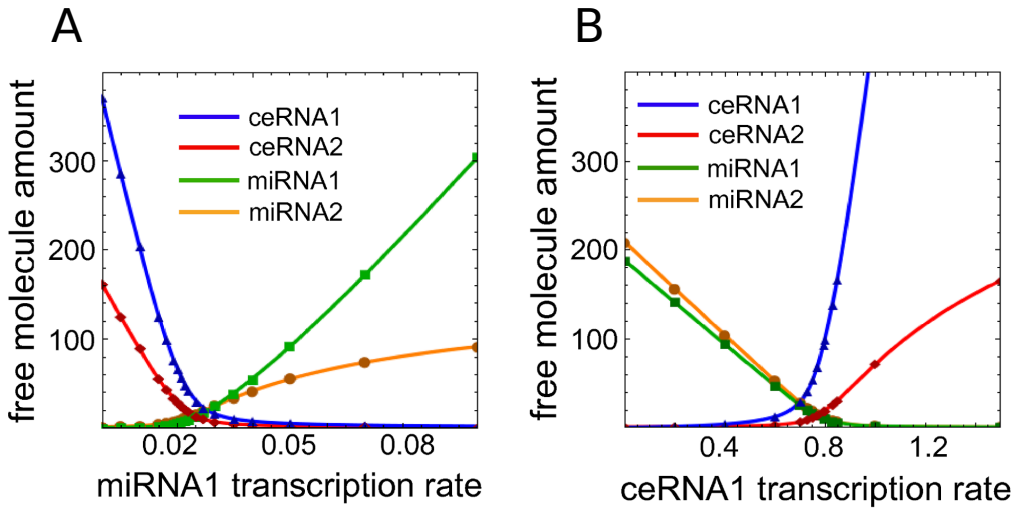

Fig. 2.7 Network composed of two genes competing for two miRNAs. Mean values as a function of (A) the transcription rate of miRNA1 and (B) the transcription rate of ceRNA1. Lines are obtained as approximate solution of the model described in [7], dots are numerical simulations. Adapted from [7].

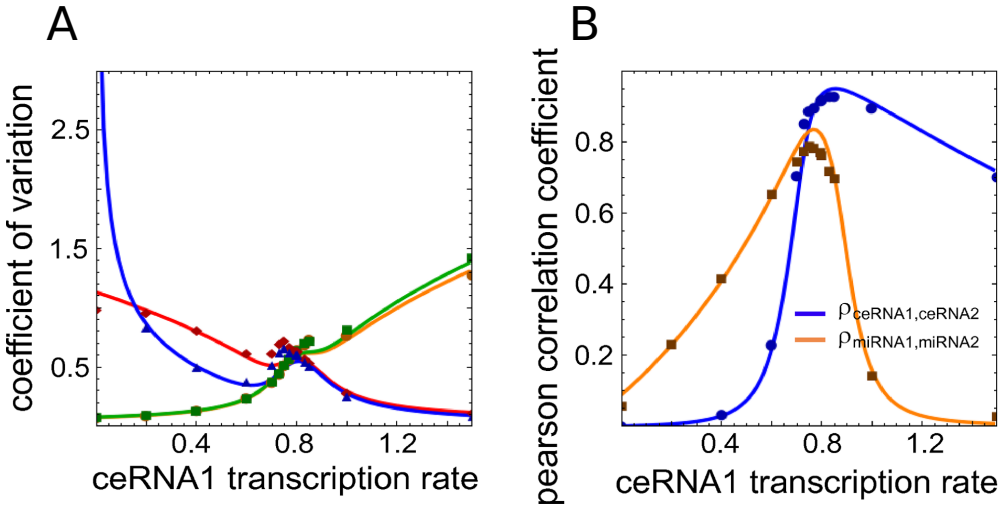

Fig. 2.8 Network composed of two genes competing for two miRNAs. (A) Coefficient of variation as a function of the transcription rate of ceRNA1. (B) Pearson correlation coefficient as a function of the transcription rate of ceRNA1. Lines are obtained as approximate solution of the model described in [7], dots are numerical simulations. Adapted from [7].

and noise and correlation maxima can be observed as an out-of-equilibrium property of the system.

Finally, some hints were obtained about the out-of-equilibrium behavior of a system of ceRNAs in response to the sudden switch on or off of one of the targets. In particular, the response times, defined as the time needed by the system to reach half of the way between the initial and final steady state, were studied through the numerical integration of the deterministic set of equations associated to the model. In a system of two targets competing for one miRNA, it was shown that, as a function of the miRNA transcription rate, the response times to the switch-on and to the switch-off of a target presented a maximum and a minimum respectively. These peaks were located at a value of the miRNA transcription rate corresponding to the threshold. These intriguing results demand further investigation and could be in principle validated through synthetic biology experiments.

2.2.5 Experimental evidences

The ability of miRNAs to induce threshold responses was shown through synthetic biology experiments in [4]. In this work the authors developed a two-color fluorescent reporter system, consisting in a plasmid, i.e. circular filament of DNA, containing two genes coding for fluorescent proteins: mCherry (red) and eYFP (yellow). The two genes were controlled by a bidirectional promoter, so that they could be transcribed by RNA polymerase with equal probability. On the 3'UTR of the gene coding for mCherry, a fixed number of binding sites (0,1,4,7) for the miRNA miR-20a was engineered, while the gene coding for eYFP was left untouched, see figure 2.9. The higher the number of binding sites, the stronger the interaction between the mCherry mRNA and miR-20a.

The plasmids containing the fluorescent reporter system were inserted in Human Embryonic Kidney cells 293 (HEK 293) by performing a transient transfection. In this way, the plasmids enter the cells and can be transcribed, but they are not permanently integrated in the cell genome, meaning that they are not replicated and are diluted when the cell duplicates. HEK 293 cells endogenously express the miRNA miR-20a which targets the mCherry mRNA. At the moment of transfection, the number of plasmids that enter each cell is random, then the expression of the fluorescent proteins is different from cell to cell. Nonetheless, due to the bidirectional

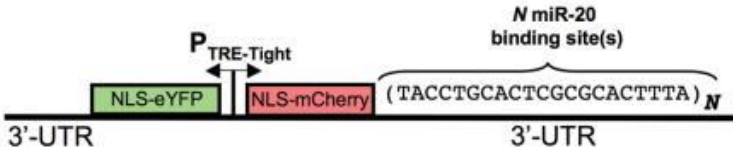

Fig. 2.9 Two-color fluorescent reporter system used in [4]. Bidirectional plasmid coding for mCherry and eYFP. The mCherry gene is engineered to contain a fixed number N of binding sites for the miRNA mir-20 on its 3'UTR. Adapted from [4].

promoter, the mCherry and eYFP genes are always expressed at the same level inside the cell. At the steady state after transfection, the fluorescence intensity of each cell in the channels of mCherry and eYFP was measured both through fluorescence microscopy and flow cytometry. The latter technique exploits microfluidic channels to measure the fluorescence intensity of each cell in the two channels, by exciting them with lasers with fixed wavelength. In this way the number of measured cells can reach the order of 10^5-10^6.

As pointed out above, the mCherry 3'UTR is engineered to contain 0, 1, 4 or 7 binding sites for miR-20a. When binding sites are present, the mCherry mRNA is targeted by the miRNA and the translation of the corresponding fluorescent protein is repressed. The expression of the eYFP protein, not regulated by the miRNA, is used as a control to monitor the level of transcriptional activity of the target mCherry, i.e. its expression level in absence of miRNA regulation. Transcriptional activity is different from cell to cell, due to the random number of plasmids entered into the single cells, and this allows to study a wide range of expression levels.

As expected, in absence of miRNA binding sites on the target mCherry, the scatter plot of the fluorescence intensity of mCherry as a function of eYFP presented a linear dependence. Increasing the number of binding sites, a nonlinearity and then the threshold-like behavior appeared. As previously described, the level of transcriptional activity corresponding to the threshold was found to be set by the amount of miRNAs. This evidence was obtained experimentally by modulating the amount of miRNAs through the transfection of either miR-20a mimic oligonucleotides or miRNA sponges inhibiting miR-20a activity. Furthermore, the steepness of the threshold was shown to be controlled by the miRNA-target interaction strength, in terms of the number of miRNA binding sites present on the 3'UTR.

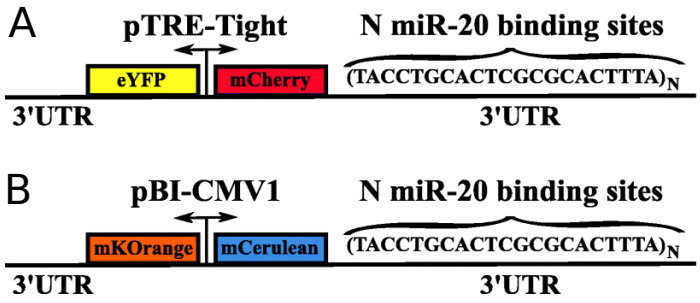

Fig. 2.10 Fluorescent reporter system used in [8]. (A) Bidirectional plasmid described in figure 2.9. (B) Bidirectional plasmid coding for mCerulean and mKOrange. The mCerulean gene is engineered to contain a fixed number N of binding sites for the miRNA mir-20 on its 3'UTR. Adapted from [8].

These results gave important insights on how miRNAs can regulate their targets and suggested that the mechanism of interaction is based on molecular sequestration. Recent studies [8, 117] investigated with a similar experimental setup the phenomenon of the competition between targets of a common miRNA (ceRNA effect). In particular, the work by Bosia and coauthors [8] aimed at validating experimentally some of the theoretical results obtained in [7] and discussed in the previous section. The authors built a system composed of two plasmids: the previously described one, used in [4], and a second plasmid with the same structure but coding for two different fluorescent proteins, mCerulean and mKOrange. Exactly as described above, the 3'UTR of the gene coding for mCerulean was engineered to contain 0, 1, 4 or 7 binding sites for the miRNA miR-20a, while the gene of mKOrange was left as a control to monitor the transcriptional activity, see figure 2.10.

The two plasmids constituted a system of two synthetic targets of the same miRNA. With this system, the ceRNA effect was investigated by performing transfection experiments with plasmids with different combinations of miRNA binding sites. The fluorescence intensity of each cell was then measured through flow cytometry. The results confirmed the model predictions on cross regulation. It was shown that the expression of one target and the steepness of its threshold response can be modulated by varying the interaction strength of the miRNA with the other target. Moreover, the results showed that the miRNA-mediated regulation increases cell-to-cell variability and in case of strong repression of one target (many binding sites on this target, few on the other one) its coefficient of variation presents a local maximum in proximity to the threshold, exactly as predicted by the model.

In addition to these results, it was found that also the correlation between the targets is maximal around the threshold and that an optimal range of values of miRNA-target interaction strength exists. This correlation originates from the fact that the fluctuations of one target are coupled to the fluctuations of the other target through the extrinsic noise of the miRNA, a phenomenon called retroactivity [118–120].

Finally, as an additional result of these experiments, it was shown that, in case of strong repression of one of the targets, bimodal distributions of fluorescence intensity can appear near the threshold. This result supports the interpretation that miRNA-mediated regulation increases cell-to-cell variability, but the fact that this variability is organized in two distinct phenotypes represents an intriguing phenomenon that deserves further investigation. In the following chapter we will theoretically address this question, relating it to the role played by extrinsic noise.

Chapter 3

Extrinsic noise induces bimodal distributions in miRNA-regulated genes

As described in the previous chapter, gene regulation mediated by microRNAs is an example of molecular sequestration. This mechanism induces a threshold-like response that makes the system ultrasensitive in proximity to the equimolarity point between miRNAs and mRNAs. An intriguing experimental result related to this system is the possibility of observing bimodal distributions of target expression in a population of identical cells. This phenomenon, observable without the need of any peculiar regulatory link, is thought to be related to the threshold-like behavior of the system and has not been characterized yet.

In this chapter we will address the question of how these bimodal distributions can emerge, showing the role that extrinsic noise can play in enhancing the phenomenon. We will first introduce a minimal model of molecular sequestration, that reproduces well the threshold behavior, and show analytically that the addition of extrinsic noise can lead to bimodal distributions. We will then analyze a more complex model which describes in more detail the miRNA-mediated gene regulation and reproduces most of its features. Through an analytic and a numerical approach we will characterize the onset of bimodal distributions in this system, highlighting the difference between behaviors at the single-cell and at the population level.

Part of the results discussed in this chapter have been published in [9] and [10].

3.1 Minimal model of molecular sequestration

In chapter 2 we presented the features of molecular sequestration, with particular attention to miRNA-mediated interactions. In this section, we will introduce a minimal model of molecular sequestration and use it to have an insight on how bimodal distributions can occur in presence of an extrinsic source of noise.

The model consists of two molecular species: T, the target and S, the sequester. The species S binds to T with rate k_+ forming a complex $\overline{TS}$ that sequesters both the molecules. The complex can then dissociate with rate k_- releasing the molecules in the environment where they can interact again. This first order reaction network is defined as follows:

$$T + S \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \overline{TS}. \tag{3.1}$$

For sake of simplicity, we will assume that the total number of molecules of the two species, i.e. the sum of the free and the bound molecules, is conserved. This leads to the following conservation laws:

$$S_T = S + \overline{TS} = const\,, \tag{3.2}$$

$$T_T = T + \overline{TS} = const\,. \tag{3.3}$$

Thanks to the conservation laws above, given T_T and S_T, S is simply a function of T ($S = S_T - \overline{TS} = S_{_T} - T_T + T$). This model has then only one free variable and, from now on, we will assume it to be T.

In order to study the features of this minimal model, we will first follow a deterministic approach, then a stochastic one. With the first approach we will have an insight on the behavior of the means, completely neglecting fluctuations, which are then taken into account by the stochastic analysis. Since we are here interested in the long term behavior of the system, in the following we will focus on the steady state solutions.

3.1.1 Deterministic approach: threshold characterization

In the deterministic approach, the observables of the system are the concentrations of the molecular species involved in the reaction network. This approximation is useful in the macroscopic limit, when the number of molecules composing the system is high. Nonetheless, we can use it to have an insight on the behavior of the mean values of the observables.

The rate equation describing the average behavior of our model is:

$$\frac{d[T]}{dt} = -k_+[T][S] + k_-[\overline{TS}]\,, \tag{3.4}$$

which can be easily solved at the steady state by exploiting the conservation laws for the concentrations $[S_T] = [S] + [\overline{TS}]$ and $[T_T] = [T] + [\overline{TS}]$. The only acceptable solution of the resulting second order equation is:

$$[T] = \frac{[T_T] - [S_T] - K + \sqrt{([T_T] - [S_T] - K)^2 + 4K[T_T]}}{2}\,, \tag{3.5}$$

where $K = k_-/k_+$ is the dissociation constant [1].

The concentration of T, as a function of its total concentration, presents the threshold-like behavior typical of the sequestration mechanism. Indeed, it shows a repressed regime, in which most T molecules are bound to S molecules, and an unrepressed regime, where the concentration of T increases linearly with that of T_T. In this regime nearly no molecule S can bind to T because most of them are already part of a complex with a T molecule.

In figure 3.1 some curves describing the mean value of T as a function of its total amount are shown, together with the analytic stochastic solution that we will present in the next section. The dissociation constant K controls the steepness of the threshold and, as a consequence, the ultrasensitivity of the system. The threshold becomes steeper as K decreases, meaning that S is more likely to bind the target T, while it becomes smoother, tending to a linear behavior, as K increases, meaning that the complex has a higher propensity to unbind. When the threshold is steep the system is ultrasensitive near the threshold point. This means that, in this region, a small variation of the total target concentration $[T_T]$ can produce a big fold-change in the concentration of free target T. The location of the theoretical threshold, i.e.

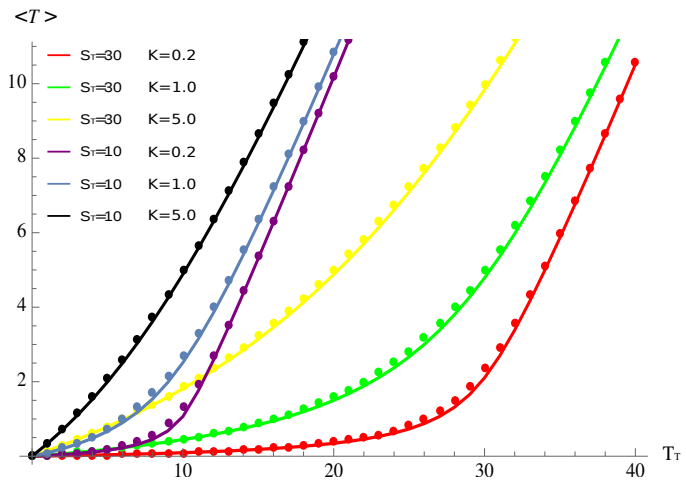

Fig. 3.1 $\langle T \rangle$ as a function of T_T for $S_T = 10$ and $S_T = 30$ and different values of the dissociation constant K. Dots correspond to the solution of the rate equation (3.5), while solid lines are obtained from the solution of the master equation (3.6).

the threshold obtained when $k_+ \gg k_-$, is set by the total amount of the sequester S_T and coincides with the equimolarity point between T and S.

3.1.2 Stochastic approach: master equation solution

The deterministic analysis allowed to study the average behavior of the minimal model, without taking into account any fluctuation. However, such fluctuations can be extremely relevant when the system is composed by only few elements. This is precisely the case of our model, where fluctuations play an essential role.

To take into account the stochastic behavior, we first write the master equation of our system [121, 122]. This equation describes the time evolution of the probability of having T free molecules at time t:

$$\frac{dP(T,t)}{dt} = k_+(T+1)(S+1)P(T+1,t) + k_-(\overline{TS}+1)P(T-1,t) - [k_+TS + k_-\overline{TS}]P(T,t)\,. \tag{3.6}$$

The master equation (3.6) includes the positive terms of binding and unbinding, when in the system there are $T+1$, $S+1$ and $T-1$, $S-1$ free molecules respectively, and the corresponding negative terms, when there are already T and S free molecules.

By implementing the conservation laws, eqs. (3.2) and (3.3), the master equation can be written in terms of T only:

$$\frac{dP(T,t)}{dt} = k_+(T+1)(S_T - T_T + T + 1)P(T+1,t) + k_-(T_T - T + 1)P(T-1,t) + \\ - [k_+T(S_T - T_T + T) + k_-(T_T - T)]P(T,t)\,. \tag{3.7}$$

Since the state of the network (3.1) is completely defined by the value of one variable and the reactions involved are reversible, we are in the presence of a reversible Markov Chain. Indeed, at the steady state a probability distribution $P(T)$ exists, so that the following relation holds:

$$P(T = i)\pi(j,i) = P(T = j)\pi(i,j)\,, \tag{3.8}$$

where $\pi(i,j) = P(T(t+1) = i|T(t) = j)$ is the transition probability from state j to state i. Equation (3.8) is also known as detailed balance relation and states that the flux of probability from state i to state j is equal to the one from state j to state i. This relation implies that at the steady state the system reaches equilibrium and $P(T)$ is the equilibrium distribution.

The detailed balance relation specific for our model reads:

$$P(T+1)k_+(T+1)(S+1) = P(T)k_-\overline{TS}\,, \tag{3.9}$$

where $P(T) = \lim_{t\to\infty} P(T,t)$. By recalling the conservation laws, eqs. (3.2) and (3.3), eq. (3.9) can be written in terms of T only:

$$P(T+1)k_+(T+1)(S_T - T_T + T + 1) = P(T)k_-(T_T - T)\,, \tag{3.10}$$

which gives:

$$P(T+1) = \frac{k_-}{k_+}\frac{(T_T - T)}{(T+1)(S_T - T_T + T + 1)}P(T)\,. \tag{3.11}$$

Given the fact that in our model the total number of molecules is conserved for both the species, the minimal number of molecules of T depends on the value of

S_T. If $T_T \leq S_T$, all T molecules can in principle be bound to S molecules, then the minimal number of T molecules allowed is 0. On the contrary, if $T_T > S_T$, even if all S molecules are bound in complex with T, there are $T_T - S_T$ free target molecules. We can then define the minimal number of target molecules T_{min} as:

$$T_{min} = \begin{cases} 0 & \text{for} \quad T_T \leq S_T \\ T_T - S_T & \text{for} \quad T_T > S_T \end{cases} . \tag{3.12}$$

In the following we will derive the equilibrium solution of the master equation (3.7) through a recursive argument. Equation (3.11) can be used to write the probability $P(T)$, starting from the probability of having the minimal number of target molecules $P(T_{min})$:

$$\begin{aligned}
P(T_{min}+1) &= \frac{k_-}{k_+}\frac{(T_T - T_{min})}{(T_{min}+1)(S_T - T_T + T_{min}+1)}P(T_{min}) \\
P(T_{min}+2) &= \frac{k_-}{k_+}\frac{(T_T - T_{min}-1)}{(T_{min}+2)(S_T - T_T + T_{min}+2)}P(T_{min}+1) \\
P(T_{min}+3) &= \frac{k_-}{k_+}\frac{(T_T - T_{min}-2)}{(T_{min}+3)(S_T - T_T + T_{min}+3)}P(T_{min}+2) \\
&\ldots \\
P(T_{min}+n) &= \frac{k_-}{k_+}\frac{(T_T - T_{min}-n+1)}{(T_{min}+n)(S_T - T_T + T_{min}+n)}P(T_{min}+n-1) ,
\end{aligned}$$

obtaining:

$$\begin{aligned}
P(T) = P(T_{min}+n) = \left(\frac{k_-}{k_+}\right)^n \frac{(T_T - T_{min})(T_T - T_{min}-1)\ldots(T_T - T_{min}-n+1)}{(T_{min}+n)(T_{min}+n-1)\ldots(T_{min}+1)} \cdot \\
\cdot \frac{1}{(S_T - T_T + T_{min}+n)(S_T - T_T + T_{min}+n-1)\ldots(S_T - T_T + T_{min}+1)} P(T_{min}) .
\end{aligned} \tag{3.13}$$

The above relation can be written in a more compact form making use of factorials:

$$P(T) = P(T_{min}+n) = \left(\frac{k_-}{k_+}\right)^n \frac{(T_T - T_{min})!T_{min}!(S_T - T_T + T_{min})!}{(T_T - T_{min} - n)!(T_{min}+n)!(S_T - T_T + T_{min} + n)!} P(T_{min}) . \tag{3.14}$$

Finally, by implementing the substitution $n = T - T_{min}$, introducing again the dissociation constant $K = k_-/k_+$ and including constant elements in the normalization term, we obtain:

$$P(T) = K^T \binom{T_T}{T} \frac{N}{(S_T - T_T + T)!} , \tag{3.15}$$

defined in the interval $T_{min} \leq T \leq T_T$, with the normalization term N given by:

$$N = \frac{1}{\sum_{T=T_{min}}^{T_T} K^T \binom{T_T}{T} \frac{1}{(S_T - T_T + T)!}} . \tag{3.16}$$

Equation (3.15) is the solution of the master equation (3.7) and represents the equilibrium probability distribution of having T free molecules.

In figure 3.2 the shape of the probability distribution $P(T)$ is shown for some values of the parameters. Keeping fixed the total amount of sequester molecules S_T, we vary (i) the dissociation constant K (which coincides with the effective strength of the interaction between T and S), and (ii) the total amount of target molecules T_T (above and below threshold). The distribution has a truncated Gaussian-like shape which turns into a more exponential-like one, centered in $T = 0$, as long as T_T is decreased below the threshold value $T_T = S_T$. Indeed, when the total amount of target T_T is small with respect to the total amount of sequester S_T, the most probable state is the one with all T bound in complex with S and then $T = 0$. Moreover, a decrease in the dissociation constant K induces the distribution to take an exponential-like shape peaked on the minimal value of allowed free T. This happens because small K means high interaction strength and, consequently, the most probable state is the one with $T = T_{min}$ (i.e. most S molecules tend to be bound to T). As a remark, we have to notice that the shape of the probability distribution always presents a single mode, no hints of multimodality can be found. This aspect of the distribution will be studied in more detail in the next section.

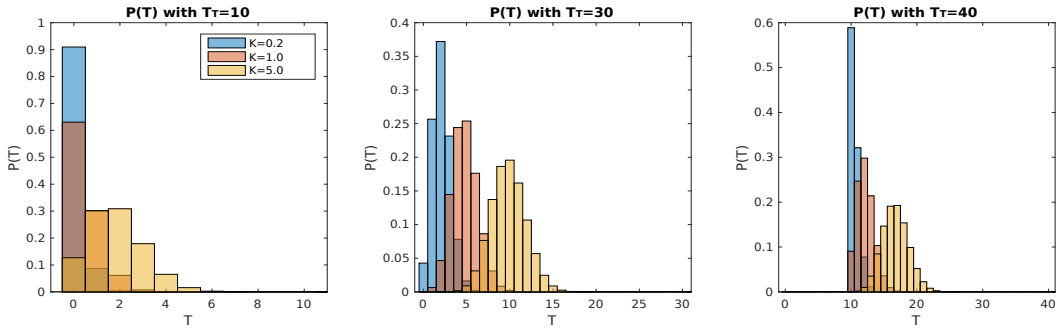

Fig. 3.2 $P(T)$ for $S_T = 10, 30, 40$ and for different values of the dissociation constant K.

To conclude, we present the analytic solution for the mean, studied in the previous section through the deterministic approach. By taking the ensemble average of T over the probability distribution $P(T)$, we obtain the exact expression for the mean of T, written in terms of Hypergeometric functions [123, 124]: [1]

$$\langle T \rangle = T_{min} \frac{{}_2F_2(1, T_{min} - T_T; T_{min}, 1 + T_{min} - T_T + S_T; -K)}{{}_2F_2(1, T_{min} - T_T; 1 + T_{min}, 1 + T_{min} - T_T + S_T; -K)} \,. \tag{3.17}$$

The above expression can be simplified if we separately consider the two regimes $T_T \leq S_T$ and $T_T \geq S_T$. When $T_T \leq S_T$ and $T_{min} = 0$ we have:

$$\langle T \rangle = K \frac{T_T}{1 - T_T + S_T} \frac{{}_1F_1(1 - T_T, 2 - T_T + S_T; -K)}{{}_1F_1(-T_T, 1 - T_T + S_T; -K)} \,, \tag{3.18}$$

when $T_T \geq S_T$ and $T_{min} = T_T - S_T$:

$$\langle T \rangle = (T_T - S_T) \frac{{}_1F_1(-S_T, T_T - S_T; -K)}{{}_1F_1(-S_T, 1 + T_T - S_T; -K)} \,. \tag{3.19}$$

as found in [125].

In figure 3.1 we compare the analytic behavior of the mean as a function of T_T to the deterministic solution. The correct threshold-like behavior is well approxi-

[1]
$${}_xF_y(a_1, ..., a_x; b_1, ..., b_y; z) = \sum_{n=0}^{\infty} \frac{(a_1)_n ... (a_x)_n}{(b_1)_n ... (b_y)_n} \frac{z^n}{n!}$$

where $(a)_0 = 1$ and $(a)_n = a(a+1)(a+2)...(a+n-1) \quad n \geq 1$

mated by the deterministic approximation, with larger deviations in proximity to the threshold, where nonlinearities are stronger.

3.2 Minimal model with extrinsic noise

Probability distributions presenting two modes, repressed and unrepressed, can be a feature of systems based on molecular sequestration. The phenomenon, although lacking a proper theoretical and experimental characterization, may have relevant biological implications regarding phenotype control, differentiation and related diseases. In order to deepen the understanding of how bimodal distributions arise in the context of molecular sequestration, and in particular to determine the minimal ingredients necessary for bimodality, we will develop the study of this phenomenon in the analytically well defined framework of our toy model. The work described in the following sections originates from the intuition that combining a threshold response and a fluctuating parameter can make the system explore two distinct states: a repressed and an unrepressed one. We will justify the fluctuating parameter as the result of an extrinsic source of noise. In our context, as introduced in chapter 1, extrinsic noise consists of all the fluctuations that affect the environment in which the stochastic reactions of biochemical networks take place. Variations, among identical cells, in the amount of cellular components and molecular machineries, for example ribosomes, can be considered as extrinsic noise, as well as gradients of signaling molecules in the extracellular environment. Moreover, variations related to cell cycle can also be considered as extrinsic noise when looking at a population of cells heterogeneous with respect to cell cycle. In our system, we will model extrinsic noise relaxing the assumption that the total amount of sequester molecules S_T is constant and allowing it to be described by a probability distribution $P(S_T)$. The following sections will define this probability and derive the analytic solution by implementing the law of total probability.

3.2.1 Discrete gaussian kernel

The minimal model of molecular sequestration in presence of extrinsic noise coincides with the model described in section 3.1 with the only difference that S_T is not a constant value, but has a probability distribution $P(S_T)$. The functional form of

$P(S_T)$ can be important because the shape of this distribution strongly influences the resulting distribution $P(T)$. In this section and in the following ones we will define and discuss the effects of a Gaussian probability distribution for the total amount of sequester, $P(S_T)$. The implications of using other kinds of extrinsic noise will be discussed in section 3.3.

Given that the variable T, being a number of molecules, can only assume discrete values, we chose $P(S_T)$ as defined by the discrete Gaussian kernel $T(n,x)$ which coincides with the scaled regular modified cylindrical Bessel function of integer order n [126]. In detail we have:

$$T(n,x) = e^{-|x|} I_n(x)\,, \tag{3.20}$$

where x is a real number and $I_n(x)$ is the regular modified cylindrical Bessel function of the first kind of integer order n [123]:

$$I_n(x) = i^{-n} J_n(ix)\,, \tag{3.21}$$

with $J_n(y)$ being the Bessel function of the first kind of integer order n. Functions in eq. (3.21) are the solutions of the modified Bessel equation, i.e. the Bessel equation in which the variable is changed from x to ix [123].

The probability distribution $P(S_T)$ is finally defined as follows:

$$P(S_T) = T(S_T - \langle S_T \rangle, x)\,, \tag{3.22}$$

where $\langle S_T \rangle$ is the mean value of the distribution and x controls its width: the greater x, the larger the distribution. Given the definition of eq. (3.22) the probability distribution $P(S_T)$ is always properly normalized.

3.2.2 Law of total probability and the full solution

We can think of our model in presence of extrinsic noise as composed of a population of identical systems, each of them with a constant S_T whose value is picked from the probability distribution $P(S_T)$. This suggests that the steady state probability distribution $P(T)$ is the superposition of single probability distributions obtained

for a given value of S_T, weighted with the probability of picking that value of S_T. In a more rigorous way, by exploiting the law of total probability [127], $P(T)$ can be written by the superposition of the conditional probabilities $P(T|S_T)$ (with S_T running from zero to infinity), weighted by $P(S_T)$:

$$P(T) = \sum_{S_T=0}^{\infty} P(T|S_T)P(S_T)\,. \tag{3.23}$$

The conditional probability $P(T|S_T)$ is the solution of the master equation (3.7) obtained for a constant S_T. Given the fact that S_T ranges from zero to infinity, in presence of extrinsic noise the minimal value of T is always $T_{min} = 0$ and the distribution $P(T)$ is then defined in the interval $0 \leq T \leq T_T$.

3.2.3 The probability distribution can be bimodal

Even though the expression of $P(T)$ in presence of extrinsic noise is difficult to derive analytically, the sum over S_T is straightforward to compute numerically. We thus have access to the full solution and we can study the effects of extrinsic noise without any approximation. In this section we will first focus on the shape of the resulting distribution $P(T)$ and then discuss the consequences of the introduction of extrinsic noise on the behavior of the mean of T.

The distribution $P(T)$ is obtained by the weighted superposition of probability distributions with constant S_T.

In section 3.1.2 we showed that the target distribution in case of pure intrinsic noise can take either a Gaussian-like or an exponential-like shape, depending on the value of the dissociation constant K and on the distance from the threshold, fig. 3.2. The superposition of this kind of curves presents a non trivial shape that in turn can be tuned by K, T_T and by the level of extrinsic noise. The intensity of noise can be quantified by the Coefficient of Variations (CV), a noise estimator defined as:

$$CV = \frac{\sigma_{S_T}}{\langle S_T \rangle}\,, \tag{3.24}$$

where σ_{S_T} is the standard deviation of the probability distribution of S_T with mean $\langle S_T \rangle$.

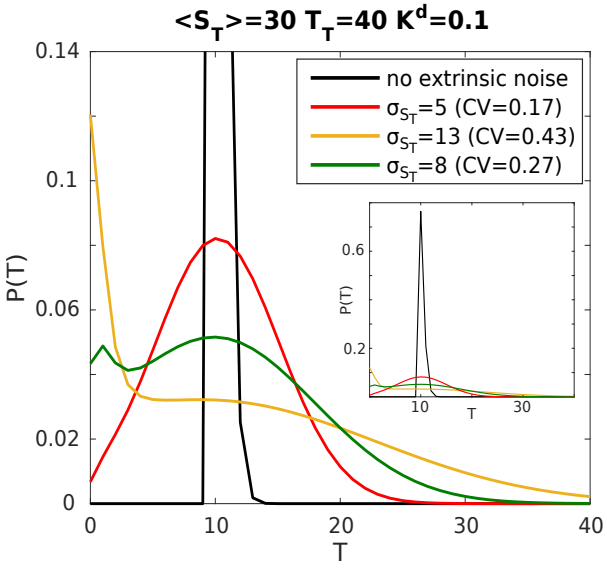

Fig. 3.3 Examples of probability distribution of T in presence of extrinsic noise. The presence of a bimodal distribution can be modulated both by varying T_T and by varying the noise level. Adapted from [9].

As an example, in figure 3.3 we show the probability distribution $P(T)$ that can be obtained by fixing $\langle S_T \rangle$ and varying the extrinsic noise by changing the variance of the distribution $P(S_T)$. We also show the probability distribution obtained with pure intrinsic noise: the addition of some extrinsic noise increases the variance of the distribution. As can be seen, the distribution can either have a single mode (peaked on $T = 0$ or on a finite value of T) or two modes (a repressed one, peaked close to $T = 0$ and an unrepressed one). The repressed peak is due to the absorbing border in $T = 0$: all the values of S_T that bring the system below threshold contribute to populate the repressed peak, regardless of their actual value. On the contrary, values of S_T that bring the system beyond threshold populate the entire range of values of T but, being the values around $\langle S_T \rangle$ more probable, they accumulate and form the unrepressed peak.

The possibility to obtain bimodality at the population level is an important feature of this minimal model, which combines a simple sequestration interaction between two molecular species and an extrinsic source of noise in terms of a fluctuating parameter. The bimodal shape of the distribution can be enhanced or disrupted both by an increase or decrease of each variable that we mentioned above. Nevertheless, we can point out the presence of some privileged directions of variation that lead to an increase of the range of values of the parameters in which bimodality is

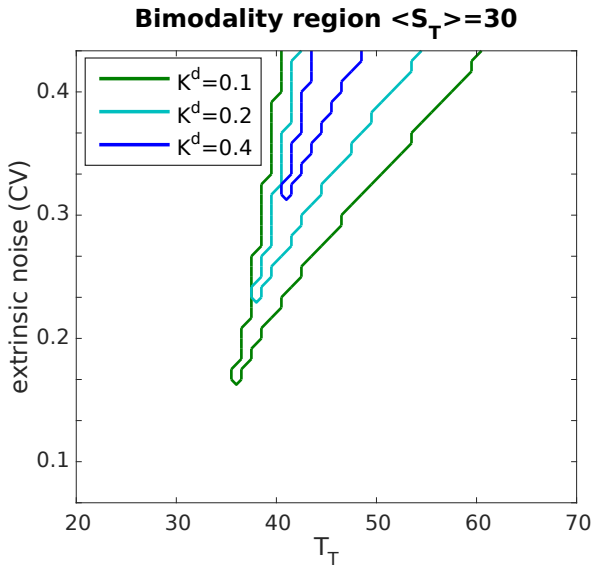

Fig. 3.4 Contour plot of bimodality as a function of T_T and noise intensity, for different dissociation constants K. $S_T = 30$. Adapted from [9].

found. In particular, we focus on the values of T_T for which we can find bimodal distributions for a given configuration of the parameters. The contour plot in figure 3.4 represents the region of bimodality as a function of T_T and noise intensity for different dissociation constants K. As we can see, as noise increases, the range of values of T_T allowing for bimodality increases and shifts towards higher values. Moreover, at a fixed noise level, the range of bimodality can be slightly increased by decreasing the dissociation rate. Summarizing, bimodality is favored by a high extrinsic noise and a strong interaction strength.

In biology, bimodal distributions are a common outcome of gene expression data, ranging from cell differentiation to cancer [49–54]. A high variability among cells, channeled by threshold mechanisms, may be useful to achieve different phenotypes, but it should as well be kept under control if these phenotypes refer to disease states.

We here showed that the combination of a threshold-like response produced by the sequestration mechanism and some normally-distributed extrinsic noise is a simple and robust mechanism to achieve bimodality at the population level, i.e. in an ensemble of systems that are identical with exception for the fluctuating parameter. The sequestration mechanism acts as an inbuilt noise filter, coherently shaping the resulting distribution without the need of further regulatory mechanisms.

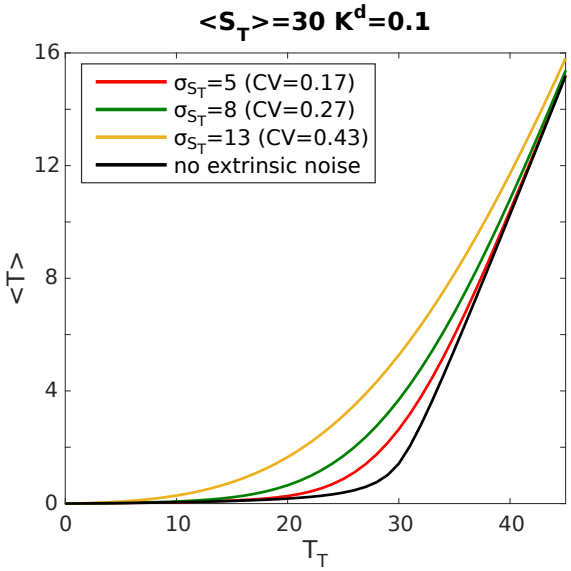

Fig. 3.5 $\langle T \rangle$ as a function of T_T for $\langle S_T \rangle = 30$ and different noise levels, $K = 0.2$. Adapted from [9].

We now turn our attention towards the effects of extrinsic noise on the target T. The introduction of extrinsic noise has a disruptive effect on the threshold: as the noise increases, the threshold becomes less steep. We underline that, since S_T is a fluctuating parameter, the position of the threshold is not well defined. As we wrote above, each element of the population has its own constant value of S_T and, consequently, its own theoretical threshold, defined by $T_T = S_T$. For this reason, each time we will mention the threshold in a system with extrinsic noise, we will refer to the one defined by the mean of the distribution of S_T, namely $T_T = \langle S_T \rangle$.

Figure 3.5 shows the mean of T, $\langle T \rangle$, for different values of $\langle S_T \rangle$ and different levels of noise. The curves with extrinsic noise are always located above their reference curve with intrinsic noise only. This is due to the fact that the system has the above mentioned absorbing border in $T = 0$. Increasing T_T, the distribution shifts towards higher values of T, loses the repressed peak and takes a symmetrical shape with the same mean of the case with pure intrinsic noise.

3.2.4 Extrinsic noise shapes the correlation between T and S

In the model of molecular sequestration with pure intrinsic noise, the correlation between the target and the sequester is trivially equal to 1. When the amount of free

molecules T increases by 1 due to the dissociation of a complex, the same happens to S, and when T decreases by 1 due to a binding reaction, again the same happens to S. The introduction of extrinsic noise in the system breaks this argument and results in a non trivial correlation between T and S.

We measure the correlation in terms of the Pearson correlation coefficient, which is defined as:

$$\rho \equiv \rho_{T,S} = \frac{\sigma_{TS}}{\sigma_T \sigma_S}\,, \tag{3.25}$$

where σ_{TS} is the covariance between T and S and σ_T, σ_S the standard deviations of the distributions $P(T)$ and $P(S)$ respectively.

To confirm what written above, we find that, without extrinsic noise on S_T, since the system is described by a single variable, $\sigma_{TS} = \sigma_T^2 = \sigma_S^2$ and the correlation is always equal to 1.

When adding extrinsic noise to the system, we can use the law of total probability to write the expectation values. The covariance can then be written as:

$$\sigma_{TS} = \sum_{S_T=0}^{\infty} \langle TS|S_T\rangle P(S_T) - \sum_{S_T=0}^{\infty} \langle T|S_T\rangle P(S_T) \sum_{S_T=0}^{\infty} \langle S|S_T\rangle P(S_T)\,. \tag{3.26}$$

From the conservation laws (eqs. 3.2 and 3.3) we have that $S = S_T - T_T + T$, then:

$$\sum_{S_T=0}^{\infty} \langle TS|S_T\rangle P(S_T) = \sum_{S_T=0}^{\infty} \langle T(S_T - T_T + T)|S_T\rangle P(S_T) = \langle TS_T\rangle - T_T\langle T\rangle + \langle T^2\rangle\,,$$

and:

$$\sum_{S_T=0}^{\infty} \langle S|S_T\rangle P(S_T) = \langle S_T\rangle - T_T + \langle T\rangle\,.$$

Combining these two expressions, we obtain:

$$\sigma_{TS} = \langle TS_T \rangle - T_T \langle T \rangle + \langle T^2 \rangle - \langle T \rangle \langle S_T \rangle + T_T \langle T \rangle - \langle T \rangle^2 = \sigma_T^2 + \sigma_{TS_T} \,. \quad (3.27)$$

In the same way, the variance of S is:

$$\sigma_S^2 = \sum_{S_T=0}^{\infty} \langle S^2 | S_T \rangle P(S_T) - \left(\sum_{S_T=0}^{\infty} \langle S | S_T \rangle P(S_T) \right)^2 ,$$

again, through the conservation laws:

$$\begin{aligned} \sum_{S_T=0}^{\infty} \langle S^2 | S_T \rangle P(S_T) &= \sum_{S_T=0}^{\infty} \langle (S_T - T_T + T)^2 | S_T \rangle P(S_T) = \\ &= \langle S_T^2 \rangle + T_T^2 - 2T_T \langle S_T \rangle + \langle T^2 \rangle + 2\langle TS_T \rangle - 2T_T \langle T \rangle \,, \end{aligned}$$

and:

$$\begin{aligned} \left(\sum_{S_T=0}^{\infty} \langle S | S_T \rangle P(S_T) \right)^2 &= \left(\sum_{S_T=0}^{\infty} \langle (S_T - T_T + T) | S_T \rangle P(S_T) \right)^2 = \\ &= \langle S_T \rangle^2 + T_T^2 - 2T_T \langle S_T \rangle + \langle T \rangle^2 + 2\langle T \rangle \langle S_T \rangle - 2T_T \langle T \rangle \,. \end{aligned}$$

Combining the two expressions we obtain:

$$\sigma_S^2 = \sigma_{S_T}^2 + \sigma_T^2 + 2\sigma_{TS_T} \,. \quad (3.28)$$

We can now finally write the expression for the Pearson correlation coefficient:

$$\rho = \frac{\sigma_{TS}}{\sigma_T \sigma_S} = \frac{\sigma_T^2 + \sigma_{TS_T}}{\sigma_T^2 \sqrt{\sigma_{S_T}^2 + \sigma_T^2 + 2\sigma_{TS_T}}} \,. \quad (3.29)$$

The extrinsic noise is able to shape the correlation profile between T and S, eventually inducing a significant anticorrelation in proximity to the threshold. This effect is also tuned by the value of the dissociation constant K^d.

3.3 Effect of other kinds of extrinsic noise

So far, we studied the effects of a normally distributed extrinsic noise on the total amount of sequester. Here, we turn our attention to different extrinsic noise distributions. In particular, we investigate the effects of a fluctuating S_T with a uniform distribution $P(S_T)$ defined on the interval $S_{Tmin} \leq S_T \leq S_{Tmax}$:

$$P(S_T) = \frac{1}{S_{Tmax} - S_{Tmin} + 1} \quad . \tag{3.30}$$

Mean and variance of the discrete uniform distribution are defined as:

$$\langle S_T \rangle = \frac{S_{Tmax} - S_{Tmin}}{2} \quad , \tag{3.31}$$

$$\langle S_T^2 \rangle - \langle S_T \rangle^2 = \frac{(S_{Tmax} - S_{Tmin} + 1)^2 - 1}{12} \quad . \tag{3.32}$$

As described in the previous section for an S_T with a Gaussian distribution, we derive the free target probability distribution as a weighted superposition of conditional probabilities that are solution of the master equation (eq. 3.7):

$$P(T) = \sum_{S_T=0}^{\infty} P(T|S_T)P(S_T) \quad . \tag{3.33}$$

In Fig. 3.6 we present some examples of $P(T)$ originated from sequester probability distributions with mean and variance comparable to the ones of Fig. 3.3.

A uniform extrinsic noise is not able to induce bimodal distributions of the target. Systems that pick a value of S_T below threshold are concentrated by the threshold response into a repressed peak with value of T close to 0. Differently than the Gaussian case, the expressed peak, which corresponded to the peak of the distribution of S_T, cannot be obtained in the uniform case. Indeed, each value of S_T has the same probability and the threshold response does not have any effect in the expressed regime. As a result, the free target probability distribution presents a flat plateau corresponding to the expressed regime.

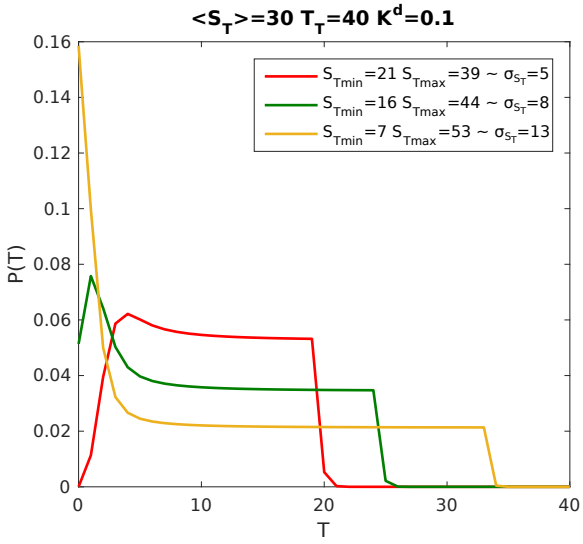

Fig. 3.6 Examples of probability distribution of T in presence of extrinsic noise with uniform distribution. Bimodal distributions cannot be obtained with this kind of extrinsic noise. Adapted from [9]

In order to obtain bimodal distributions of the free target amount, the extrinsic noise must have a peaked distribution, sufficiently broad to sample both below and above threshold.

3.4 A model of miRNA-regulated gene: RNA and protein

In the previous sections we studied a minimal model of molecular sequestration, analyzing the effects of the introduction of an external source of noise and showing that bimodal distributions can be achieved. The purpose was to give some insights about systems based on molecular sequestration through a toy model under analytic control. An example of a system based on molecular sequestration is the miRNA-mediated gene regulation, as discussed in chapter 2. We here introduce a more realistic stochastic model that describes this process and analyze its features in presence of extrinsic noise, using both analytic approximations and numerical simulations.

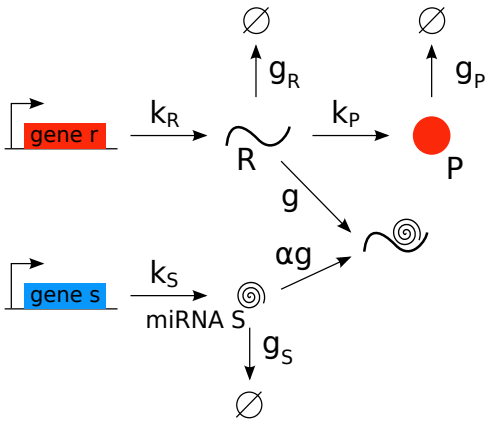

Fig. 3.7 MiRNA-target interaction network. The miRNA and the target mRNA are transcribed from independent genes with rate k_S and k_R respectively and degrade with rate g_S and g_R. The interaction is governed by the effective parameter g. The target mRNA is always degraded after binding, while the miRNA can be recycled with probability $1-\alpha$. Free mRNAs are translated with rate k_P into proteins which can in turn degrade with rate g_P. Adapted from [10].

The model that we consider is the one introduced by Bosia et al. in [7] and discussed in section 2.2.4, with only one miRNA and one target. We here recall the details of this model and then derive the approximated analytic solution.

We consider the case of one miRNA that regulates one target, but the results can be easily generalized to the case of multiple miRNAs and targets. The circuit representing the model is depicted in figure 3.7. The molecular species involved are miRNAs (S), target messenger RNAs (R) and proteins (P), product of the translation of the target mRNAs. MiRNAs and mRNAs are transcribed from independent genes. For simplicity we neglect all the intermediate reactions leading to the synthesis of mRNAs and miRNAs, and assume that they are produced at constant rates k_R and k_S respectively. MiRNAs and mRNAs can also be degraded by the action of specialized enzymes. Hereby, we assume these reactions to be governed by mass-action law with rates g_S and g_R. The molecular reactions associated to these processes are:

$$\emptyset \underset{g_R}{\overset{k_R}{\rightleftharpoons}} R\,, \qquad \emptyset \underset{g_S}{\overset{k_S}{\rightleftharpoons}} S\,. \tag{3.34}$$

MiRNAs act as post-transcriptional regulators by binding the target mRNAs in a complex RS that can be subsequently degraded. The interaction between miRNAs and mRNAs is quantified by the effective parameter g. This parameter accounts for

the strength of the miRNA-target coupling and for the complex degradation rate. From a biochemical point of view, this parameter depends on the number of miRNA binding sites present on the target molecule [4]. The formation of a miRNA-mRNA complex then reads:

$$R + S \xrightarrow{g} RS\,. \tag{3.35}$$

While mRNAs are assumed to be always degraded as a consequence of sequestration, miRNAs can be recycled with probability $1-\alpha$, also called catalyticity, in the following way:

$$RS \xrightarrow{1-\alpha} S\,. \tag{3.36}$$

Whenever the mRNAs are not bound to miRNAs, they can be translated by ribosomes into proteins with translation rate k_P and, as assumed for the other molecular species, proteins can be as well degraded by mass-action kinetics with rate g_P, i.e.:

$$R \xrightarrow{k_P} R + P\,, \qquad\qquad P \xrightarrow{g_P} \emptyset\,. \tag{3.37}$$

The behavior of this effective model of miRNA regulation can be analyzed through different analytic approximations. In the following sections we will discuss the various approaches on the basis of the results of previous works [4, 7].

3.4.1 Mean field description

The most immediate way to analyze this model is to neglect its intrinsic stochastic nature and develop a macroscopic description. In order to do that, we write the rate equations for the concentrations of the molecular species involved:

$$\frac{d[R]}{dt} = k_R - g_R[R] - g\,[R]\,[S]\,, \tag{3.38}$$
$$\frac{d[S]}{dt} = k_S - g_S[S] - \alpha\, g\,[R]\,[S]\,, \tag{3.39}$$
$$\frac{d[P]}{dt} = k_P[R] - g_P[P]\,. \tag{3.40}$$

The set of differential equations above describes the process of miRNA-mediated gene regulation deterministically. As done before, we are here interested in studying the long term behavior of the system. Differently from the minimal model of molecular sequestration, detailed balance does not hold for this system, for this reason, when the amount of free molecules of the three species does not change over time, we do not consider the system at equilibrium, but at the steady state. At the steady state, the system of equations can be easily solved obtaining the following expressions:

$$[R] = \frac{-g_R g_S + \alpha\, g\, k_R - g\, k_S + A}{2\alpha\, g\, g_R}\,, \tag{3.41}$$
$$[S] = \frac{-g_R g_S - \alpha\, g\, k_R + g\, k_S + A}{2 g\, g_S}\,, \tag{3.42}$$
$$[P] = \frac{k_P[-g_R g_S + \alpha\, g\, k_R - g\, k_S + A]}{2\alpha\, g\, g_R\, g_P}\,, \tag{3.43}$$

with A defined as:

$$A = \sqrt{4\alpha\, g\, g_R g_S k_R + (-g_R g_S + \alpha\, g\, k_R - g\, k_S)^2}\,.$$

The state of the proteins is uniquely determined by the state of the mRNAs, indeed $[P] = \frac{k_P}{g_P}[R]$.

Figure 3.8 shows the behavior of the concentration of mRNAs, miRNAs and proteins as a function of the target transcription rate k_R. We can see the clear threshold-like profile of the curves, already discussed in section 3.1.1, with the

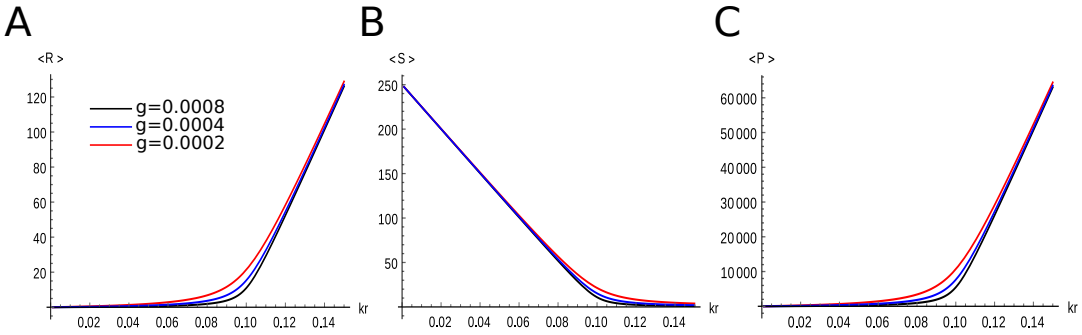

Fig. 3.8 Concentration of (A) R, (B) S and (C) P as a function of the transcription rate k_R for different values of g.

threshold that becomes steeper as the interaction strength g increases. The position of the theoretical threshold can be found by taking the limit of strong miRNA-mRNA coupling ($\lim_{g\to\infty}$) in the equations for the concentrations. In this limit, the expression for $[R]$ becomes:

$$[R] = \frac{k_R}{g_R} - \frac{k_S}{\alpha\, g_R}\,. \tag{3.44}$$

The above expression is negative for $k_R < \frac{k_S}{\alpha}$, but being $[R]$ a concentration of molecules, we assume it to be identically equal to zero in this regime. Indeed, since this approach is completely deterministic, when the target transcription rate is smaller than $\frac{k_S}{\alpha}$, all the molecules of R are bound to S molecules and the concentration of their free amount is zero. This result locates the threshold at the equimolarity point $\alpha k_R = k_S$, where the amount of free molecules of R and S are both equal to zero.

3.4.2 Approximate analytic solution: gaussian approximation and system size expansion

The model of miRNA regulation is much more complex than the toy model that we used to understand the main features of molecular sequestration. A complete description of this system would imply the analytic solution of the following master equation, describing the dynamics of the probability of finding R free mRNA molecules, S free miRNA molecules and P proteins at time t:

$$\begin{aligned}\frac{dP(R,S,P,t)}{dt} &= k_R\Big[P(R-1,S,P,t)-P(R,S,P,t)\Big]+\\ &\frac{g_R}{V_{cell}}\Big[(R+1)P(R+1,S,P,t)-RP(R,S,P,t)\Big]+\\ &k_S\Big[P(R,S-1,P,t)-P(R,S,P,t)\Big]+\\ &\frac{g_S}{V_{cell}}\Big[(S+1)P(R,S+1,P,t)-SP(R,S,P,t)\Big]+\\ &\frac{k_P R}{V_{cell}}\Big[P(R,S,P-1,t)-P(R,S,P,t)\Big]+\\ &\frac{g_P}{V_{cell}}\Big[(P+1)P(R,S,P+1,t)-PP(R,S,P,t)\Big]+\\ &\frac{g\alpha}{V_{cell}^2}\Big[(S+1)(R+1)P(R+1,S+1,P,t)-SRP(R,S,P,t)\Big]+\\ &\frac{g(1-\alpha)S}{V_{cell}^2}\Big[(R+1)P(R+1,S,P,t)-RP(R,S,P,t)\Big]. \qquad (3.45)\end{aligned}$$

The rates have been rescaled due to the fact that the variables are now numbers of molecules instead of concentrations. Indeed, we have $R=[R]V_{cell}$, $S=[S]V_{cell}$ and $P=[P]V_{cell}$, where V_{cell} is the volume of the cell in which the process takes place.

We mentioned before that detailed balance does not hold for this network. One could think that an analytic solution of the master equation could be obtained by introducing the moment generating function and solving the resulting system of partial differential equations. Unfortunately this approach is not feasible, even at the steady state. Each moment computed through the generating function depends on higher order ones and the system of partial differential equations is not closed. An analytic approximation is then needed, together with numerical simulations of the network to check the results.

Among the possible methods, we mention the Gaussian approximation [7] and the van Kampen's system-size expansion [121]. The first approach consists in assuming that the joint probability $P(R,S,P)$ describing the system is a multivariate Gaussian. This distribution has the property that all the moments with order higher then the second can be written in terms of the first and the second. This allows the system of partial differential equations to be closed and solved analytically, reconstructing the approximate joint probability distribution. The van Kampen's system-size expansion represents a different approach to obtain an approximate solution of the master

equation. In the following we will describe in detail this method and present its analytic predictions for the model under study.

The idea below the van Kampen's expansion is that the variables describing a stochastic system can be split into two contributions: the one describing the average behavior and the one describing fluctuations, small compared to the first one. In order to perform a power series expansion, we must select a small parameter to use in it, which has to be related to the size of fluctuations. Following van Kampen's notation we name the expansion parameter Ω and assume that it represents the size of the system, so that when Ω is large relative fluctuations are small. In order to present the van Kampen's approach step by step, we will accurately follow the notation and description developed by Elf and Ehrenberg in [128].

We consider a multidimensional system of size Ω composed of N different molecular species interacting through M chemical reactions. The state of the system is defined by the vector X, which contains the numbers of molecules of the different species. The corresponding molecular concentrations are defined by the vector $x = \frac{X}{\Omega}$ and spacial homogeneity by fast diffusion is assumed. The chemical master equation describing the time evolution of the probability distribution $P(X,t)$ can be written as:

$$\frac{dP(X,t)}{dt} = \Omega \sum_{j=1}^{M} \left(\prod_{i=1}^{N} E^{-S_{ij}} - 1 \right) f_j(x,\Omega) P(X,t)\,, \tag{3.46}$$

where $f_j(x,\Omega)$ are the transition rates. S_{ij} is the $N \times M$ stoichiometric matrix whose elements indicate how many molecules of species i are added or subtracted from the system in reaction j. $E^{-S_{ij}}$ is a step operator defined as $E^{-S_{ij}} g(X_1,...,X_N) = g(X_1,...,X_i - S_{ij},...X_N)$.

The assumption at the basis of the approximation is introduced by the following change of variables:

$$X_i \equiv \Omega \phi_i + \Omega^{1/2} \xi_i\,, \tag{3.47}$$

where ϕ_i is the macroscopic (deterministic) concentration of species i and ξ_i is a new random variable describing the contribution of fluctuations around the mean, whose width is assumed to be of order $\Omega^{1/2}$.

As a consequence of the change of variable, eq. (3.47), $P(X,t)$ is substituted by $\Pi(\xi,t)$, the probability distribution describing the state of the system in terms of the new random variable, which is defined as:

$$P(X,t) = P(\Omega\phi + \Omega^{1/2}\xi, t) = \Pi(\xi,t)\,. \tag{3.48}$$

The master equation (3.46) has now to be written in terms of the probability distribution $\Pi(\xi,t)$. Starting from the left-hand side and neglecting the arguments of the distributions for a lighter notation we have:

$$\frac{\partial P}{\partial t} = \frac{\partial \Pi}{\partial t} + \sum_{i=1}^{N} \frac{\partial \xi_i}{\partial t}\frac{\partial \Pi}{\partial \xi_i} = \frac{\partial \Pi}{\partial t} - \Omega^{1/2}\sum_{i=1}^{N} \frac{\partial \phi_i}{\partial t}\frac{\partial \Pi}{\partial \xi_i}\,, \tag{3.49}$$

where the differentiation has been made keeping constant the number of molecules, then $\frac{\partial X_i}{\partial t} = 0$ which implies $\frac{\partial \xi_i}{\partial t} = -\Omega^{1/2}\frac{\partial \phi_i}{\partial t}$.

To deal with the right-hand side of the master equation, we perform a Taylor expansion of the transition rates and the step operator:

$$f_j(x) = f_j(\phi + \Omega^{-1/2}\xi) = f_j(\phi) + \Omega^{-1/2}\sum_{i=1}^{N}\frac{\partial f_j(\phi)}{\partial \phi_i}\xi_i + O(\Omega^{-1})\,, \tag{3.50}$$

$$\prod_{i=1}^{N} E^{-S_{ij}} = 1 - \Omega^{-1/2}\sum_{i=1}^{N} S_{ij}\frac{\partial}{\partial \xi_i} + \frac{1}{2}\Omega^{-1}\sum_{i=1}^{N}\sum_{k=1}^{N} S_{ij}S_{kj}\frac{\partial^2}{\partial \xi_i \partial \xi_k} + O(\Omega^{-3/2})\,. \tag{3.51}$$

The master equation can then be written for the new variables by inserting equations (3.47), (3.49), (3.50) and (3.51) in eq. (3.46), obtaining:

$$\frac{\partial \Pi(\xi,t)}{\partial t} - \Omega^{1/2} \sum_{i=1}^{N} \frac{\partial \phi_i}{\partial t} \frac{\partial \Pi(\xi,t)}{\partial \xi_i} =$$
$$\Omega \sum_{j=1}^{M} \left(\Omega^{-1/2} \sum_{i=1}^{N} S_{ij} \frac{\partial}{\partial \xi_i} + \frac{1}{2} \Omega^{-1} \sum_{i=1}^{N} \sum_{k=1}^{N} S_{ij} S_{kj} \frac{\partial^2}{\partial \xi_i \partial \xi_k} + O(\Omega^{-3/2}) \right) \cdot$$
$$\cdot \left(f_j(\phi) + \Omega^{-1/2} \sum_{i=1}^{N} \frac{\partial f_j(\phi)}{\partial \phi_i} \xi_i + O(\Omega^{-1}) \right) \Pi(\xi,t) \,. \tag{3.52}$$

By collecting and equating all terms of order $\Omega^{1/2}$ we find:

$$\sum_{i=1}^{N} \frac{\partial \phi_i}{\partial t} \frac{\partial \Pi(\xi,t)}{\partial \xi_i} = \sum_{i=1}^{N} \sum_{j=1}^{M} S_{if} f_j(\phi) \frac{\partial \Pi(\xi,t)}{\partial \xi_i} \,. \tag{3.53}$$

The above equation is satisfied, meaning that the terms of order $\Omega^{1/2}$ cancel, since

$$\frac{\partial \phi_i}{\partial t} = \sum_{j=1}^{M} S_{if} f_j(\phi) \tag{3.54}$$

is the macroscopic equation that ϕ_i is assumed to follow.

Finally, by collecting and equating all terms of order Ω^0 we obtain:

$$\frac{\partial \Pi(\xi,t)}{\partial t} = \sum_{j=1}^{M} \left(-\sum_{i,k} S_{ij} \frac{\partial f_j(\phi)}{\partial \phi_k} \frac{\partial (\xi_k \Pi(\xi,t))}{\partial \xi_i} + \frac{1}{2} f_j(\phi) \sum_{i,k} S_{ij} S_{kj} \frac{\partial^2 \Pi(\xi,t)}{\partial \xi_i \partial \xi_k} \right) . \tag{3.55}$$

We can write the above expression in a more compact form which clearly shows that we are dealing with a linear Fokker-Planck equation:

$$\frac{\partial \Pi(\xi,t)}{\partial t} = -\sum_{i,k} A_{ik} \frac{\partial (\xi_k \Pi(\xi,t))}{\partial \xi_i} + \frac{1}{2} \sum_{i,k} [BB^T]_{ik} \frac{\partial^2 \Pi(\xi,t)}{\partial \xi_i \partial \xi_k} \,, \tag{3.56}$$

where $A_{ik} = \sum_{j=1}^{M} S_{ij} \frac{\partial f_j(\phi)}{\partial \phi_k}$ and $[BB^T]_{ik} = \sum_{j=1}^{M} S_{ij} S_{kj} f_j(\phi)$.

The stationary solution of the linear Fokker-Planck equation is known [129, 130] and corresponds to a multivariate Gaussian:

$$\Pi(\xi) = \left((2\pi)^{N/2}\sqrt{det\Xi}\right)^{-1} \exp(-\xi^T \Xi \xi/2)\,, \tag{3.57}$$

with zero mean and covariance matrix $\Xi = \langle \xi\, \xi^T \rangle$ given by the Lyapunov equation:

$$A\Xi + \Xi A^T + BB^T = 0\,. \tag{3.58}$$

The van Kampen's system-size expansion is also called Linear Noise Approximation, because only terms of order Ω^0 are taken into account to compute the behavior of fluctuations.

We can now apply this formalism to our model in order to find an approximate solution to the master equation (3.45). Defining n_X the vector describing the state of the system in terms of the number of molecules of the three species, $n_X = (R,S,P)$, and ρ_X the corresponding vector of the concentrations, $\rho_X = ([R],[S],[P])$, we can write the assumption of the van Kampen's expansion as:

$$n_X = V_{cell}\rho_X + V_{cell}^{1/2}\xi_X\,, \tag{3.59}$$

where ξ_X is the new random variable describing fluctuations. We now implement in the master equation (3.45) the change of variable, eq. (3.59) and follow the steps described above. At the leading order we recover the deterministic equations (3.38), (3.39) and (3.40). At the next-to-leading order we obtain the linear Fokker-Planck equation, whose steady-state solution is a trivariate normal distribution. The Lyapunov equation, eq. (3.58), corresponds to the following system of equations:

$$\begin{aligned} 0 = {} & k_r - 2g_r\langle\xi_R^2\rangle + g_r[R] - 2\alpha g[R]\langle\xi_S\xi_R\rangle - 2\alpha g[S]\langle\xi_R^2\rangle + \\ & \alpha g[R][S] - 2(1-\alpha)g[S]\langle\xi_R^2\rangle + (1-\alpha)g[R][S] - 2(1-\alpha)g[R]\langle\xi_S\xi_R\rangle\,, \end{aligned}$$

$$0 = k_s - 2g_s\langle\xi_S^2\rangle + g_s[S] - 2\alpha g[R]\langle\xi_S^2\rangle - 2\alpha g[S]\langle\xi_S\xi_R\rangle + \alpha g[R][S]\,,$$

$$0 = k_p[R] - 2g_p\langle\xi_P^2\rangle + g_p[P] + 2k_p\langle\xi_P\xi_R\rangle\,,$$

$$\begin{aligned}0 = &-g_r\langle\xi_S\xi_R\rangle - g_s\langle\xi_S\xi_R\rangle - \alpha g[R]\langle\xi_S^2\rangle - \alpha g[S]\langle\xi_S\xi_R\rangle - \\ &\alpha g[R]\langle\xi_S\xi_R\rangle - \alpha g[S]\langle\xi_R^2\rangle + [R][S]\alpha g - (1-\alpha)g[S]\langle\xi_S\xi_R\rangle - (1-\alpha)g[R]\langle\xi_S^2\rangle\,,\end{aligned}$$

$$\begin{aligned}0 = &-g_P\langle\xi_P\xi_R\rangle + k_P\langle\xi_R^2\rangle - \alpha g[R]\langle\xi_P\xi_S\rangle - \alpha g[S]\langle\xi_P\xi_R\rangle - \\ &(1-\alpha)g[S]\langle\xi_P\xi_R\rangle - (1-\alpha)g[R]\langle\xi_P\xi_S\rangle - g_r\langle\xi_P\xi_R\rangle\,,\end{aligned}$$

$$0 = -g_P\langle\xi_P\xi_S\rangle - g_s\langle\xi_P\xi_S\rangle + k_P\langle\xi_S\xi_R\rangle - \alpha g[R]\langle\xi_P\xi_S\rangle - \alpha g[S]\langle\xi_P\xi_R\rangle\,.$$

By including the deterministic equations (3.38), (3.39) and (3.40), the above system of equations can be solved, obtaining the covariance matrix and then the probability distribution $P(R,S,P)$ which is the approximate solution of the master equation (3.45).

3.4.3 Numerical simulations through the Gillespie's algorithm

In order to check the validity of the approximation, we need to have access to the real probability distribution that describes the system. This can be done with the help of numerical simulations.

The algorithm that we will use to reconstruct the probability distribution is the one introduced by Gillespie [131, 132] and named Stochastic Simulation Algorithm (SSA). In its direct form (the one that we will use for all the simulations present in this thesis), the SSA simulates the stochastic time evolution of a system of interacting

chemical species, described by the process $X(t)$, generating "exact numerical realizations of the process" [133]. In the SSA all the reactions that happen in a certain time interval are described, leading to the explicit simulation of the Markovian random walk in the space of the chemical species. In this sense the SSA is equivalent to the master equation. As a consequence, the probability distribution obtained from an ensemble of many trajectories generated by the SSA coincides with the probability distribution whose time evolution is described by the master equation.

Besides the molecular species that define the state of the system, the SSA, as well as the master equation, is based on two fundamental elements: the state-change vector and the propensity function. The state-change vector is defined as "the change in the vector of the species' molecular populations induced by a single occurrence of a particular reaction" [133]. The propensity function $a_j(x)$ of reaction j is "the function whose product with dt gives the probability that a particular reaction (in this case reaction j) will occur in the next infinitesimal time dt" [133], given $X(t) = x$. The propensity function of a given reaction is made by the product of a constant part, the reaction constant, and the number of molecules of the species involved in the reaction. The reaction constant corresponds to the intrinsic probability of the reaction and the number of molecules "weighs" this probability.

The backbone of the SSA is the Monte Carlo method. In summary, assuming a well mixed system, the flowchart of the algorithm is the following:

1. Initialize the number of molecules of each species and the values of the reaction constants.

2. Calculate and store the value of the propensity $a_j(x)$ of each reaction, together with the value of $a_0(x)$, defined as the sum of all the propensities.

3. Generate two random numbers r_1 and r_2 from a uniform distribution in the unit interval.

4. Use r_1 and r_2 to compute the interval of time τ for the next reaction to occur and the index j of that reaction. In particular, by using the standard Monte Carlo inversion generating method [134] we have:

$$\tau = \frac{1}{a_0(x)} \ln\left(\frac{1}{r_1}\right), \tag{3.60}$$

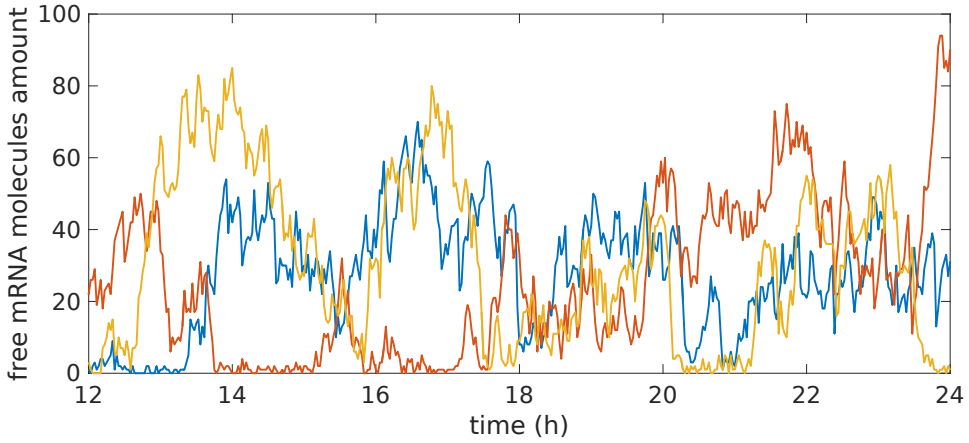

Fig. 3.9 Examples of steady-state trajectories over time of the free mRNA amount. $k_S = 1.2 \times 10^{-3}$ nM min^{-1}, $k_R = 2.7 \times 10^{-3}$ nM min^{-1}, $g_S = 1.2 \times 10^{-2}$ min^{-1}, $g_R = 2.4 \times 10^{-2}$ min^{-1}, $g = 1.5 \times 10^{3}$ nM^{-1} min^{-1} and $\alpha = 0.5$

and:

$$j = \text{the smallest integer satisfying } \sum_{j'=1}^{j} a_{j'}(x) > r_2 a_0(x)\,. \tag{3.61}$$

5. Increase time by τ and update the numbers of molecules according to the state-change vector of the reaction that occurred.

6. Repeat from step 2 or end the simulation.

Since we are interested in the steady-state distribution, we will end each run of the SSA only when the system reaches thermalization.

As an example, in figure 3.9 we show some steady-state trajectories of the number of free mRNAs.

The probability distributions obtained through the SSA can be compared to the analytic approximation and can be used to compute mean values and size of the fluctuations of the species of the network, see Fig. 3.10. As can be seen in figure, both the mean and the coefficient of variations are well approximated by the van Kampen's expansion. In addition to the threshold behavior of the mean, we can see a local maximum of the coefficient of variations in proximity of the threshold. This aspect, already discussed in section 2.2.4, is related to the fact that around

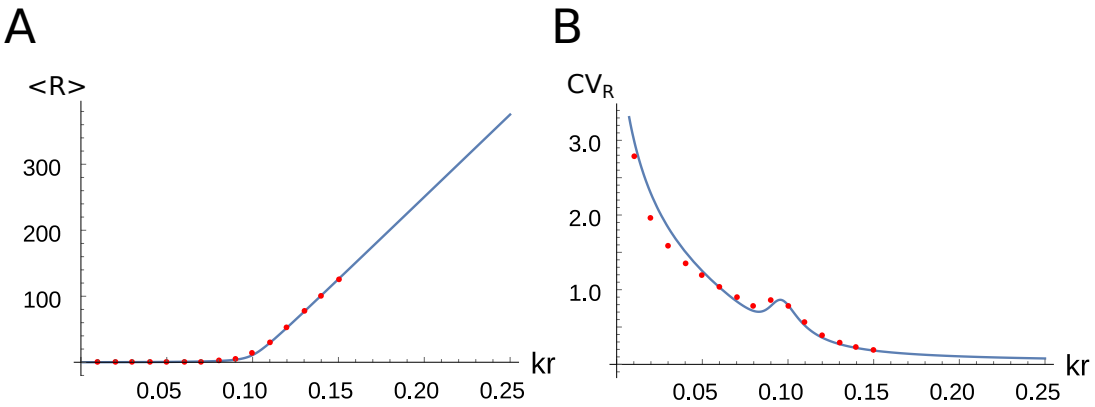

Fig. 3.10 Mean and coefficient of variation of the mRNA as a function of k_R. Solid lines approximate solution, dots numerical simulations.

the equimolarity point, i.e. the threshold, the system is ultrasensitive, then a small fluctuation in the relative amount of mRNAs and miRNAs can be amplified in a big fold-change.

3.5 MiRNA-target interaction model with extrinsic noise

So far we presented a model of miRNA-mediated gene regulation and analyzed it through analytic approximations and numerical simulations. In this system the only source of stochasticity was intrinsic, due to the inherent randomness of molecular reactions. As it was done for the minimal model of molecular sequestration, our goal is now to investigate the effects on the system of an additional source of noise. This extrinsic noise will be introduced by letting the value of one parameter fluctuate across a population of identical cells. We will then recover the full probability distribution of the system by applying the law of total probability and compare it to the exact one as obtained from the SSA.

3.5.1 Law of total probability on the approximate analytic solution

In order to study the potential consequences of environmental fluctuations, we introduce an extrinsic source of noise in terms of a fluctuating miRNA transcription

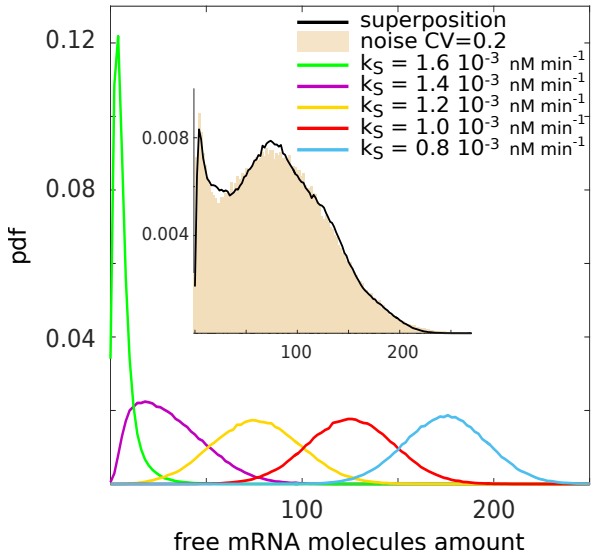

Fig. 3.11 Comparison between the bimodal mRNA noisy distribution and the weighted superposition of distributions obtained without noise for different miRNA transcription rates. The black line is the result of the weighted superposition of the distributions represented in the main plot, which are obtained through numerical simulations. Adapted from [10].

rate k_S. The network that describes the model is represented in figure 3.7. Following the approach used for the minimal model of molecular sequestration, we assume this parameter to be picked from a normal distribution defined for positive values of k_S:

$$P(k_S) = \frac{1}{\sqrt{2\pi\sigma_{k_S}^2}} e^{-\frac{(k_S-\langle k_S\rangle)^2}{2\sigma_{k_S}^2}}, \tag{3.62}$$

where $\langle k_S\rangle$ and $\sigma_{k_S}^2$ are respectively the mean and variance of k_S. The intensity of the extrinsic noise is governed by the value of the variance of the distribution and can be quantified in terms of the coefficient of variations: $CV = \sigma_{k_S}/\langle k_S\rangle$.

The master equation (3.45) that we introduced for the model with pure intrinsic noise does not hold anymore in presence of a fluctuating parameter. Moreover, these fluctuations on k_S do not allow to write a similar equation for this system in a simple way. However, as proposed in [127] and previously described for the minimal model, the probability distribution of the full system in presence of extrinsic noise, $P(R,S,P)$, can be written as a weighted superposition of conditional probabilities by using the law of total probability, see Fig. 3.11:

$$P(R,S,P) = \int P(k_S)P(R,S,P|k_S)dk_S\,, \tag{3.63}$$

where $P(k_S)$ is the Gaussian distribution of k_S and $P(R,S,P|k_S)$ is the conditional probability of observing a certain configuration of the system, given a specific value of k_S. This conditional probability can be obtained as solution of the master equation (3.45) for any given k_S. We can therefore apply again the van Kampen's system-size expansion on the master equation and obtain all the moments of the distribution. These moments are functions of the fluctuating parameter k_S and the full solution can be obtained by averaging the result over all the values of k_S as in equation (3.63).

3.5.2 Gillespie's simulations with extrinsic noise

The approximate analytic solution of the master equation in presence of extrinsic noise, as obtained from eq. (3.63), can be compared to the exact one making use of numerical simulations. The fluctuating miRNA transcription rate k_S is inserted in the framework of the SSA previously described, by picking a value for k_S from the distribution, eq. (3.62), at the beginning of each run of the algorithm. The picked value of k_S is kept constant throughout the realization of the algorithm, simulating the behavior of a cell belonging to a population in which the distribution of k_S is given by eq. (3.62).

Since the mRNA transcription rate k_R is fixed, the fact that the system explores different values of k_S means that it can experience both the repressed state, below threshold, and the unrepressed state, above threshold. In the repressed state, the number of free mRNA molecules is close to zero, because the total amount of miRNAs is bigger than the total amount of mRNAs. In the unrepressed state, the total amount of mRNAs is bigger than that of miRNAs and their mean amount increases linearly with the transcription rate k_R. Due to the absorbing border created by the threshold, all the values of k_S that bring the system below threshold contribute to the formation of a peak around zero in the resulting distribution of free mRNA. Conversely, values of k_S that lead the system to the unrepressed state, result in amounts of free mRNAs that generate the unrepressed peak. The position of the unrepressed peak is directly linked to the mean of the miRNA transcription rate ditribution. Indeed, being $\langle k_S \rangle$ the most likely value of k_S, it is mapped in the most

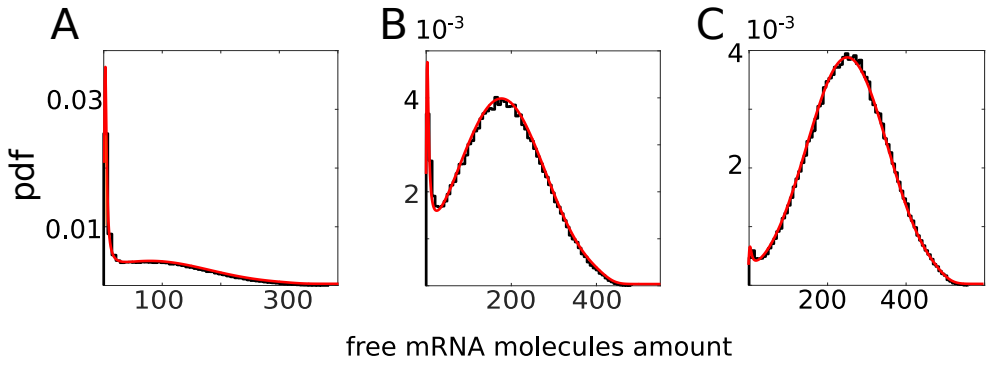

Fig. 3.12 Examples of free mRNA distribution for different values of k_R. From (A) to (C): $k_R = 3.1 \times 10^{-3}$ nM min^{-1}, 4.0×10^{-3} nM min^{-1}, 4.8×10^{-3} nM min^{-1}. Solid red lines correspond to the analytic approximation, solid black lines are numerical simulations. [10].

likely number of free mRNAs of the unrepressed region. The combination of this point with the absorbing border in zero creates the basis of bimodality.

As discussed for the minimal model, we can find configurations of the parameters in which the repressed and unrepressed peak are separated. In figure 3.12 we show some examples of free mRNA distributions obtained through numerical simulations and compare them to the approximate analytic solution. As can be seen, for these values of the parameters, the analytic solution approximates quite well the exact one, even in case of bimodal distribution. Our approximate analytic approach can then be used to have insights on the shape of the free mRNA distribution in presence of extrinsic noise, nonetheless numerical simulations are needed in order to check the results.

The kind of bimodality described above arises at the population level, due to a parameter (k_s) whose value fluctuates across an ensemble of otherwise identical systems. Nevertheless, numerical simulations showed that a bimodal mRNA distribution can appear also in absence of extrinsic noise. This bimodality at the single-cell level originates when the system is very close to a steep threshold, meaning that the miRNA-mRNA interaction strength g must be high. In proximity of a steep threshold the system is ultrasensitive. In these conditions, even a small intrinsic fluctuation can make the system jump from the repressed to the unrepressed state and vice versa. This happens continuously over time in each identical system and can give rise to bimodal distributions. Our approximate analytic approach cannot reproduce bimodal distributions in case of pure intrinsic noise. Indeed, the outcome of the van Kampen's expansion is always a Gaussian distribution, which is by definition

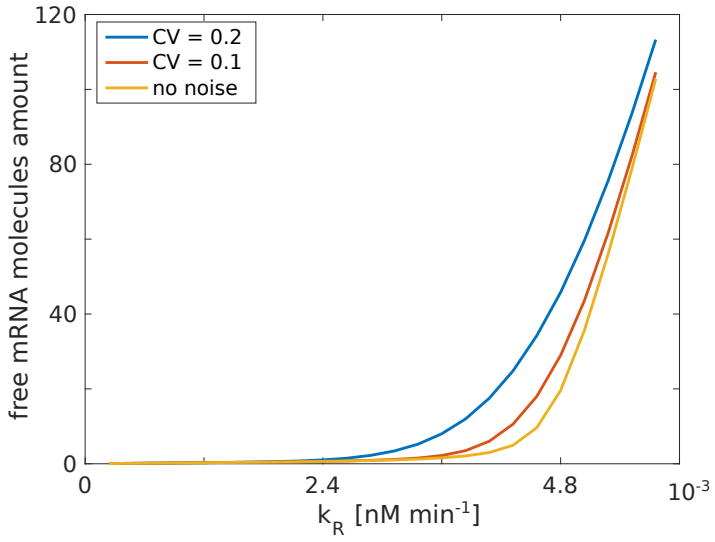

Fig. 3.13 Free mRNA molecules amount as a function of the transcription rate k_R for different noise levels.

unimodal. In order to investigate bimodal distributions at the single-cell level in this system, numerical simulations are essential.

We will discuss in more detail the differences between bimodality at the population and single-cell level in the following sections, where we will point out the duality between the miRNA-target interaction strength and the extrinsic noise.

To conclude, we now mention the effect of extrinsic noise on the mean values. Exactly as we discussed in section 3.2.3 for the minimal model, the threshold becomes smoother as extrinsic noise increases. Moreover, the mean in presence of extrinsic noise is always higher than the pure intrinsic noise case, see fig. 3.13.

3.6 Discussion of the results

In previous sections we introduced a model of miRNA-mediated gene regulation composed of one miRNA and one target. We proposed an approximate solution of the system's master equation based on the van Kampen's system-size expansion and used the Stochastic Simulation Algorithm to have access to the exact solution. We then turned our attention towards the effects of the addition of some extrinsic noise, in terms of a fluctuating miRNA transcription rate. We derived an approximate solution

by exploiting the law of total probability and checked the results through numerical simulations. As a result of this analysis, we showed that mRNA bimodal distributions can appear, both at a population and single-cell level. In the following we will discuss in detail the features of these different bimodal distributions, analyzing the role played by the parameters of the model. We will then examine how the shape of the mRNA distribution influences the distribution of proteins. The considerations that we will present are based on what we published in [10].

3.6.1 Single-cell vs population-induced bimodality

The understanding of the underlying mechanisms that allow cell diversity and variability is of fundamental importance to increase our knowledge about phenomena like cell fate determination, differentiation, organism development and related diseases. MiRNAs are known to be strongly involved in developmental mechanisms [76, 135–137] and can play an important role in controlling cell variability. We here discuss the details of the appearance of bimodality in the expression distribution of a gene controlled by a miRNA.

As presented in previous sections, the molecular sequestration mechanism acting between miRNAs and target mRNAs, gives rise to non-trivial threshold effects in condition of quasi equimolarity between the two molecular species. Keeping fixed the mRNA transcription rate to some value k_R^*, the threshold value of the miRNA transcription rate k_S is defined by: $k_S^* = \alpha k_R^*$.

If miRNAs are produced at a rate above threshold, $k_S > k_S^*$, the system becomes enriched in miRNAs, which tend to bind most of the available mRNA molecules, therefore preventing their translation. In this condition, from the point of view of the mRNA, the system can be considered below threshold, in the above mentioned repressed state. When the miRNA transcription rate is below its threshold value, $k_S < k_S^*$, the system is enriched in target mRNAs. This condition corresponds to the unrepressed state of the system, where the mean amount of free mRNA molecules increases linearly with their transcription rate k_R. The steepness of the threshold-like transition between the repressed and unrepressed state gets stronger as the miRNA-target coupling g increases. In proximity to the threshold, the intrinsic stochasticity of chemical reactions makes the system switch randomly between the repressed and unrepressed state. This stochastic switching can give rise to bimodal distributions

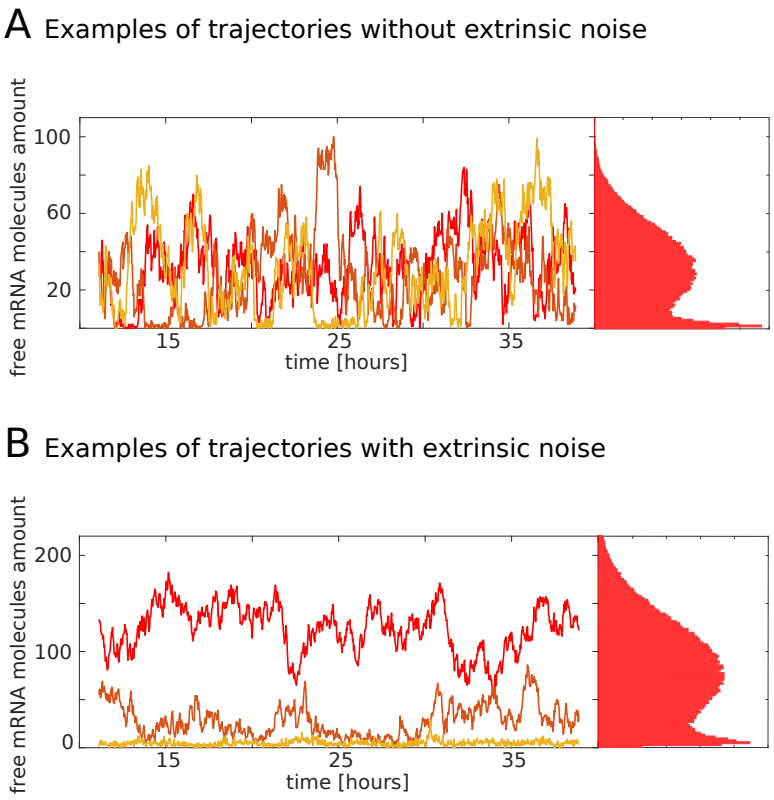

Fig. 3.14 Examples of trajectories and corresponding distribution. (A) Strong miRNA-target coupling, no extrinsic noise. (B) Weak miRNA-target coupling, extrinsic noise. Adapted from [10].

of the amount of free target. However, the presence of bimodality is confined to a very narrow range of values of the mRNA transcription rate k_R and requires a high interaction strength. In figure 3.14 (A) we show an example of this kind of bimodal distribution, together with some steady-state trajectories in which jumps between the repressed and unrepressed state can be recognized. This bimodality is a characteristic of each cell in which the miRNA network is present. Indeed, every single cell can switch from the repressed to the unrepressed state, provided that the interaction strength between miRNA and target is high enough.

When some extrinsic noise is added to the system, the value of the miRNA transcription rate k_S is different from cell to cell and is described by a Gaussian distribution centered in $\langle k_S \rangle$. The value of $\langle k_S \rangle$ sets the position of the threshold in the space of the mRNA transcription rates, indeed $k_R^* = \langle k_S \rangle / \alpha$. Given the mean $\langle k_S \rangle$ and standard deviation σ_{k_S} of the miRNA transcription rate distribution $P(k_S)$, following the scheme depicted in figure 3.15, we can describe the extrinsic-noise

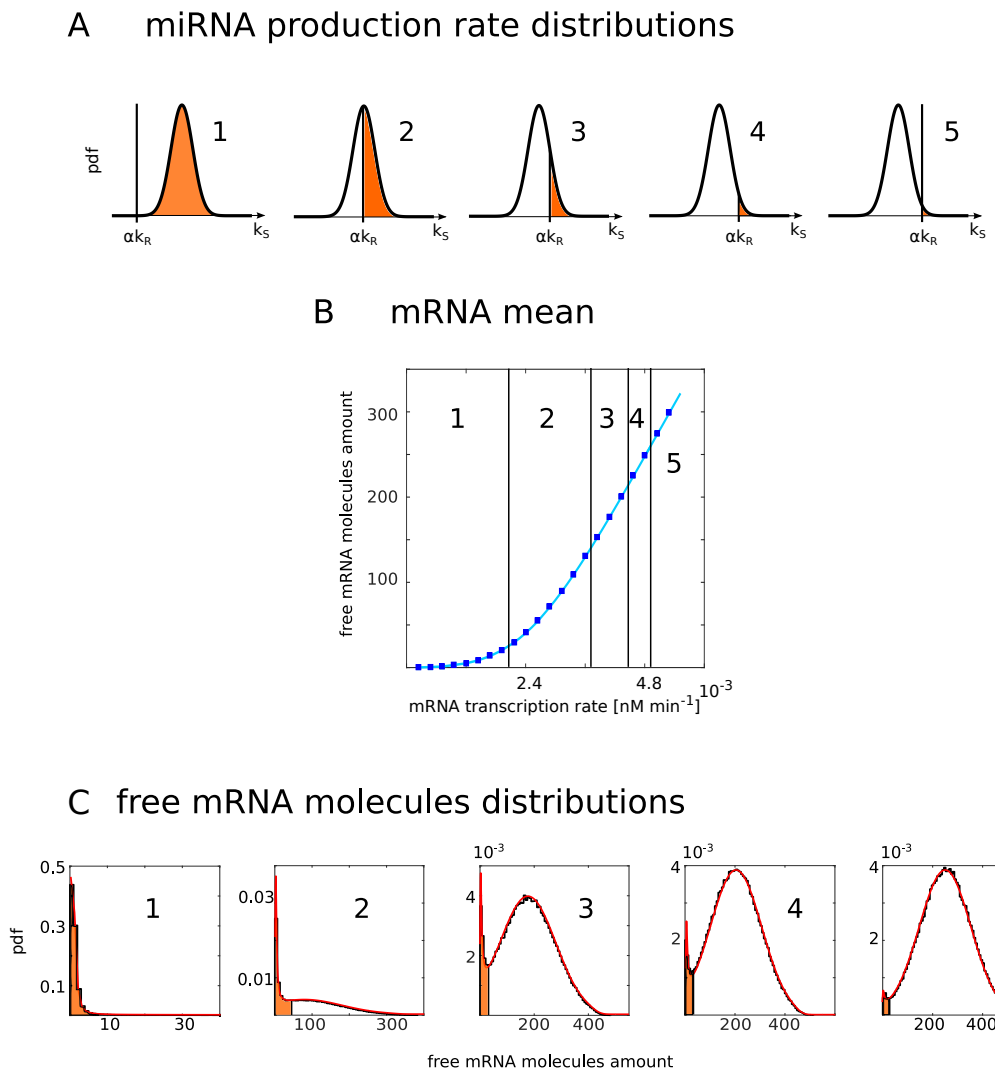

Fig. 3.15 Extrinsic-noise-induced bimodality. (A) Qualitative representation of the distribution of the miRNA transcription rate k_S. The black vertical line indicates the value of the miRNA transcription rate $k_S = \alpha k_R$ for different values of the target transcription rate k_R . The distributions represent the different regions labeled from 1 to 5 shown in (B). The region of the distribution contributing to the repressed state is colored in orange. (B) Free mRNA molecules amount as a function of the target transcription rate k_R . Solid lines are analytic predictions while blue squares correspond to numerical simulations. (C) Free mRNA molecules distributions corresponding to the regions in (B). Solid black lines correspond to numerical simulations while solid red lines are analytic predictions. The repressed region is coloured in orange. Adapted from [10].

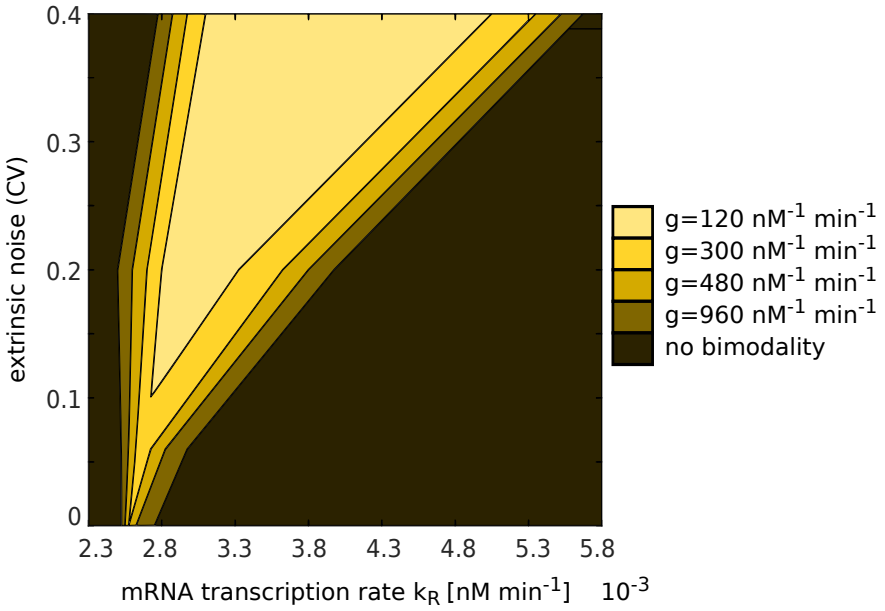

Fig. 3.16 Phase diagram that identifies the bimodality region as a function of the transcription rate k_R and of the level of extrinsic noise for different values of g. Adapted from [10].

scenario as follows. When the target transcription rate k_R is well below its threshold value, $k_R \ll k_R^*$, most of the miRNA transcription rates drawn from $P(k_S)$ will bring the system to the repressed state, where the target molecules are all bound to miRNAs and their free amount is close to zero. For values of k_R closer to the threshold, the miRNA transcription rates extracted from the left-tail of the distribution will result in some unbound targets. Above the threshold, as $k_R > k_R^*$, the majority of the drawn k_S will lead to the unrepressed state while the right tail of the distribution will sample from the repressed region. This heterogeneity of the miRNA transcription rates, leading some cells to be in the repressed state and others in the unrepressed one, is the key to obtain bimodal distributions at the population level. In figure 3.14 (B) an example of this population-induced bimodality is shown, together with three representative steady-state trajectories. The yellow one corresponds to a k_S that holds the system in the repressed state, the red to a k_S that allows some unbound mRNA molecules and the orange to a k_S, close to the threshold value, that lets the system jump between the repressed and unrepressed state.

By analyzing the effect of different levels of extrinsic noise, we found that, increasing the intensity of noise (i.e. increasing σ_{k_S}), the range of values of the

mRNA transcription rate for which bimodality is present gets wider. This effect is illustrated in figure 3.16, where the bimodality region of the distribution is shown as a function of the mRNA transcription rate k_R and the extrinsic noise level, for different miRNA-target coupling strengths. In addition to that, we found that also the separation between the two peaks of the distributions increases with the noise intensity.

An important take-home message of this analysis is that, differently from the pure intrinsic noise case, a fine tuning of the target transcription rate is no longer necessary in order to obtain bimodal expression distributions in a population of cells. Indeed, in figure 3.16 we can see that, even in case of a high k_R, there could be a non-negligible fraction of the randomly picked k_S that leads the system to the bound state, building the repressed peak of the bimodal distribution.

Finally, we want to notice how the distributions in figure 3.15 (C), obtained through numerical simulations, are well approximated by the solution of the system-size expansion. This proves the goodness of this approximation for the extrinsic noise case, anyhow numerical simulations are always mandatory to verify the analytic predictions.

3.6.2 Interplay between coupling strength and extrinsic noise

In the previous section we showed that the existence of a bimodal distribution is enhanced by extrinsic noise, as the range of bimodality, in terms of the values of k_R, increases with the noise intensity. Figure 3.16 takes into account different values of the miRNA-target interaction strength g. Focusing on the x axis of this figure, which describes the pure intrinsic noise case, we see that only high interaction strengths give rise to bimodality and its range of existence gets wider as the coupling increases. This behavior is maintained also in presence of extrinsic noise.

Both the interaction strength g and the extrinsic noise affect the range of values of k_R in which bimodality is present. Nonetheless, we know that the coupling acts at a single-cell level, while the extrinsic noise induces a population effect. From now on, we will neglect the differences of the underlying mechanisms and we will focus only on the resulting distribution of free mRNAs that can be measured across a population of cells. From this point of view, the interaction strength and the extrinsic noise can be seen as different “handles” to control the same phenomenon.

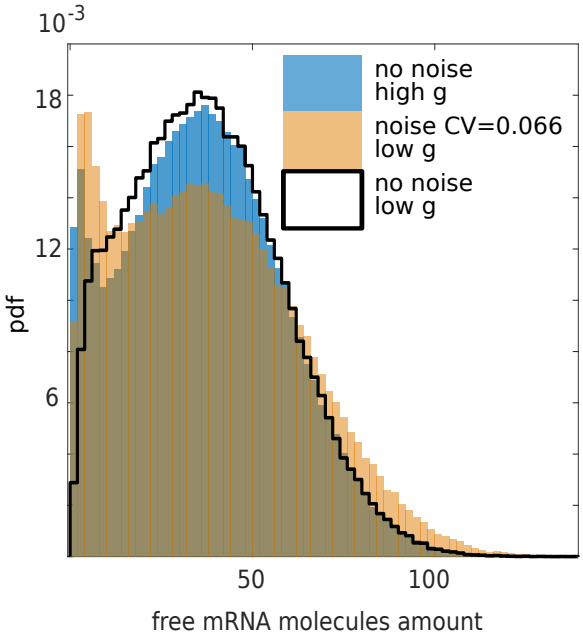

Fig. 3.17 Interplay between interaction strength and extrinsic noise. Target mRNA probability distributions obtained from numerical simulations. Solid black line: unimodal reference case with low g and no extrinsic noise. Blue: no extrinsic noise, high g. Orange: extrinsic noise, low g. Adapted from [10].

In order to investigate this possible duality between the two quantities, we ran numerical simulations aimed at finding whether it is possible to obtain rather the same bimodal distribution acting on just one handle. As shown in Figure 3.17, this is the case. Indeed, let's consider first a reference case with pure intrinsic noise and low miRNA-target coupling (black line). Starting from this reference condition, we see that similar bimodal distributions can be obtained either by increasing the coupling between miRNA and mRNA (blue histogram) or by introducing a stronger extrinsic noise (orange histogram).

In conclusion, our results show that the miRNA-mRNA coupling and the extrinsic noise act at a similar level with respect to bimodality. In fact, a higher level of extrinsic noise can compensate for a weak coupling in order to obtain bimodal expression distributions and, consequently, two different phenotypes.

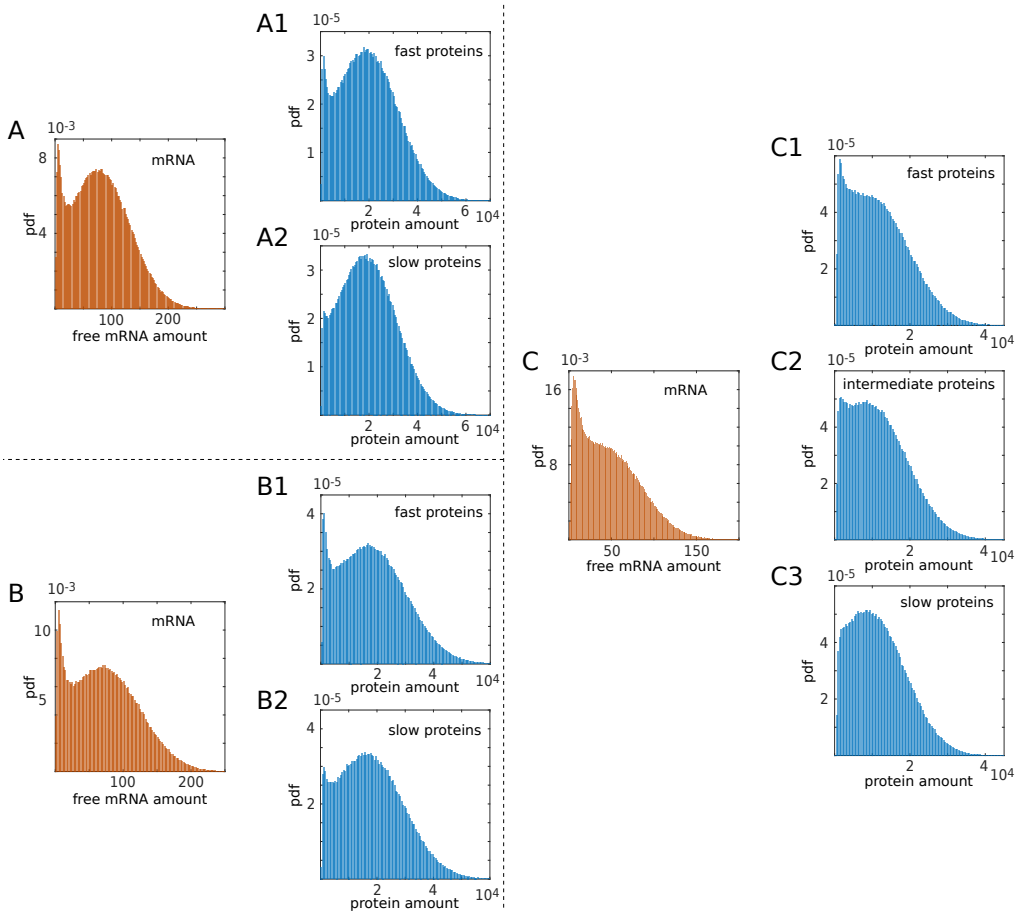

Fig. 3.18 Protein stability alters the shape of the distribution. Free mRNA distributions are represented in orange, while protein distributions corresponding to different levels of protein stability are represented in blue. Fast proteins distributions are obtained for a protein half life comparable to the mRNA one; in this condition the state of the protein is copying the mRNA one and the distributions almost coincide. Slow protein distributions are obtained for a protein half life up to 10 times longer than the mRNA one; as a consequence of the higher protein stability different outcomes can be achieved, depending on the level of extrinsic noise, the miRNA-target interaction strength and the closeness to the threshold (k_R). Starting with a well defined bimodal distribution (A1) and (B1), for a fixed level of extrinsic noise, the repressed peak can be buffered (A2) or not (B2) depending on the value of k R . If the initial distribution is unimodal repressed (C1), for a given range of parameters, it can be converted into a unimodal unrepressed (C3), crossing a bimodal state (C2), by increasing the protein stability. Adapted from [10].

3.6.3 Bimodality on proteins

So far we analyzed in detail the features of the free mRNA distribution, focusing our study on bimodality. Our model of miRNA-mediated gene regulation is built to include also the dynamics of the proteins translated from the mRNAs. In this section, we will follow the analysis performed on the mRNAs, in order to characterize the shape of the protein distributions. Particular attention will be paid on the role of time scales.

From a deterministic perspective we know that the mean amount of proteins is proportional to the mean amount of free mRNAs, in particular we have $[P] = k_P/g_P[R]$. The behavior of the distribution is controlled by the degree of separation between the time scale of the mRNA and protein dynamics. The parameters that can be used to evaluate the separation between the time scales are the mRNA and protein degradation rates, g_R and g_P, since they control the stability of the respective molecules. When the proteins' dynamics is fast, at least comparable to the one of the mRNAs, $g_P \simeq g_R$, their distribution follows closely the shape of the mRNA one. As the proteins' dynamics gets slower, $g_P < g_R$, the intrinsic fluctuations of the mRNA tend to be filtered out and this leads to narrower protein distributions. In the limit of time-scale separation, $g_P \ll g_R$, proteins "see" only the mean value of free mRNAs and the protein distribution is a Poissonian with mean $\langle P \rangle = k_P/g_P V_{cell}[R]$.

This effect acts at a single-cell level, but it affects also the shape of the protein distribution in presence of extrinsic noise. Indeed, for any miRNA transcription rate k_S, the associated protein distribution gets narrower as the protein dynamics gets slower. As a consequence, also the protein distribution in presence of extrinsic noise concentrates around its mean. This behavior can have opposite effects on the shape of the protein distribution, when the mRNA one is bimodal or in proximity to bimodality.

When the target distribution is bimodal, a slower proteins dynamics will make the protein distribution more peaked around its mean, favoring in this way the unrepressed peak, which is located close to the mean. This can lead the protein distribution to a complete loss of the repressed peak and, consequently, of its bimodal shape. The result can then be a unimodal protein distribution associated to a bimodal mRNA one. In other cases, protein bimodality can resist to the noise reduction induced by the slower dynamics. Indeed, as we wrote above, this is a single-cell

mechanism and, for strongly bimodal target distributions, it is not necessarily able to compensate the effects of the extrinsic noise.

When the mRNAs present a repressed unimodal distribution, the noise reduction due to a slower proteins dynamics can induce the rise of a second, unrepressed, peak. This would imply a bimodal protein distribution associated to a unimodal repressed mRNA one. Finally, even slower proteins dynamics can make the unrepressed peak overcome the repressed one, leading a unimodal unrepressed protein distribution in presence of a unimodal repressed mRNA one.

Some examples of the conditions described above are reported in figure 3.18. The consequence of these results is that a slower protein dynamics, favoring the expression in proximity to the mean, can produce completely different effects on the shape of the protein distribution. The occurrence of a specific condition depends on the interplay among the extrinsic noise level, the miRNA-mRNA interaction strength and the transcription and degradation rates. Despite that, the fact that in a population of cells the protein distribution can present a completely different shape with respect to the mRNA, can have interesting biological implications from the point of view of the coexistence of two phenotypes at the protein and mRNA level.

3.6.4 Bimodality in endogenous conditions: fold repression and time-varying extrinsic noise

One my wonder whether the results presented above hold in endogenous conditions. In order to make a first check, we can focus on two important quantities: the number of molecules and the fold repression. If the magnitude of these two quantities is comparable to the one observed in experiments, we can conclude that the phenomena described in our analysis are likely to be observed in nature.

In the experimental framework [4, 8], fold repression is measured as the ratio between the unregulated and the regulated level of expression of the miRNA target. It is then a quantity that takes into account only mean expression levels and indicates the average level of repression with respect to the case without regulation (constitutive expression).

As it can be observed in Fig. 3.19, the fold-repression profile is significantly sensitive to offsets in the data. Indeed, if the offset is zero or close to zero (panel (A)),

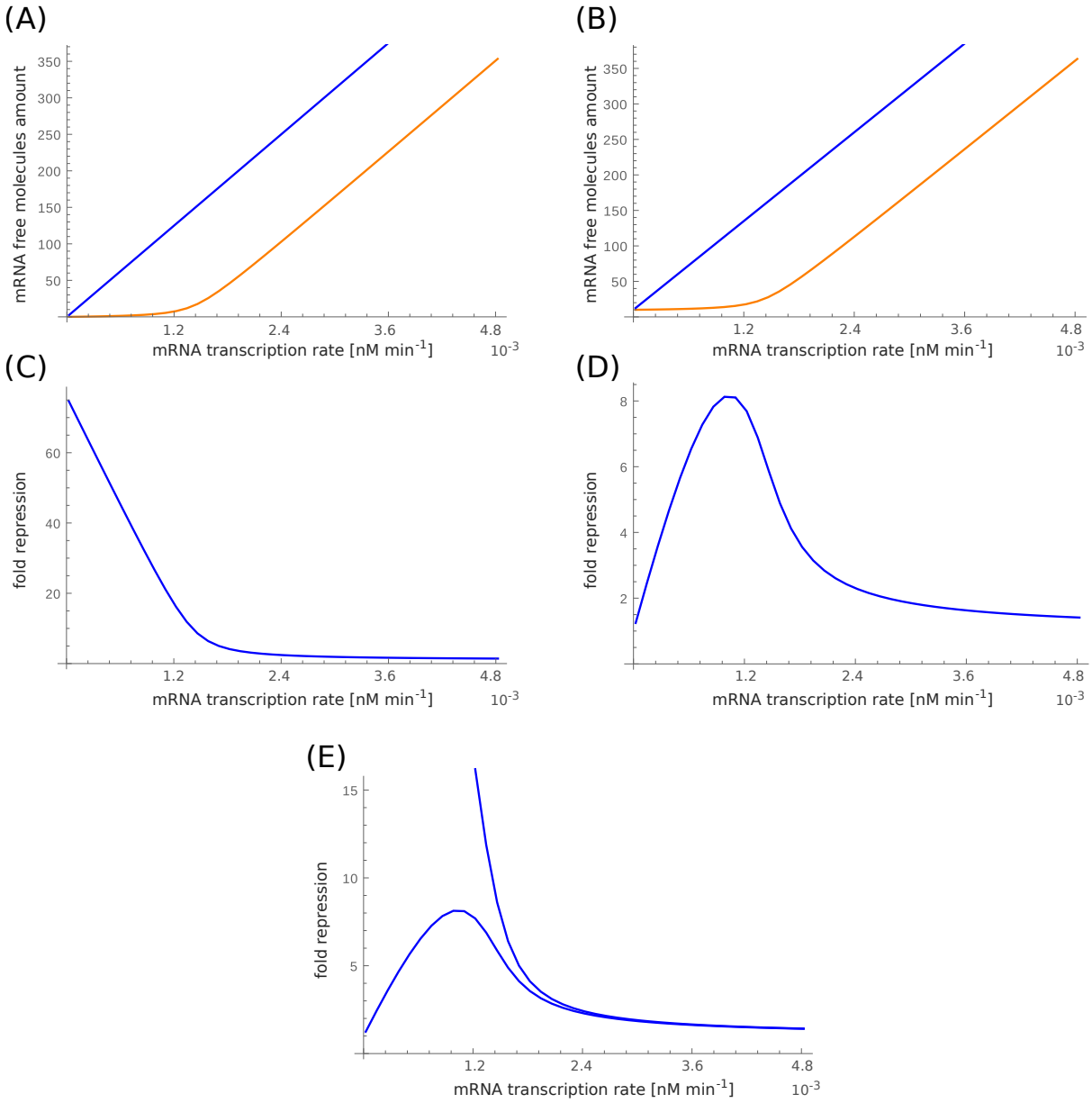

Fig. 3.19 The role of the offset. (A) Example of two average mRNA profiles, for a regulated (orange) and an unregulated (blue) mRNA. (B) Same profile as panel A but with the curves shifted upwards by an arbitrary offset of 10 mRNA molecules. (C) Fold repression (ratio of blue to orange curve from panel A) without the offset. (D) Fold repression (ratio of blue to orange curve from panel B) with the offset. (E) Comparison of the plots of fold repression with offset (from panel D) and without offset (from panel C). Adapted from [10].

for low values of constitutive expression, the fold repression, i.e. the ratio between the blue and the orange line in the plot, is high and then monotonically decreases (panel (C)). With an initial offset (panel (B)), the fold repression is equal to 1 in the beginning, then shows a maximum in proximity to the threshold and finally tends to a constant value (panel (D)). The two profiles, then, differ strongly, as shown in the comparison given in panel (E). The mechanism generating the profile is the same but the fold repression is an observable strongly affected by the free mRNA offset, which in fluorescence reporter experiments, is not a clearly controlled quantity. Since our model has no offset, the fold repression we measured should be taken as an upper bound to the ones normally obtained in experiments.

Figure 3.20 shows the bimodality region (red lines) for two sets of parameters, plotted against the mean amount of free mRNAs and proteins (Fig. 3.20A,C) and against the fold repression (Fig. 3.20B,D). As can be observed, in the bimodality region the mean values of free mRNAs are of the order of hundreds in Fig. 3.20A and of the order of tens in Fig. 3.20C and the fold repression ranges between 2 and 10. The amount of free miRNAs in this regime is of the order of tens, while its total amount (measured as the ratio between its transcription and degradation rates) is within 250 molecules per cell. These values are in agreement with experimental results in literature [138, 139] and suggest that titration interactions and extrinsic noise may play a role in inducing bimodal distributions also in endogenous regimes.

Another important aspect that has to be taken into account in order to check the validity of our theoretical predictions concerning bimodality, is the time scale of the reactions involved in the system. All the analysis described in previous sections was performed treating the extrinsic noise as a static quantity. Indeed, the miRNA trascription rate k_S was picked randomly from a Gaussian distribution for each realization of the system and its value was not allowed to vary in time. This approach is supposed to well approximate the case in which fluctuations of the value of k_S have a time scale much longer than the ones of the other reactions of the system. In endogenous conditions this is not necessarily the case, since changes of gene expression may happen on time scales of minutes or hours. Moreover, the time scale of the fluctuations of k_S should not be longer than the one set by the process of cell division. In the following, we will address these points and justify the use of our model to investigate the variability induced by cell cycle in the context of miRNA regulation.

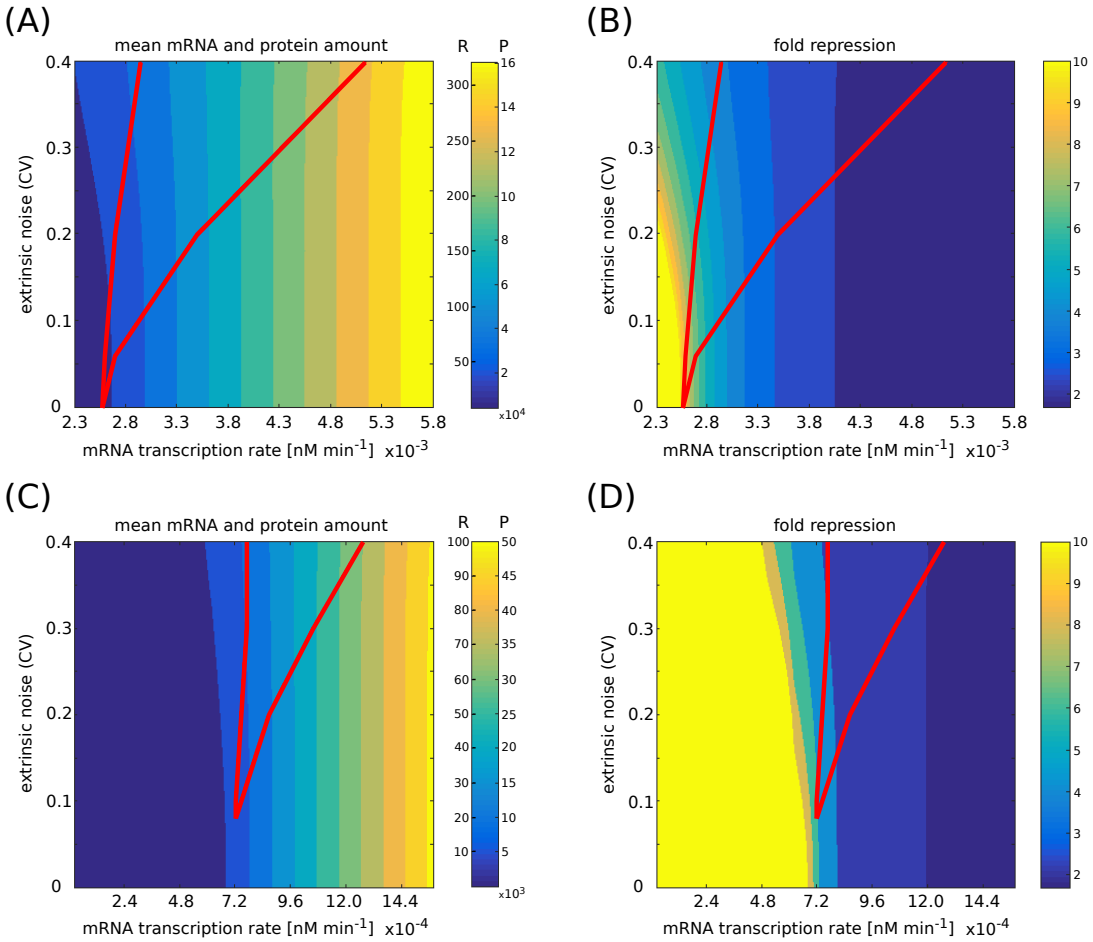

Fig. 3.20 Mean molecules amounts and fold repression. (A,C) Mean mRNA free amount (R) and protein amount (P) for two different sets of parameters, as a function of mRNA transcription rate and extrinsic noise level. The red line indicates the bimodality region. (B,D) Fold repression, i.e. ratio between the unregulated and regulated expression level, as a function of mRNA transcription rate and extrinsic noise level. The red line indicates the bimodality region. Mean values and fold repression are computed through the analytic approximation, while the bimodality region is obtained from numerical simulations. The set of parameters of panels (C) and (D) resembles an endogenous scenario, where the mean values of free mRNAs are of order of tens and the fold repression ranges between 2 and 6. Adapted from [10]

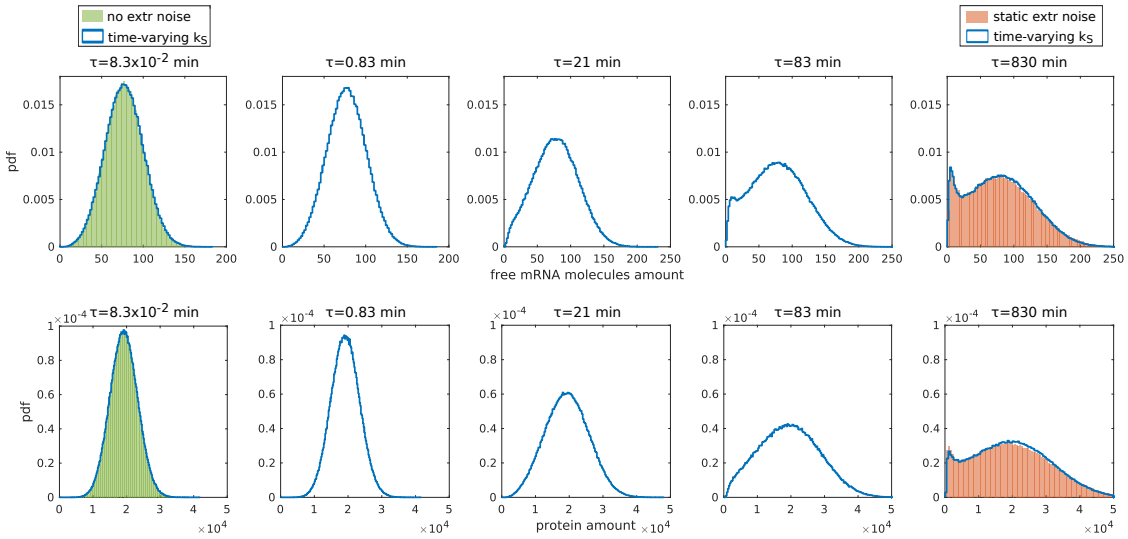

Fig. 3.21 Bimodality appearance in presence of time-dependent extrinsic noise. Solid cyan lines represent free mRNA molecules distributions with a time-dependent extrinsic noise on the miRNA transcription rate k_S. The transcription rate is coupled to a birth and death process with finite pool $N = 100$. The steady-state distribution of k_S is a nearly Gaussian distribution with mean $\overline{k}_S = 1.2 \times 10^{-3}$ nM min^{-1} and standard deviation $\sigma = 2.4 \times 10^{-4}$ nM min^{-1}. The time scale of this process is tuned by changing the values of the birth and death rates, keeping their ratio fixed. The time scales of the fluctuations of the miRNA transcription rate, from left to right are: 8.3×10^{-2} min, 0.83 min, 21 min, 83 min and 830 min. The green histogram in the leftmost plot represents the free mRNA molecules distribution in absence of extrinsic noise. The orange histogram in the rightmost plot represents the free mRNA molecules distribution with static extrinsic noise introduced as described in the main text; the k_S distribution used in this case is a Gaussian with mean $\overline{k}_S = 1.2 \times 10^{-3}$ nM min^{-1} and standard deviation $\sigma = 2.4 \times 10^{-4}$ nM min^{-1}. All the free mRNA molecules distributions are obtained from numerical simulations. Adapted from [10].

We first consider the constraints set by cell division on the degradation rates of the system. If we assume a doubling time of 24 hours, this implies a typical corresponding degradation rate of the order of $ln(2)/(24*60) \sim 4.8 \times 10^{-4}$ min^{-1}. The degradation rates we have used in the case studies, as reported in the legend of the figures, vary between 1.2×10^{-2} min^{-1} and 2.4×10^{-2} min^{-1} for most cases and, when investigating slow protein dynamics, are pushed to 2.4×10^{-3} min^{-1}. These degradations are at least one order of magnitude faster then the lower bound set by cell division, which is then, clearly, not a concern for our results.

In order to investigate the effects of extrinsic-noise temporal fluctuations, we first need to evaluate the typical time scales of the reactions involved in the system. One can estimate these time scales by looking at the rate equation and inserting the steady state concentrations. The rate at which one observes an appreciable change in a concentration is then obtained by dividing the propensities by the concentrations and the typical time scale is just its inverse. We chose a set of parameters for which bimodality was observed in case of static extrinsic noise, in particular the parameters used in Fig. 3.18A. In this case, the solution of the rate equation gives an mRNA concentration of $[R] = 0.03$ nM and a miRNA concentration of $[S] = 6.6 \times 10^{-4}$ nM. Note that these are just indicative values and differ slightly from the averages stemming from the numerical simulations with extrinsic noise. For the case of the miRNA-target interaction, we evaluated the time scales related to a change of concentration of both the miRNA and the mRNA. The values of the time scales are reported in Table 3.1.

To study the dynamic noise case, we performed Gillespie simulations allowing the transcription rate of the microRNA to fluctuate in time. We realised a dynamically fluctuating microRNA transcription rate via a birth and death process with finite pool $N = 100$. This is done via the auxiliary variable X that obeys the master equation:

$$\frac{dP(X)}{dt} = k_X(N-X+1)P(X-1) + g_X(X+1)P(X+1) - [k_X(N-X) + g_X X]P(X), \tag{3.64}$$

which has a binomial steady-state distribution with average 20 and standard deviation 4, closely resembling to a Gaussian. The rate k_S is obtained by multiplying $X * 6 \times 10^{-5}$ nM min^{-1} and has a steady-state distribution that closely approximates a Gaussian distribution with mean $\langle k_S \rangle = 1.2 \times 10^{-3}$ nM min^{-1} and standard devi-

reaction	time scale (min)
k_R	10
g_R	42
k_S	0.5
g_S	83
g (mRNA)	13
g (miRNA)	0.3
k_P	42
g_P	42
k_X	$8.3 \times 10^{-2} - 830$
g_X	$8.3 \times 10^{-2} - 830$
cell division	1440

Table 3.1 Time scales of all reactions involved in the network. Each reaction is identified by its corresponding rate.

ation $\sigma_{k_S} = 2.4 \times 10^{-4}$ nM min^{-1}, i.e. the same distribution from which we drew the rates in the static case. As discussed for the other reactions, the time scales of the extrinsic fluctuations can be estimated by dividing the average amount of the auxiliary species by the average propensities of birth and death at the steady state. This identifies a typical time scale on which the abundance of the auxiliary species X varies appreciably, $\tau = \frac{1}{g_X} = \frac{<X>}{(N-<X>)k_x}$. Keeping the ratio $k_X/g_X = 0.25$ nM fixed and varying k_X and g_X we can explore different time scales (τ) of the fluctuations of the microRNA transcription rates. To probe regimes in which these fluctuations are faster, comparable and slower than the typical time scales of the reactions we let τ take values of 8.3×10^{-2} min, 0.83 min, 21 min, 83 min and 830 min, see Fig. 3.21. These should be compared with the typical time scales of the various reactions at play in the system that are presented in Table 3.1. In case of fast dynamic noise ($\tau = 8.3 \times 10^{-2}$ min) the extrinsic fluctuations are averaged out and the resulting distribution coincides with the one with pure intrinsic noise. Conversely, with slow dynamic noise ($\tau = 830$ min) the resulting distributions are practically indistinguishable from the static extrinsic noise case treated in the previous sections. This time scale is comparable to (shorter than) 24 hours (1440 min). Then, we expect the results obtained for static noise to be relevant for settings in which the extrinsic noise is caused by variations along the cell cycle, in cells dividing every 24 hours. For intermediate time scales bimodality is modulated by the frequency of extrinsic fluctuations.

3.7 Implications and conclusions

In this long chapter we described the core of our work about the appearance of bimodal distributions, induced by extrinsic noise, in circuits based on the mechanism of molecular sequestration. We first introduced a minimal model composed of two molecular species that bind and unbind with constant rates. In this framework we were able to derive the analytic solution of the master equation and we showed that a normally distributed extrinsic source of noise, in terms of a fluctuating total amount of molecules of one species, can induce bimodal distributions. We then moved to a more complex model that describes the most important reactions involved in the process of miRNA-mediated gene regulation. We studied this system through an approximate analytic approach, based on the van Kampen's system-size expansion, and numerical simulations that implemented the Stochastic Simulation Algorithm. From this analysis we were able to characterize the rise of bimodal distributions, that in this system can appear also in absence of extrinsic noise. In particular, we highlighted the difference between single-cell bimodality and population bimodality induced by the extrinsic noise. The first one appears only in case of strong miRNA-mRNA coupling for a small range of the parameters, while the second one can appear also for small couplings in a wide range of values of the mRNA transcription rate. We then discussed the role played by the protein dynamics in reshaping, at the protein level, the associated distribution of mRNAs and showed that our results can be relevant in endogenous conditions.

In conclusion, if the goal of a biological system is to achieve two differentially expressed phenotypes across a population of identical cells, the combination of a mechanism inducing a threshold-like response, with some heterogeneity produced by extrinsic noise, can be an efficient strategy. Given the strong involvement of miRNAs in the process of cell differentiation and their ability to induce threshold responses, we can think that a similar mechanism could be involved. In our model, the source of extrinsic noise was a fluctuating miRNA transcription rate, but similar results are expected in case of fluctuations affecting the target. From an experimental perspective, it would be interesting to investigate networks related to cell differentiation that involve miRNAs, trying to find out whether their genes can be more transcriptionally noisy across cells than other classes of genes.

Chapter 4

Competition for the same sequester: crosstalk and correlations

In the previous chapter we analyzed the mechanism of molecular sequestration, focusing on its ability to give rise to bimodal distributions in combination to a source of extrinsic noise. We studied a minimal model composed of two molecular species that bind and form a complex which can then dissociate. Afterwards, we introduced a more complex model, based on molecular sequestration, describing the network of interactions involving a miRNA that regulates one target. Both the models involved only two molecular species: a sequester and a target. In this chapter we will extend this analysis to a system of two targets competing for the binding to the same sequester. Following the outline of the previous chapter, we will first extend the minimal model with the addition of one target species, then we will study a model of miRNA-mediated gene regulation with two genes competing for the same miRNA. Both the models will be studied in presence and absence of a Gaussian source of extrinsic noise. The analytic control on the minimal model will allow us to study in detail the behavior of the correlations induced by the competition between the targets. Furthermore, the model of miRNA-target interaction will give us insights on how bimodal distributions can be tuned by the crosstalk between the targets.

Part of the results discussed in this chapter have already been published in [9] and [10].

4.1 Minimal model of competition for the same sequester

As a first step towards the study of the implications of the competition between two targets for the same sequester, we will here introduce a minimal model of competition for a limiting resource that will be completely characterized analytically.

The model is composed of two molecular species, T_1 and T_2, that compete for binding to a third one, the sequester S. The species T_1 binds S with rate k_+^1, forming a complex $\overline{T_1S}$ that dissociates with rate k_-^1. The species T_2 in turn binds S with rate k_+^2 and forms a complex $\overline{T_2S}$ that dissociates with rate k_-^2.

The reactions that define the model are then:

$$T_1 + S \underset{k_-^1}{\overset{k_+^1}{\rightleftharpoons}} \overline{T_1S}\,, \tag{4.1}$$

$$T_2 + S \underset{k_-^2}{\overset{k_+^2}{\rightleftharpoons}} \overline{T_2S}\,. \tag{4.2}$$

As done previously, the total amount of molecules of each species (T_{1T}, T_{2T} and S_T) is assumed to be constant. This leads to the following conservation laws:

$$T_{1T} = T_1 + \overline{T_1S} = const\,, \tag{4.3}$$

$$T_{2T} = T_2 + \overline{T_2S} = const\,, \tag{4.4}$$

$$S_T = S + \overline{T_1S} + \overline{T_2S} = const\,. \tag{4.5}$$

As a consequence of the conservation laws, the number of independent variables of the system is limited to 2 instead of 5. In the following, we will use T_1 and S in order to describe the system. Indeed, the number of free molecules of the species T_2 is obtained as $T_2 = T_{2T} - S_T + S + T_{1T} - T_1$, while the number of complexes is obtained inverting the conservation laws.

In the following section we will present the exact analytic solution of the master equation associated to this system.

4.1.1 Stochastic approach with analytic solution

As in the case with a single target, stochastic fluctuations play an important role in this system. In order to analytically investigate the behavior of these fluctuations and of the correlation between the targets, we need to adopt a stochastic approach.

The master equation that describes the time evolution of the probability distribution of observing T_1 and S free molecules at time t is:

$$\begin{aligned}\frac{dP(T_1,S,t)}{dt} &= (\overline{T_1S}+1)k_-^1P(T_1-1,S-1)+(T_1+1)(S+1)k_+^1P(T_1+1,S+1)+\\ &+(\overline{T_2S}+1)k_-^2P(T_1,S-1)+(T_2+1)(S+1)k_+^2P(T_1,S+1)+\\ &-(\overline{T_1S}k_-^1+T_1Sk_+^1+\overline{T_2S}k_-^2+T_2Sk_+^2)P(T_1,S)\,. \end{aligned} \tag{4.6}$$

By exploiting the conservation laws, eqs. (4.3), (4.4) and (4.5), we can write it in terms of T_1 and S only:

$$\begin{aligned}\frac{dP(T_1,S,t)}{dt} &= (R_{1T}-T_1+1)k_-^1P(T_1-1,S-1)+(T_1+1)(S+1)k_+^1P(T_1+1,S+1)+\\ &+(S_T-S-T_{1T}+T_1+1)k_-^2P(T_1,S-1)+\\ &+(T_{2T}-S_T+S+T_{1T}-T_1+1)(S+1)k_+^2P(T,S+1)+\\ &-[(T_{1T}-T_1)k_-^1+T_1Sk_+^1+(S_T-S-T_{1T}+T_1)k_-^2+\\ &+(T_{2T}-S_T+S+T_{1T}-T_1)Sk_+^2]P(T_1,S)\,. \end{aligned} \tag{4.7}$$

For this circuit of interactions detailed balance holds. Indeed, by considering a closed loop that links together four different states of the system we see that the clockwise product of the transition probabilities is equal to the counterclockwise one:

$$\frac{k_-^1(T_{1T}-T_1)k_+^2(S+2)(T_{2T}-S_T+S+T_{1T}-T_1+1)}{k_+^2(S+1)(T_{2T}-S_T+S+T_{1T}-T_1+1)k_-^1(T_{1T}-T_1)}\cdot$$
$$\cdot\frac{k_+^1(T_1+1)(S+1)k_-^2(S_T-S-T_{1T}+T_1)}{k_-^2(S_T-S-T_{1T}+T_1)k_+^1(T_1+1)(S+2)}=1\,.$$

Given detailed balance, each reaction is balanced by its reverse one, then the following relations hold:

$$k_-^1(T_{1T}-T_1)P(T_1,S)=k_+^1(T_1+1)(S+1)P(T_1+1,S+1)\,, \tag{4.8}$$

$$k_-^2(S_T-T_{1T}-S+T_1)P(T_1,S)=k_+^2(S+1)(T_{1T}+T_{2T}-S_T+S-T_1+1)P(T_1,S+1)\,. \tag{4.9}$$

Dividing eq. (4.9) by $P(T_1)$ we obtain a recursive relation for the probability of observing S free molecules conditioned on having a generic number T_1 of free molecules:

$$P(S+1|T_1)=\frac{k_-^2}{k_+^2}\frac{S_T-T_{1T}-S+T_1}{(S+1)(T_{1T}+T_{2T}-S_T+S-T_1+1)}P(S|T_1)\,. \tag{4.10}$$

In this system, the minimal number of free molecules of S can either be $S_{min}=0$ if $S_T\leq T_{1T}-T_1+T_{2T}$, or $S_{min}=S_T-T_{1T}+T_1-T_{2T}$ when $S_T>T_{1T}-T_1+T_{2T}$. In the latter case, even if all T_2 molecules are bound, there is a finite number of free S molecules. Eq. (4.10) can be used to recursively write the expression of the conditional probability $P(S|T_1)$ starting from the probability $P(S_{min}|T_1)$:

$$P(S|T_1)=\left(\frac{k_-^2}{k_+^2}\right)^{S-S_{min}}\frac{(S_T-T_{1T}+T_1-S_{min})!}{(S_T-T_{1T}+T_1-S)!}\frac{S_{min}!}{S!}\cdot$$
$$\cdot\frac{(S_{min}+T_{1T}+T_{2T}-S_T-T_1)!}{(S+T_{1T}+T_{2T}-S_T-T_1)!}P(S_{min}|T_1)\,. \tag{4.11}$$

We can now write the first relation of detailed balance (eq. 4.8) in terms of the conditional probability $P(S|T_1)$:

$$k_-^1(T_{1T}-T_1)P(S|T_1)P(T_1) = k_+^1(T_1+1)(S+1)P(S+1|T_1+1)P(T_1+1)\,. \quad (4.12)$$

Defining $S_{min} \equiv S^*_{min} + T_1$ and inserting the expression of the conditional probability in equation 4.12, the resulting relation can be made explicit for the marginal probability $P(R_1+1)$ thus obtaining:

$$P(T_1+1) = \frac{k_-^1}{k_+^1}\frac{T_{1T}-T_1}{(T_1+1)(S^*_{min}+T_1+1)}\frac{P(S^*_{min}+T_1|T_1)}{P(S^*_{min}+T_1+1|T_1+1)}P(T_1)\,. \quad (4.13)$$

By iterating eq. (4.13), we can write the analytic expression of the marginal probability $P(T_1)$:

$$P(T_1) = \left(\frac{k_-^1}{k_+^1}\right)^{T_1-T_{1min}}\frac{(T_{1T}-T_{1min})!T_{1min}!(S^*_{min}+T_{1min})!}{(T_{1T}-T_1)!T_1!(S^*_{min}+T_1)!}\frac{P(S^*_{min}+T_{1min}|T_{1min})}{P(S^*_{min}+T_1|T_1)}P(T_{1min})\,, \quad (4.14)$$

where $P(T_{1min})$ is the probability of having the minimal number allowed of free T_1. T_{1min} is defined by the total amount of S and T_1. The two possible cases can be either $T_{1min}=0$ when $S_T \geq T_{1T}$, which means that in principle all the molecules of T_1 can be bound to molecules of S, or $T_{1min} = T_{1T} - S_T$ when $S_T < T_{1T}$, meaning that even if all the molecules of S are bound to molecules of T_1, T_{1min} molecules are free. When $T_{1min} > 0$ we have that $S^*_{min} + T_{1min} = S_T - T_{1T} - T_{2T} + T_{1T} - S_T = -T_{2T} \leq 0$, then the conditional probability at the numerator in eq. (4.14) is $P(0|T_{1min})$. Conversely, if $T_{1min} = 0$, the conditional probability can be either $P(0|T_{1min})$ or $P(S^*_{min}|T_{1min})$, depending on the sign of S^*_{min}.

The probability $P(T_1,S)$ can be reconstructed by using the definition of joint probability $P(T_1,S) = P(S|T_1)P(T_1)$, which leads to the following expression:

$$P(T_1,S) = \left(\frac{k_-^1}{k_+^1}\right)^{T_1}\left(\frac{k_-^2}{k_+^2}\right)^{S-T_1}\frac{P(T_{1min},S^*_{min}+T_{1min})}{(S_T-T_{1T}+T_1-S)!S!(T_{1T}+T_{2T}-S_T-T_1+S)!(T_{1T}-T_1)!T_1!}\,, \quad (4.15)$$

with $P(T_{1min}, S^*_{min} + T_{1min})$ given by the normalization of the probability distribution:

$$P(T_{1min}, S^*_{min} + T_{1min}) = \sum_{T_1=T_{1min}}^{T_{1T}} \sum_{S=S^*_{min}+T_{1min}}^{S_T} \left(\frac{k^1_+}{k^1_-}\right)^{T_1} \left(\frac{k^2_+}{k^2_-}\right)^{S-T_1} (S_T - T_{1T} + T_1 - S)!S!\cdot$$
$$\cdot (T_{1T} + T_{2T} - S_T - T_1 + S)!(T_{1T} - T_1)!T_1!\,. \tag{4.16}$$

Also for the joint probability, if $T_{1min} > 0$, then $S^*_{min} + T_{1min} \equiv 0$, while when $T_{1min} = 0$, the value of S^*_{min} is given by its definition $S^*_{min} \equiv S_T - T_{1T} - T_{2T}$, being zero when the result is negative.

The solution of the master equation that we have obtained, eq. (4.15), can be written in terms of T_1 and T_2 by implementing the substitution $S = T_2 + T_1 - T_{2T} - T_{1T} + S_T$:

$$P(T_1, T_2) = \frac{(K_1)^{T_1}(K_2)^{T_2}}{N} \frac{\binom{T_{1T}}{T_1}\binom{T_{2T}}{T_2}}{[S_T + T_1 + T_2 - T_{1T} - T_{2T}]!}\,, \tag{4.17}$$

where we introduced the dissociation constants: $K_1 \equiv \frac{k^1_-}{k^1_+}$, $K_2 \equiv \frac{k^2_-}{k^2_+}$. The normalization factor N is given by:

$$N = \sum_{T_2=\max(0,T_{2T}-S_T)}^{T_{2T}} \sum_{T_1=\max(0,T_{2T}+T_{1T}-S_T-T_2)}^{T_{1T}} \frac{(K_1)^{T_1}(K_2)^{T_2}\binom{T_{1T}}{T_1}\binom{T_{2T}}{T_2}}{[S_T + T_1 + T_2 - T_{1T} - T_{2T}]!}\,. \tag{4.18}$$

The minimal values that T_1 and T_2 can assume in the joint probability $P(T_1, T_2)$ are defined by the total amount of sequester S_T: $T_{1min} = \max(0, T_{1T} - S_T)$ and $T_{2min} = \max(0, T_{2T} - S_T)$. Nonetheless, when the number of free molecules of one of the two targets, for example T_1, is chosen, the range of values that T_2 can assume is limited by the fact that there are only $S_T - T_{1T} + T_1$ molecules of the sequester S available for binding. In this case we have $T_{2min} = \max(0, T_{2T} + T_{1T} - S_T - T_1)$. Analogously, when the value of T_2 is chosen, the minimal value that T_1 can assume is: $T_{1min} = \max(0, T_{2T} + T_{1T} - S_T - T_2)$.

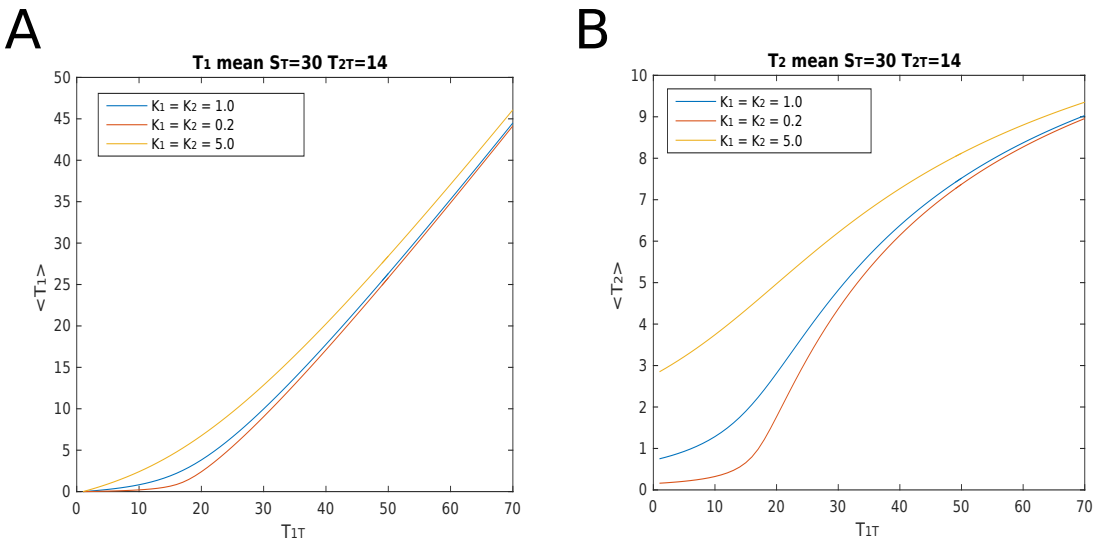

Fig. 4.1 Mean amount of free targets. (A) Mean number of molecules of target T_1 as a function of T_{1T}. (B) Mean number of molecules of target T_2 as a function of T_{1T}. $S_T = 30$ $T_{2T} = 14$.

In the following sections, the probability distribution, eq. (4.17), will be used to analyze the stochastic properties of the system.

4.1.2 Threshold and crosstalk

The analytic solution of the master equation, eq. (4.17), can be used to investigate the properties of the minimal model.

As a first step, we can check the behavior of the mean amount of free molecules of the two targets. As expected, the threshold-like behavior of the mean is maintained, see figure 4.1. The position of the threshold is located at the equimolarity point between the two targets and the sequester, i.e. when $T_{1T} + T_{2T} = S_T$. The steepness of the thresholds of the two target species and the ultrasensitivity of their response are controlled by the value of their dissociation constants K_1 and K_2. The smaller the value of the dissociation constant, the steeper the threshold. Indeed, when the dissociation constant is small, meaning that the binding rate is way higher than the unbinding one, the repression below threshold is strong and diminishes only after the equimolarity point. Our minimal model reproduces the phenomenon of crosstalk that we introduced in section 2.2.4 for a circuit of miRNA-regulated genes (ceRNA

effect). Indeed, as can be seen in figure 4.1, the amount of free molecules of a target whose total amount is kept constant, can be tuned by changing the total amount of molecules of the second target. This indirect interaction is mediated by the binding to the sequester.

In the following sections we will investigate in detail the behavior of the correlation induced by this indirect interaction between the two targets.

4.2 Correlations between competitors with and without extrinsic noise

The most important feature of this minimal model of molecular sequestration and competition is the fact that it is fully under analytic control. This feature allows us to study and characterize the coupling induced by the competition for the limiting resource, without the need of numerical simulations. In the following sections we will study this coupling in terms of the linear correlation and mutual information between the two competing targets. We will show that the presence of an extrinsic source of noise strongly influences the shape of the correlation and can lead to a change in its sign.

4.2.1 Pearson correlation coefficient

In the previous section we showed that the competition for the common sequester induces crosstalk between the two targets. Here, we continue the study of this coupling by analyzing the correlation induced by this kind of interaction. For this purpose, we first compute the Pearson correlation coefficient between the targets. This estimator measures the linear correlation between two variables, in our case the amounts of free molecules T_1 and T_2. It is defined as:

$$\rho_{T_1,T_2} = \frac{\sigma_{T_1T_2}}{\sigma_{T_1}\sigma_{T_2}}\,, \tag{4.19}$$

where $\sigma_{T_1T_2}$ is the covariance between T_1 and T_2 and σ_{T_1}, σ_{T_2} the standard deviations of T_1 and T_2 respectively.

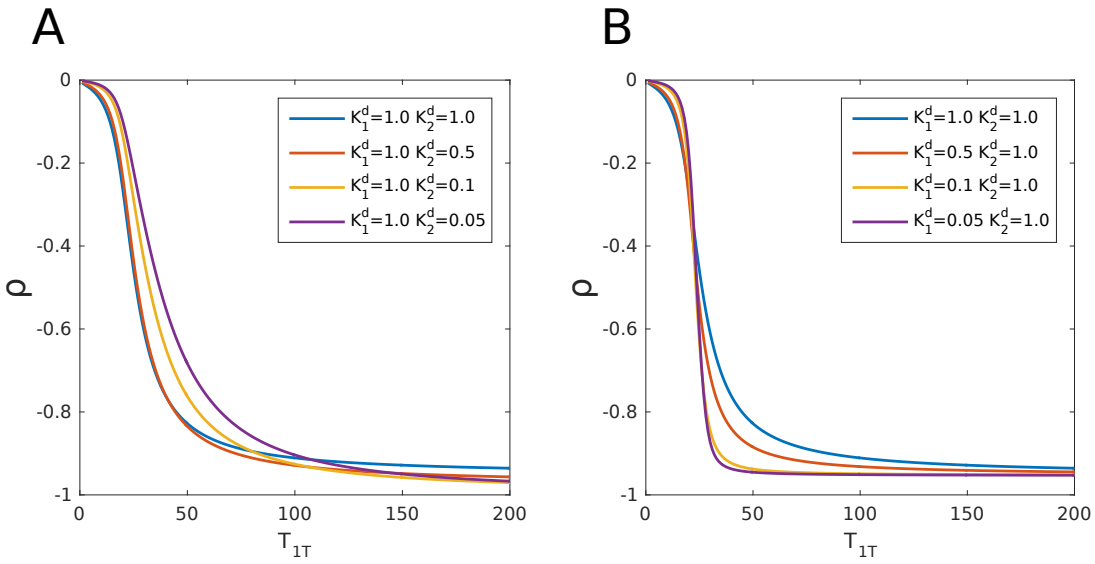

Fig. 4.2 Pearson correlation coefficient between T_1 and T_2 as a function of T_{1T} for different values of the dissociation constants. $T_{2T} = 10$ $S_T = 30$. (A) $K_2 = 1.00$. (B) $K_1 = 1.00$. Adapted from [9].

Let us consider the case in which the total number of target molecules is bigger than the one of the sequester. In this condition, almost all the sequesters are bound. This means that an increase in the number of bound molecules of one target implies a decrease of the number of bound molecules of the other target. As a consequence, in this regime the induced correlation between the two targets is negative. The Pearson correlation coefficient, as a function of the total amount of molecules of one target, assumes a sigmoidal shape, ranging from 0 to an asymptotic value of negative correlation, with the drop in proximity to the threshold, see figure 4.2. Focusing on the profile of the Pearson correlation as a function of T_{1T}, we observe that the two dissociation constants have a different influence on it. Indeed, the dissociation constant of the first target, K_1, controls only the steepness of the drop: as K_1 is decreased the profile becomes steeper. Differently, the dissociation constant of the second target, K_2, sets the lower bound of the correlation. Indeed, by decreasing K_2, the asymptotic value that the correlation reaches for large values of T_{1T} gets closer to -1. Furthermore, as K_2 is decreased, the drop becomes smoother and its location less defined. This kind of profile is conserved upon varying the total amount of molecules T_{2T} of the second target. Indeed, figure 4.3 shows the contour plot of the Pearson correlation coefficient as a function of both T_{1T} and T_{2T}. As can be seen,

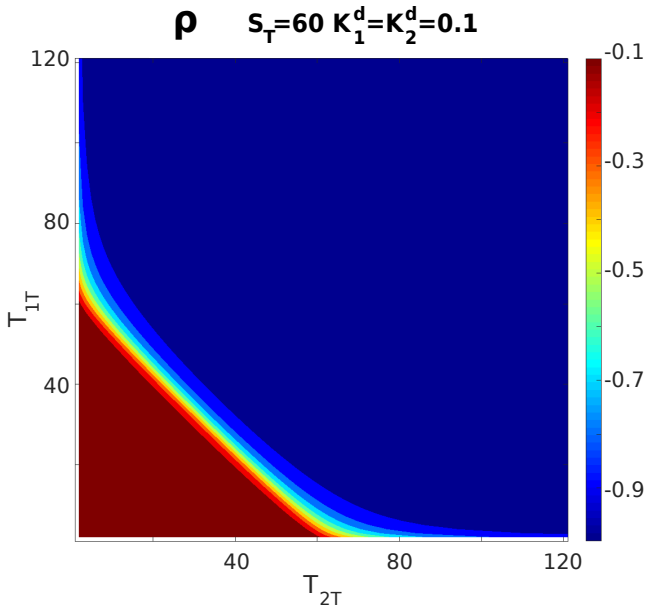

Fig. 4.3 Contour plot of the Pearson correlation coefficient between T_1 and T_2 as a function of T_{1T} and T_{2T}. $S_T = 60$ $K_1 = K_2 = 0.1$. Adapted from [9].

no local or global maxima are present and the anticorrelation between the targets increases with their total amount, asymptotically reaching the value -1.

Let us now explore the dependence of the correlation on the two dissociation constants, by keeping fixed the total amounts of molecules T_{1T} and T_{2T}. Let us focus on the case in which the total amount of molecules of one target and its dissociation constant are fixed (*e.g.* K_1 and T_{1T} fixed). By varying the dissociation constant of the other target (in this case K_2), the correlation profile can be non monotonic, displaying a minimum, see figure 4.4 (A). We shall refer to the value of the dissociation constant that minimizes the Pearson correlation as ${K_2}^*$. As can be seen in figure 4.4 (B), the value of ${K_2}^*$ depends linearly on the fixed dissociation constant of the competing target (K_1) and the slope is set by the total amount of target molecules T_{2T}. Keeping S_T fixed, the slope decreases as T_{2T} increases and vanishes when $T_{2T} = S_T$. In the regime in which $T_{2T} \geq S_T$, the correlation profile is monotonic as a function of the dissociation constant and its lowest value is reached for vanishing K_2, regardless of the value of K_1. This means that the minimum of the correlation is obtained for the highest affinity between the second target and the sequester.

Besides the slope of the linear dependence on K_1 of the minimum position ${K_2}^*$, the total amount of target molecules T_{2T} influences also the offset. When

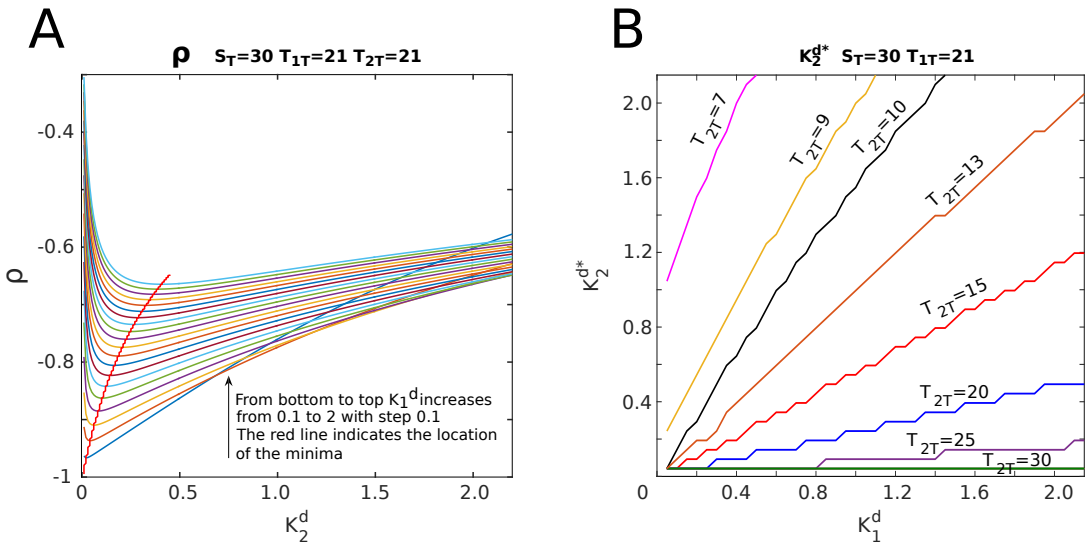

Fig. 4.4 (A) Pearson correlation coefficient between T_1 and T_2 as a function of K_2. $S_T = 30$ $T_{1T} = T_{2T} = 21$. The red line indicates the location of the minima. (B) Position of the minimum $K_2{}^*$ of the Pearson correlation coefficient as a function of K_1. $S_T = 30$ $T_{1T} = 21$, different colors indicate different values of T_{2T} (see legend). Adapted from [9].

$T_{1T} + T_{2T} \leq S_T$ at $K_1 = 0$, the correlation is minimized by a finite value of $K_2{}^*$. Conversely, for $T_{1T} + T_{2T} > S_T$ the offset disappears and a vanishing K_1 corresponds to a vanishing $K_2{}^*$. In this region of the parameters, the minimum of the correlation is reached for the highest affinity of both the targets. The existence of the offset in the position of $K_2{}^*$ as a function of K_1 indicates the presence of two sub-regimes characterized by a different joint dependence on the two dissociation constants K_1 and K_2. What determines these regimes is the total number of target molecules in the system $T_{1T} + T_{2T}$, compared to the total number of sequester molecules S_T. When the offset is present, for $T_{1T} + T_{2T} \leq S_T$, there are always some free molecules of the sequester and both the targets are in the repressed state. Nonetheless, these two targets can be correlated and their correlation presents a global minimum as a function of the two dissociation constants, see figure 4.5 (A,B). The presence of a global minimum for finite values of the dissociation constants justifies the existence of the offset in the 1-dimensional plots. For systems in which the number of target molecules is small, the global minimum can be located at relatively large values of the dissociation constants, this meaning a condition of weak interaction between the targets and the sequester.

As an example, let us consider again figure 4.5. In the first regime, for the case in which $T_{1T} = T_{2T} = 15$ and $S_T = 30$ (Fig. 4.5B), we see that the minimum of

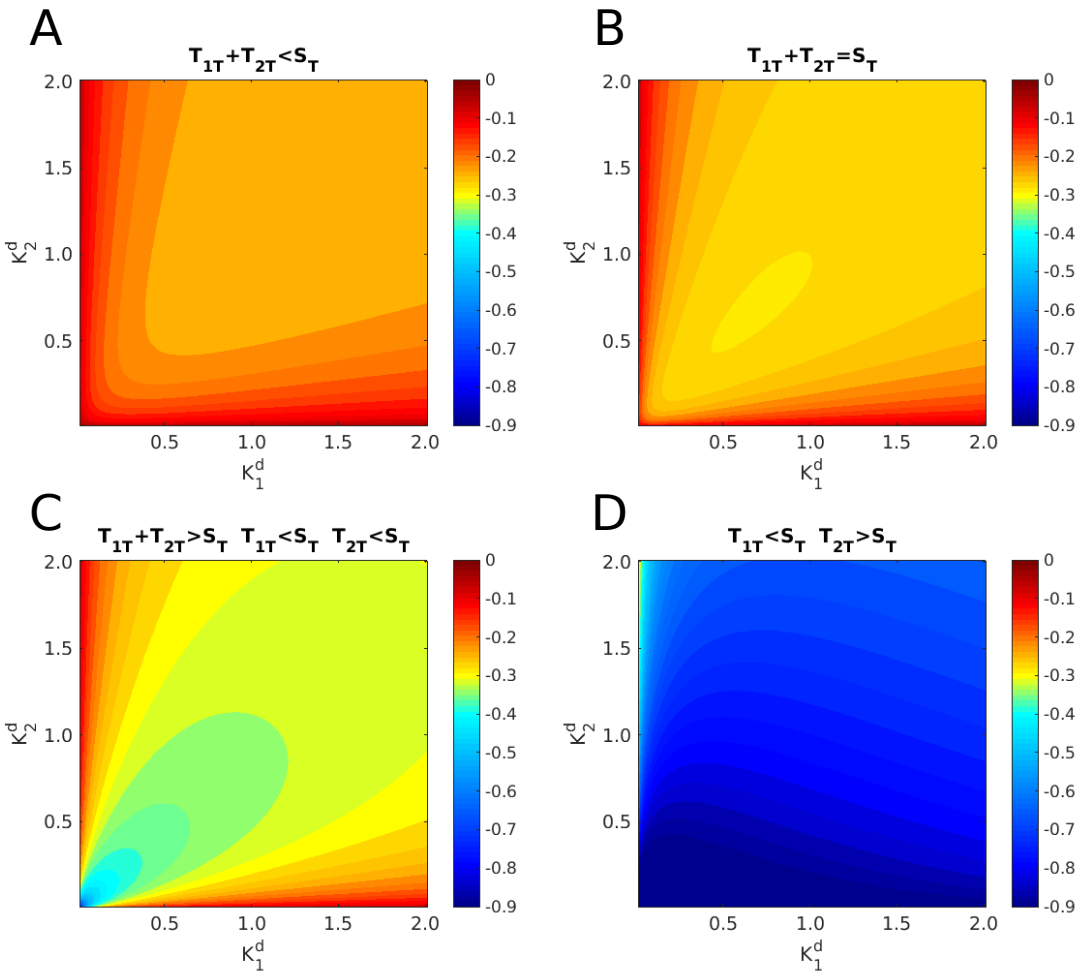

Fig. 4.5 Contour plot of the Pearson correlation coefficient between T_1 and T_2 as a function of K_1 and K_2, $S_T = 30$. (A) $T_1 = 15$ and $T_2 = 14$. (B) $T_1 = T_2 = 15$. (C) $T_1 = 15$ and $T_2 = 16$. (D) $T_1 = 15$ and $T_2 = 31$. Adapted from [9].

the correlation is present and is obtained for values of the dissociation constants for which the average number of free target is low $\langle T_1 \rangle = \langle T_2 \rangle \simeq 2$ and $\langle S \rangle \simeq 4$. Moving into the second regime, where $T_{1T} + T_{2T} \geq S_T$ (Fig. 4.5C), we see that there is no global minimum and correlation is lower for lower values of the dissociation constants. Nonetheless, even when the global minimum is absent, considering slices of the contour plot (*e.g.* for K_1 fixed) the correlation still presents a minimum if $T_{2T} < S_T$. Finally, when $T_{2T} > S_T$, the local minimum in the 1-D plot of the correlation as a function of K_2 is lost and the minimal value of correlation is reached for vanishing K_2.

4.2.2 Mutual information

By characterizing the Pearson correlation coefficient, we limited our study to the linear correlation between the two targets. In order to investigate more deeply this coupling, we can compute the mutual information between them. Mutual information (MI) is a measure of the amount of information that can be obtained on one quantity, normally unaccessible, by measuring another quantity which is instead experimentally accessible. Differently from the Pearson correlation coefficient, MI is able to capture the non-linear behavior of the interaction between two correlated quantities. It is then considered more reliable in order to infer the whole properties of the experimentally unaccessible quantity.

MI is defined as the Kullback-Leibler divergence between the joint probability of two random variables and the product of their marginal probabilities [140]:

$$I(X,Y) = D_{KL}(P(x,y)||P(x)P(y)) = \sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)} \,. \tag{4.20}$$

It measures the statistical dependence of the two variables X and Y, which are assumed to be discrete. When the logarithm to base 2 is used, MI is measured in bits. Moreover, this quantity is symmetric and always non-negative.

Given the above definition, the MI between the free amounts of molecules T_1 and T_2 of the two targets is:

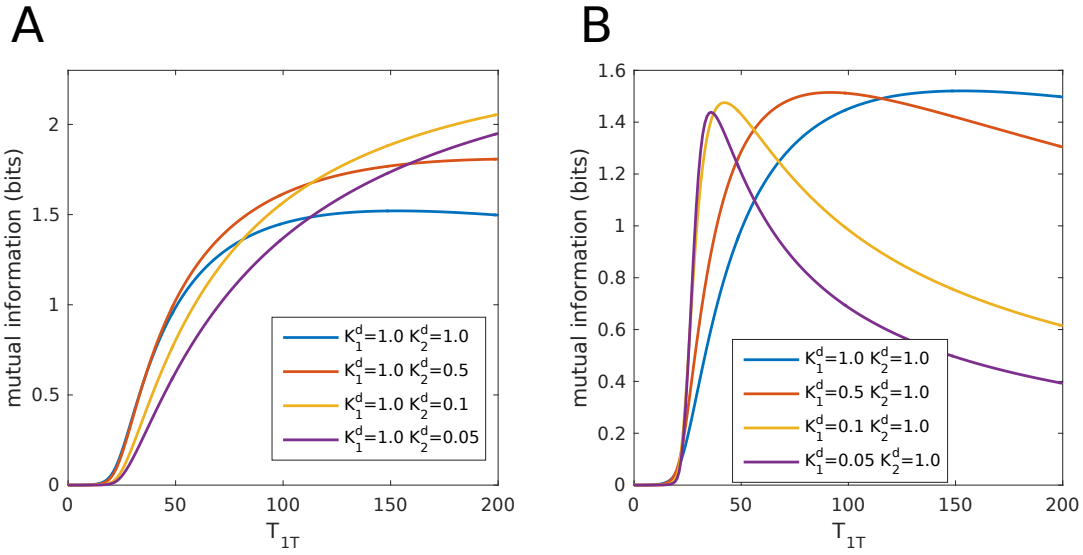

Fig. 4.6 Mutual information between T_1 and T_2 as a function of T_{1T} for different values of the dissociation constants. $T_{2T} = 10$ $S_T = 30$. (A) $K_1 = 1.00$. (B) $K_2 = 1.00$. Adapted from [9].

$$I(T_1,T_2) = \sum_{T_1,T_2} P(T_1,T_2) \log \frac{P(T_1,T_2)}{P(T_1)P(T_2)}, \tag{4.21}$$

where $P(T_1,T_2)$ is the joint probability distribution, eq. (4.17), and $P(T_1) = \sum_{T_2} P(T_1,T_2)$ and $P(T_2) = \sum_{T_1} P(T_1,T_2)$ the marginal ones.

As we did for the Pearson correlation coefficient, we first compute the MI, eq. (4.21), as a function of the total amount of molecules of one of the targets, namely T_1. As can be seen in figure 4.6, the MI presents a maximum that can either be very narrow or extremely wide, resembling a plateau, depending on the values of the two dissociation constants. In particular, keeping K_2 fixed, a decrease of K_1, meaning a stronger affinity between the first target and the sequester, favors the rise of the peak. This maximum gets closer to the threshold value ($S_T - T_{2T}$) and narrower as the value of the dissociation constant becomes smaller. Furthermore, the maximal value of MI slightly decreases as K_1 gets smaller. On the other hand, keeping K_1 fixed, a decrease of K_2 makes the peak extremely wide, moves it towards higher values of T_{1T} and increases the maximal value of MI. This effect is similar to what we showed for the Pearson correlation. Indeed, in this case the dissociation constant K_2 is the main regulator of the maximal amount of information that can be achieved, while in the Pearson case it controlled the asymptotic value of correlation.

In addition to that, also for the MI the drop becomes smoother as K_2 is decreased, while it becomes steeper as K_1 is decreased. The main qualitative difference with the behavior of the Pearson coefficient is the presence of the local maximum. Indeed, we can directly compare the MI to the Pearson correlation, by assuming that the joint probability distribution is a bivariate Gaussian. In this case, all the MI is enclosed in the correlation and the relation that connects the two quantities is:

$$I(X,Y) = -\frac{1}{2}\log(1-\rho^2_{X,Y})\,. \tag{4.22}$$

By exploiting eq. (4.22) we plotted the Pearson correlation in units of MI and compared it to the real MI. The results are shown in figure 4.7. We can see that the profile of the two quantities is similar, but the peak of MI is not reproduced. The position of this peak, in terms of T_{1T}, scales linearly with K_1 and almost linearly with the inverse of K_2. Moreover, the value of this peak is almost stationary as K_1 is varied (a part from the case of very small K_1), while it decreases when K_2 gets bigger, see figure 4.8. The reason of the existence of this peak can be related to the marked non-Gaussianity of the joint target distribution after saturation ($T_{1T}+T_{2T} > S_T$). In this region the distribution gets strongly peaked, a scenario which has been related to instances of MI lower than the one predicted by Gaussian approximations [141].

As an additional comparison between the quantities, in figure 4.9 we show the contour plots representing the value of MI and correlation (in MI units) as a function of the total amount of both the targets. In this 2-D case, no global or local maxima are found, neither for the correlation, nor for the mutual information. Nonetheless, we can recover the 1-D non-monotonic behavior of the MI by taking "slices" of the contour plot with T_{2T} fixed.

So far, we investigated the behavior of the MI as a function of the total amount of molecules of the two targets. In the following, we will focus our attention on its dependence on the value of the dissociation constants, keeping T_{1T} and T_{2T} fixed. As in the case of the Pearson correlation, by keeping fixed one of the dissociation constants (*e.g.* K_1) and varying the other one (here K_2) , the MI presents a maximum for a finite value of K_2, named ${K_2}^*$ (it was a minimum for the correlation). With respect to the Pearson, the minum ${K_2}^*$ scaled linearly with the dissociation constant K_1. In this case, there is a nonlinear dependence for small values of K_1, and a linear dependence as K_1 increases, see figure 4.10. The values of K_2 corresponding to the

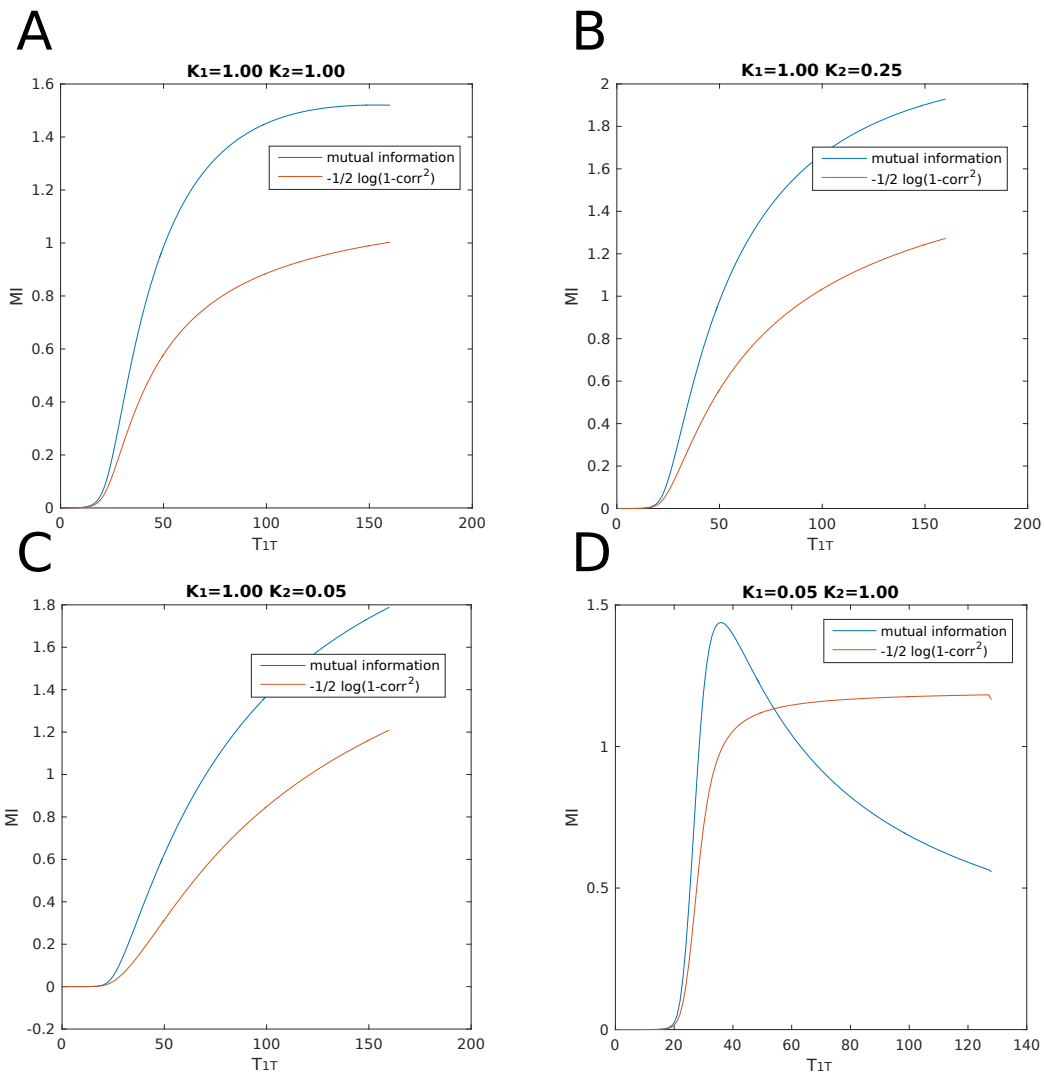

Fig. 4.7 Comparison between the Mutual Information (MI), blue line, and the Pearson correlation coefficient, orange line, (converted in MI units through eq. (4.22)). $T_{2T} = 10$ $S_T = 30$.

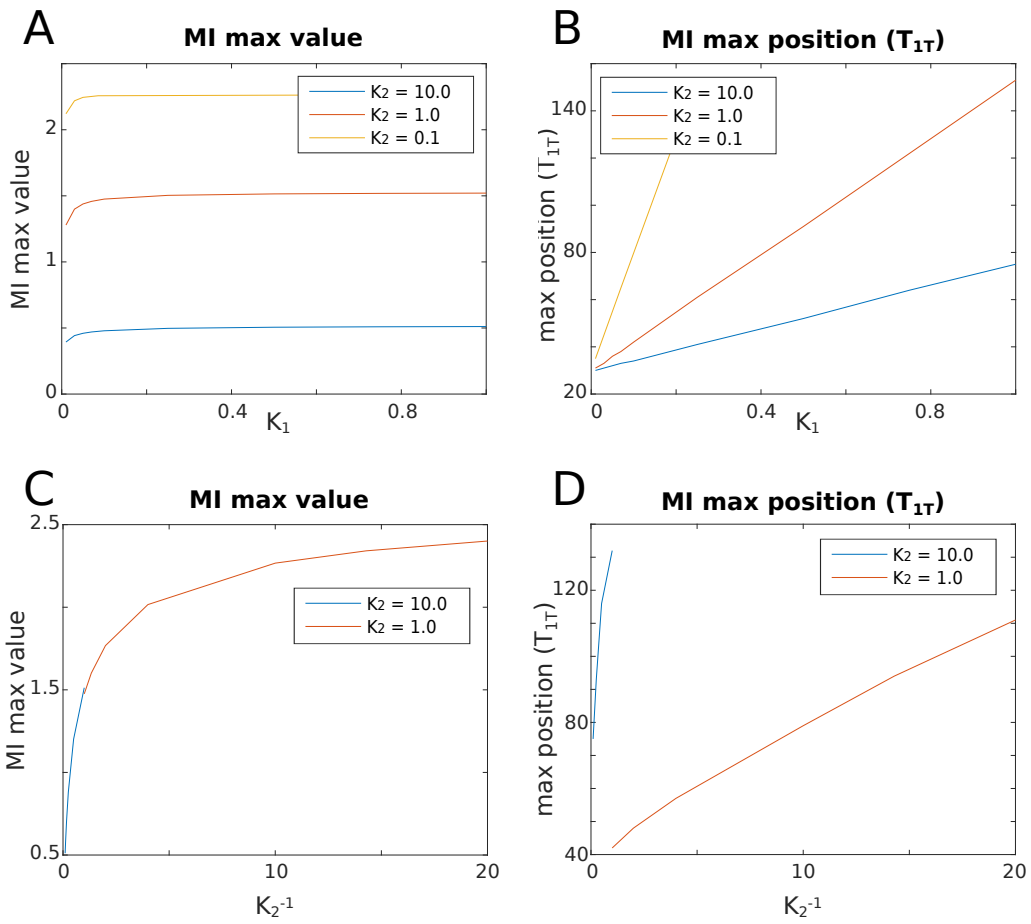

Fig. 4.8 Mutual information (MI) maximum value (A,C) and position in terms of T_{1T} (B,D) as a function of the dissociation constants. $S_T = 30$ $T_{2T} = 10$.

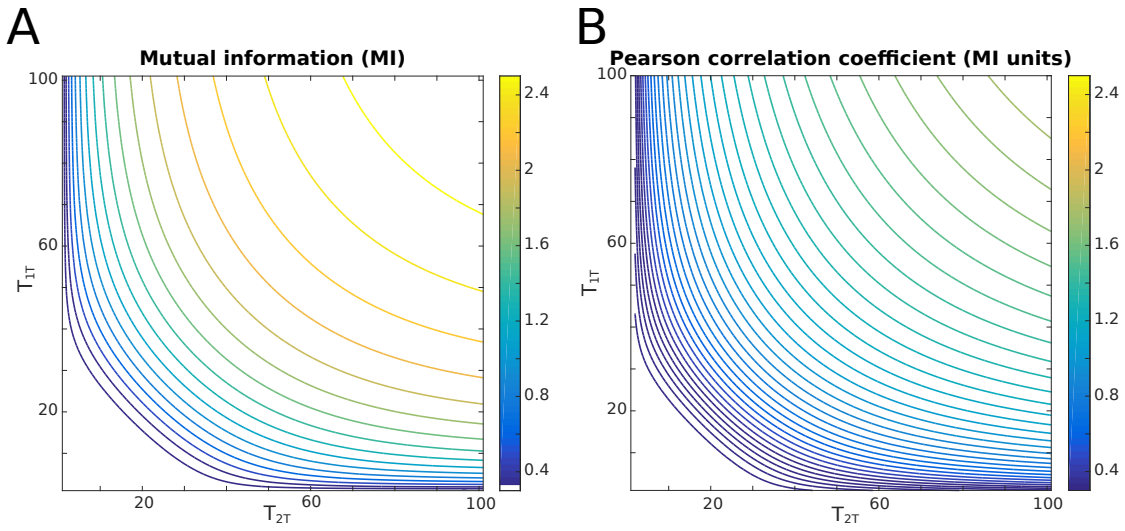

Fig. 4.9 2-D comparison between the Mutual Information (A) and the Pearson correlation coefficient (B) (converted in MI units through eq. (4.22)). $K_1 = K_2 = 1.0$ $S_T = 30$.

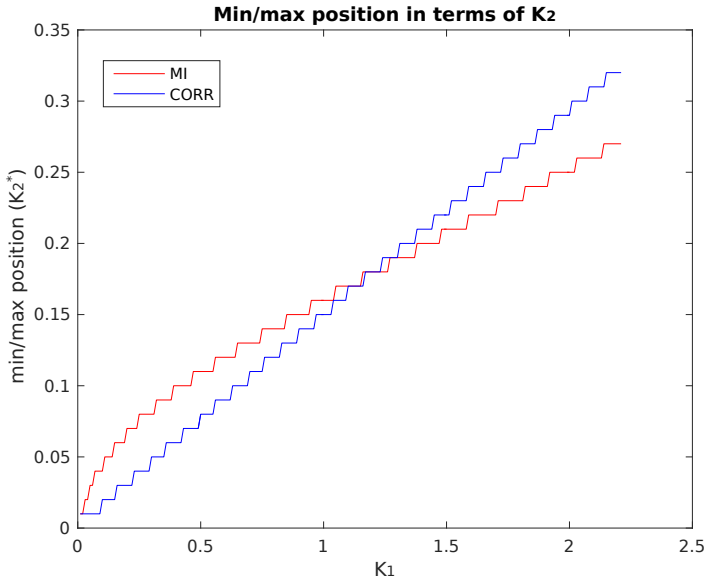

Fig. 4.10 Comparison between the position of the maximum of mutual information and the minimum of Pearson correlation, in terms of K_2, as a function of K_1. $S_T = 30$ $T_{1T} = 45$ $T_{2T} = 21$.

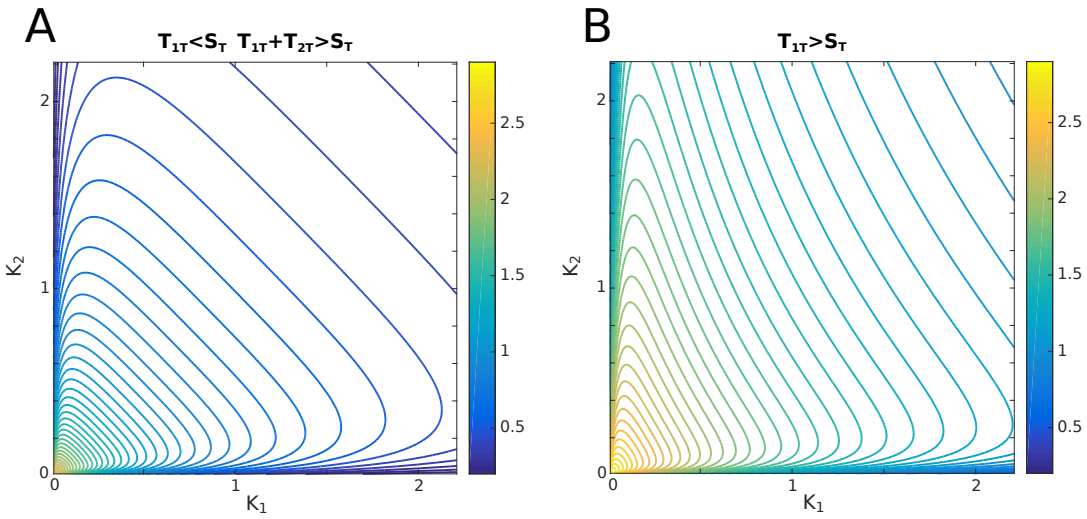

Fig. 4.11 Contour plot of the mutual information between T_1 and T_2 as a function of K_1 and K_2, $S_T = 30$ $T_{2T} = 21$. (A) $T_{1T} = 21$. (B) $T_{1T} = 45$.

maximum of MI and to the minimum of correlation are then different. This means that the regions of optimal coupling identified by the two quantities are not exactly the same. In addition to this difference, we notice that, in the mutual information case, ${K_1}^*$ does not vanish when $T_{1T} > S_T$, as can be see in the contour plots of figure 4.11.

In conclusion, we showed that there are some important differences between Pearson correlation and MI, even though the general behavior is quite similar. In the following section, regarding the role of extrinsic noise, we will mainly focus on the Pearson correlation coefficient. Nonetheless, we will show that in this case, the difference between the two quantities is almost negligible.

4.2.3 Extrinsic noise and its effects on correlation

So far, we studied a minimal model composed of two target species that compete for the same sequester. In this model the only source of noise was the one due to the intrinsic fluctuations of the system. In this section, we investigate the role played in this model by an extrinsic source of noise. In order to add this extrinsic noise to the system, we follow the same approach introduced in section 3.2.2 for the case with one target. In particular, we assume the total amount of sequester to be described by the probability distribution $P(S_T)$. This distribution is a discrete

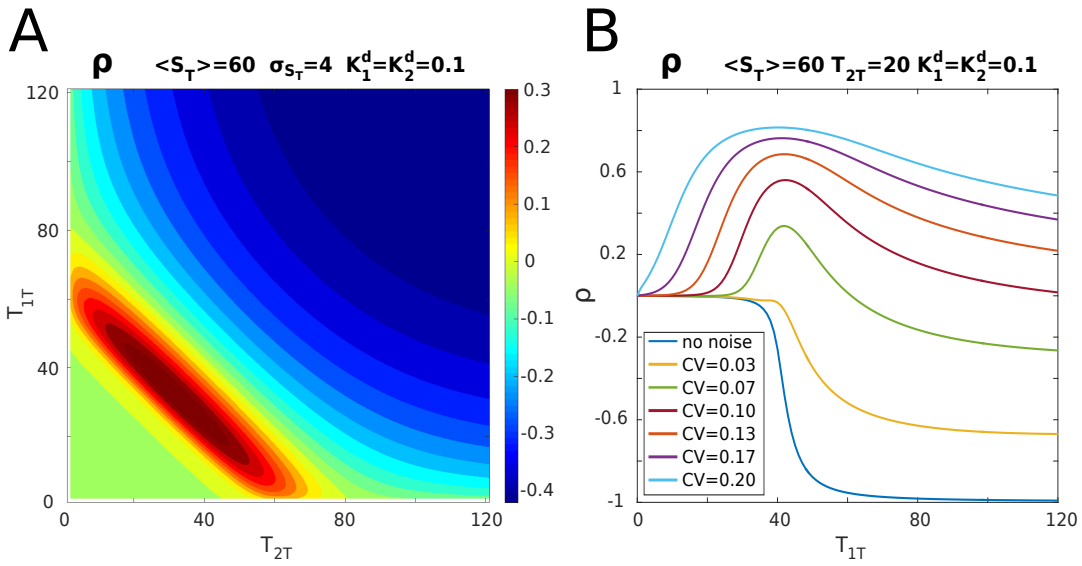

Fig. 4.12 Pearson correlation coefficient in presence of extrinsic noise. $\langle S_T \rangle = 60$ $K_1 = K_2 = 0.1$. (A) Contour plot, $\sigma = 4$. (B) As a function of T_{1T}, $T_{2T} = 20$, for different levels of extrinsic noise (see legend). Adapted from [9].

Gaussian defined as the discrete Gaussian kernel presented in section 3.2.1. Given the steady-state solution of the master equation for a fixed S_T, eq. (4.17), the full solution in presence of extrinsic noise on the value of S_T is obtained by using the law of total probability. The probability distribution $P(T_1, T_2)$ in presence of extrinsic noise is then the superposition of single probability distributions obtained for a given value of S_T weighted by the probability of picking that value of S_T:

$$P(T_1, T_2) = \sum_{S_T=0}^{\infty} P(T_1, T_2|S_T)P(S_T)\,. \tag{4.23}$$

The conditional probability $P(T_1, T_2|S_T)$ is the solution of the master equation obtained for a constant S_T, eq. (4.17). Given that S_T ranges from 0 to infinity, in presence of extrinsic noise, the minimal value of T_1 and T_2 is always 0 and the distribution $P(T_1, T_2)$ is then defined in the space $0 \leq T_1 \leq T_{1T}$, $0 \leq T_2 \leq T_{2T}$.

As a first result, we recover the smoothing of the threshold-like behavior of the means already discussed in section 3.2.3.

The main goal of this analysis is the study of the correlation between the targets: fluctuations in the total amount of sequester molecules positively correlate them. Indeed, in systems that picked a low (high) value of S_T, both the targets will tend

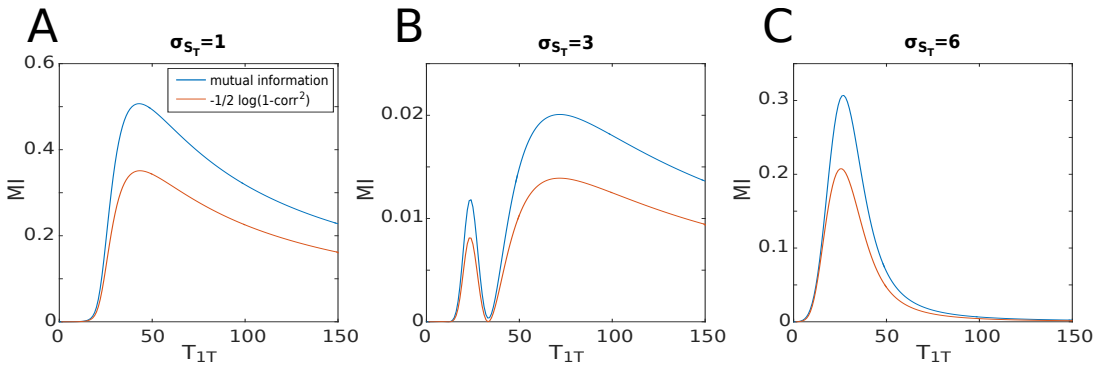

Fig. 4.13 Comparison between the Mutual Information (MI), blue line, and the Pearson correlation coefficient, orange line, (converted in MI units through eq. (4.22)). $T_{2T} = 10$ $\langle S_T \rangle = 30$ $K_1 = 0.2$ $K_2 = 1.0$. Adapted from [9].

to have more (fewer) free molecules. Such positive correlation counterbalances the negative one induced by the competition and discussed in the previous section. The negative interference between these two opposite sources of correlation can lead to almost uncorrelated systems. Nonetheless, these two conflicting sources of correlation have a different dependence on the dissociation constants and the total abundances of targets. As a consequence, it is possible to identify regions of the parameters space in which they are positively or negatively correlated. In particular, following figure 4.12 (B), we can see that small extrinsic fluctuations reduce the level of anticorrelation induced by the competition. If the extrinsic noise is increased, the positive correlation can counterbalance the negative one, resulting in a global correlation close to zero. Finally, a further increase of the extrinsic fluctuations leads to a positive global correlation, with a maximum around the threshold point.

This behavior suggests that competition can be an effective way to obtain positive correlation between the targets when the common resource is fluctuating.

Figure 4.13 shows that the MI and the Pearson correlation coefficient have almost the same profile in presence of extrinsic noise. As a consequence, the discussion above is valid also in case the MI is used to measure the coupling between the targets.

The presence of extrinsic noise influences not only the behavior of the correlations, but also the shape of the probability distribution. Indeed, as discussed for the case with one target, extrinsic fluctuations in the amount of the sequester can give rise to bimodal distributions of the targets. Bimodality can be found in the joint distribution $P(T_1, T_2)$, but it is more common in the marginal probabilities, $P(T_1)$ and $P(T_2)$. Figure 4.14 shows some examples of bimodality, both in the joint probability

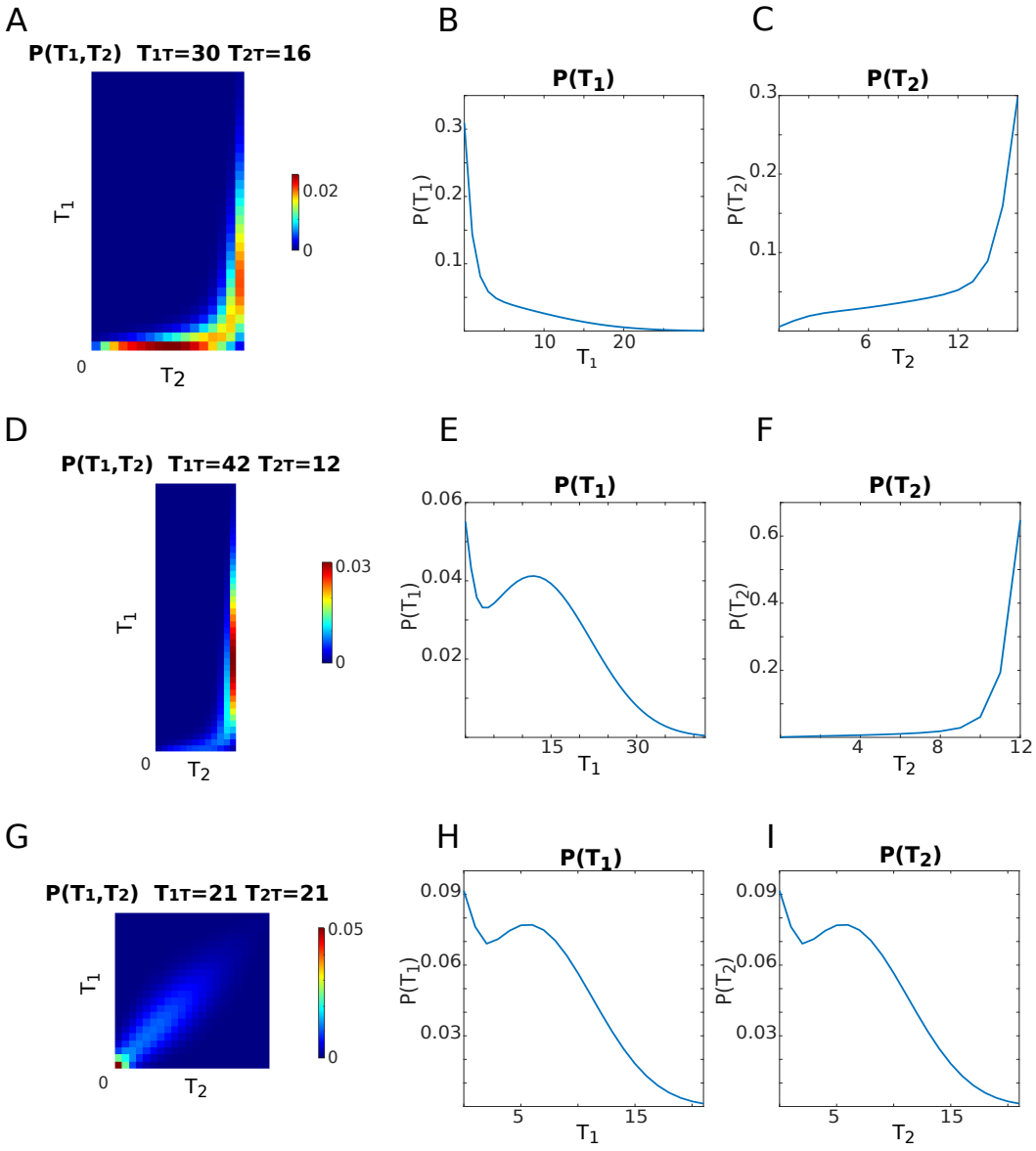

Fig. 4.14 Examples of bimodality in the joint probability $P(T_1, T_2)$ and in the marginal probabilities $P(T_1)$ and $P(T_2)$. $\langle S_T \rangle = 30$ $sd = 10$ (A-C) $T_{1T} = 30$ $T_{2T} = 16$ $K_1 = 0.01$ $K_2 = 1.00$. (D-F) $T_{1T} = 42$ $T_{2T} = 12$ $K_1 = 0.01$ $K_2 = 1.00$. (G-I) $T_{1T} = 21$ $T_{2T} = 21$ $K_1 = 0.10$ $K_2 = 0.10$.

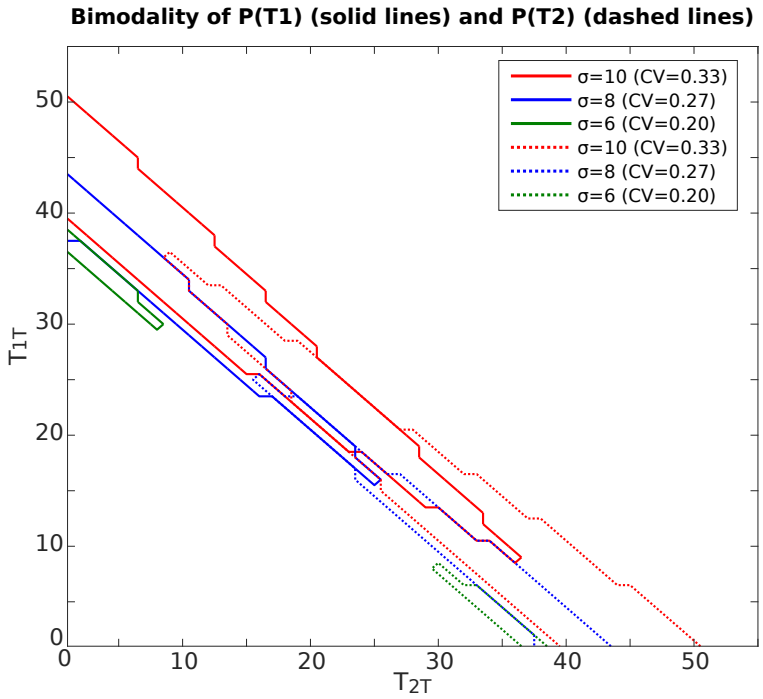

Fig. 4.15 Phase diagram that identifies the bimodality region of the marginal probability $P(T_1)$, solid lines, and $P(T_2)$, dashed lines, as a function of T_{1T} and T_{2T} for different levels of extrinsic noise. $\langle S_T \rangle = 30$ $K_1 = K_2 = 0.1$.

and in the marginals. The presence of bimodality in the marginal probabilities of the two targets can be tuned by changing their total amount.

Figure 4.15 represents a phase diagram that indicates the regions of bimodality of the two marginals for different levels of extrinsic noise and with $K_1 = K_2$. The interesting aspect is that the crosstalk between the competitors allows bimdality on one target to be tuned by the total amount of molecules of the other target. Furthermore, there are regions in which both the marginal probabilities are bimodal, regions with only one bimodal distributions and regions in which both the distributions are unimodal. The interplay between the competitors plays then a crucial role in stabilizing specific phenotypes. We will study this effect in more detail in the following sections for the model of miRNA-mediated regulation with two target genes competing for the same miRNA.

4.3 Model of competition between two mRNAs targeted by the same miRNA

In the previous section we studied a minimal model in which two target species compete for the same sequester. We exploited the complete analytic control to investigate the behavior of the correlations and the shape of the distribution, both in presence and absence of extrinsic fluctuations in the amount of sequester. We showed that the competition negatively correlates the targets, while fluctuations of the sequester positively correlate them. Furthermore, we pointed out that, in presence of extrinsic noise, the crosstalk between the targets is able to tune the shape of the distribution, from unimodal to bimodal and vice versa. In this section we will define a model of miRNA-mediated regulation similar to the one described in section 3.4, but with the addition of a second target species competing for the same miRNA. We will study this system through analytic approximations and numerical simulations, focusing on the role of the interplay between competition and extrinsic noise in stabilizing different phenotypes.

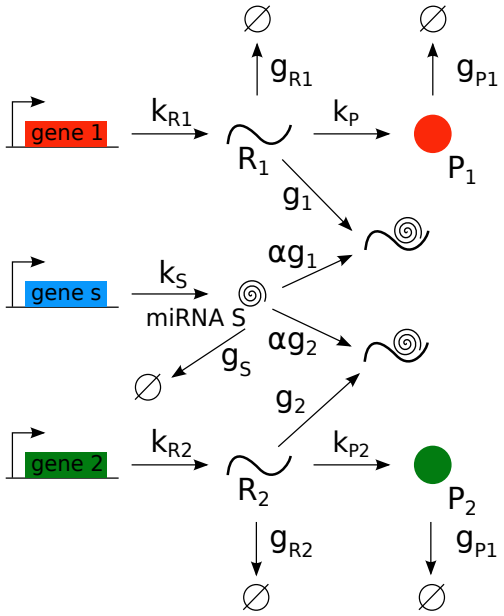

Fig. 4.16 MiRNA-target interaction network with two competing targets. The miRNA and the target mRNAs are transcribed from independent genes with rate k_S, k_{R_1} and k_{R_2} respectively and degrade with rate g_S, g_{R_1} and g_{R_2}. The interaction is governed by the effective parameter g. The target mRNAs are always degraded after binding, while the miRNA can be recycled with probability $1-\alpha$. Free mRNAs are translated with rate k_{P_1} and k_{P_2} into proteins which can in turn degrade with rate g_{P_1} and g_{P_2}. Adapted from [10].

4.3.1 Introduction to the model and approximate analytic solution

The stochastic model of miRNA-mediated regulation that we here describe, corresponds the one already introduced by Bosia et al. in [7] and discussed in section 2.2.4, with only one miRNA and two target genes.

The circuit representing the model is depicted in figure 4.16. The molecular species involved are miRNAs (S), two different messenger RNAs (R_1 and R_2) and proteins (P_1 and P_2), product of the translation of the two target mRNAs. MiRNAs and the target mRNAs are transcribed from independent genes. For simplicity we neglect all the intermediate reactions leading to the synthesis of mRNAs and miRNAs. We then assume that they are produced at constant rates: k_{R_1} and k_{R_2} for the first and second target respectively, and k_S for the miRNA. MiRNAs and mRNAs can also be

degraded with rates g_S for the miRNA and g_{R_1} and g_{R_2} for the target mRNAs. The reactions associated to these processes are:

$$\emptyset \underset{g_{R_1}}{\overset{k_{R_1}}{\rightleftharpoons}} R_1\,, \qquad \emptyset \underset{g_{R_2}}{\overset{k_{R_2}}{\rightleftharpoons}} R_2\,, \qquad \emptyset \underset{g_S}{\overset{k_S}{\rightleftharpoons}} S\,. \tag{4.24}$$

MiRNAs bind the target mRNAs forming a complex that can be subsequently degraded. This complex is assumed to reach equilibrium faster than the other species, then the interaction between the miRNA and the first and second target mRNA is quantified by the effective parameters g_1 and g_2 respectively. The strength of the miRNA-target coupling, in terms of the association and dissociation rates, together with the complex degradation rate, are contained in these parameters. The formation of the miRNA-mRNA complexes then reads:

$$R_1 + S \xrightarrow{g_1} R_1 S\,, \qquad R_2 + S \xrightarrow{g_2} R_2 S\,. \tag{4.25}$$

While mRNAs are assumed to be always degraded as a consequence of sequestration, miRNAs can be recycled with probability $1-\alpha$, also called catalyticity:

$$R_1 S \xrightarrow{1-\alpha} S\,, \qquad R_2 S \xrightarrow{1-\alpha} S\,. \tag{4.26}$$

Whenever the mRNAs are not bound to miRNAs, they can be translated by ribosomes into proteins with translation rate k_{P_1} for the first target and k_{P_2} for the second one. As assumed for the other molecular species, proteins can be degraded by mass-action kinetics with rates g_{P_1} and g_{P_2}. These processes read:

$$R_1 \xrightarrow{k_{P_1}} R_1 + P_1\,, \qquad R_2 \xrightarrow{k_{P_2}} R_2 + P_2\,. \tag{4.27}$$

$$P_1 \xrightarrow{g_{P_1}} \emptyset\,, \qquad P_2 \xrightarrow{g_{P_2}} \emptyset\,. \tag{4.28}$$

As done for the case of a single target, this effective model can be studied by a deterministic approach, investigating the behavior of the concentrations of the molecular species involved. The rate equations associated to this model are:

$$\frac{d[R_i]}{dt} = k_{R_i} - g_{R_i}[R] - g_i\,[R_i]\,[S]\,, \tag{4.29}$$

$$\frac{d[S]}{dt} = k_S - g_S[S] - \alpha \sum_{i=1}^{2} g_i\,[R_i]\,[S]\,, \tag{4.30}$$

$$\frac{d[P_i]}{dt} = k_{P_i}[R_i] - g_{P_i}[P_i]\,, \tag{4.31}$$

where i assumes value 1 and 2. This system of equations can be easily solved at the steady state, obtaining the concentrations of miRNA, mRNAs and proteins.

Since our goal is the study of how the shape of the probability distribution is influenced by the competition between the targets, we need to take into account the stochastic fluctuations of the system. The complete master equation describing the time evolution of the probability distribution of observing $\overline{n} = (R_1, R_2, S, P_1, P_2)$ molecules is:

$$\begin{aligned}
\frac{dP(\overline{n},t)}{dt} = & \sum_{i=1}^{2} k_{R_i}\Big[P(R_i-1,t) - P(\overline{n},t)\Big] + \\
& \sum_{i=1}^{2} \frac{g_{R_i}}{V_{cell}}\Big[(R_i+1)P(R_i+1,t) - R_i P(\overline{n},t)\Big] + \\
& k_S\Big[P(S-1,t) - P(\overline{n},t)\Big] + \\
& \frac{g_S}{V_{cell}}\Big[(S+1)P(S+1,t) - S P(\overline{n},t)\Big] + \\
& \sum_{i=1}^{2} \frac{k_{P_i} R_i}{V_{cell}}\Big[P(P_i-1,t) - P(\overline{n},t)\Big] + \\
& \sum_{i=1}^{2} \frac{g_{P_i}}{V_{cell}}\Big[(P_i+1)P(P_i+1,t) - P_i P(\overline{n},t)\Big] + \\
& \sum_{i=1}^{2} \frac{g_i \alpha}{V_{cell}^2}\Big[(S+1)(R_i+1)P(R_i+1,S+1,t) - S R_i P(\overline{n},t)\Big] + \\
& \sum_{i=1}^{2} \frac{g_i(1-\alpha)S}{V_{cell}^2}\Big[(R_i+1)P(R_i+1,t) - R_i P(\overline{n},t)\Big]\,,
\end{aligned} \tag{4.32}$$

where the rates have been rescaled due to the fact that the variables are now numbers of molecules instead of concentrations. Indeed, we have $R_i = [R_i]V_{cell}$, $S = [S]V_{cell}$ and $P_i = [P_i]V_{cell}$, where V_{cell} is the volume of the cell in which reactions occur.

An approximate steady-state analytic solution of the above master equation can be obtained by using again the van Kampen's system-size expansion described in section 3.4.2. In this way, the steady-state probability distribution $P(\overline{n})$ is approximated by a multivariate Gaussian. Together with the analytic approach, we performed numerical simulations through the SSA introduced in section 3.4.3.

As a first result of this analysis, all the theoretical predictions related to the crosstalk between targets described in section 2.2.4 are recovered. In the following section this crosstalk will be investigated in presence of extrinsic noise.

4.3.2 Extrinsic noise: bimodality can be tuned by the interplay between targets

As described in section 3.5.1, extrinsic noise is added to the system in terms of a fluctuating miRNA transcription rate k_S. The probability distribution of this rate, $P(k_S)$, is the Gaussian distribution of equation (3.62), with mean $\langle k_S \rangle$ and variance $\sigma^2_{k_S}$, and the level of extrinsic noise is quantified by the coefficient of variation: $CV = \sigma_{k_S}/\langle k_S \rangle$.

The steady-state probability distribution $P(R_1, R_2, S, P_1, P_2)$ describing the system with extrinsic noise is obtained by implementing the law of total probability. Indeed, we perform a superposition of probability distributions conditioned on a fixed value of k_S, weighted by the probability $P(k_S)$:

$$P(R_1, R_2, S, P_1, P_2) = \int P(k_S) P(R_1, R_2, S, P_1, P_2 | k_S) dk_S \,, \tag{4.33}$$

where the conditional probability distribution $P(R_1, R_2, S, P_1, P_2|k_S)$ is obtained as solution of the master equation, eq. (4.32), for any given k_S, through the van Kampen's approximation.

As a result of the fluctuating k_S, we find that the marginal probability distributions of the two target mRNAs (and corresponding proteins) can display two distinct

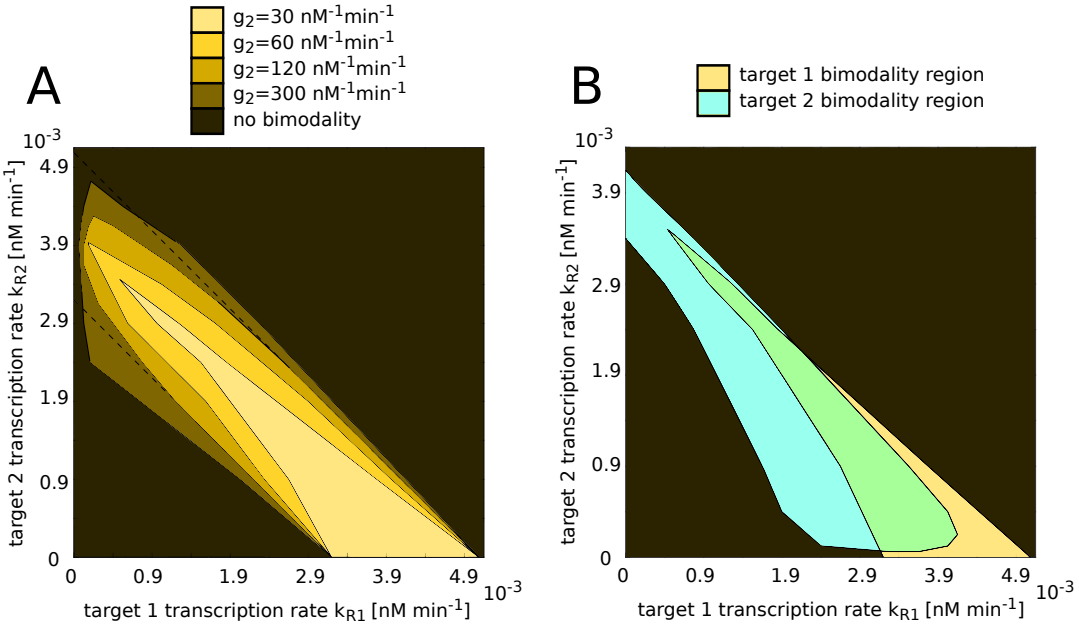

Fig. 4.17 (A) Phase diagram that identifies the bimodality region of the marginal probability of target R_1 for a fixed level of extrinsic noise, small miRNA/target 1 interaction strength $g_1 = 1.2 \times 10^2$ nM^{-1} min^{-1} and different miRNA/target 2 interaction strengths g_2, as a function of the transcription rates k_{R_1} and k_{R_2}. (B) Phase diagram that identifies the bimodality region of the marginal probabilities of target R_1 (yellow) and R_2 (light blue) for fixed values of g_1 and g_2 and a fixed level of extrinsic noise, as a function of the transcription rates k_{R_1} and k_{R_2}. Adapted from [10].

phenotypes. This bimodality at the population level is influenced by the relative expression of the two targets and by the strength of their coupling with the miRNA. In order to better characterize these trends, let us consider the phase diagram of figure 4.17 (A). This plot shows the region of bimodality of the marginal probability of target R_1 as a function of the transcription rates k_{R_1} and k_{R_2}, for a fixed level of extrinsic noise, a fixed g_1 and different interaction strengths g_2. For the particular value of g_1 chosen, in absence of extrinsic noise, the distribution of R_1 does not present a bimodal profile for any value of the parameters space k_{R_1}-k_{R_2}. As can be seen, by increasing the transcription rate k_{R_2}, the range of bimodality of $P(R_1)$ shifts towards lower values of the transcription rate k_{R_1}. This trend follows the one of the threshold point. When k_{R_2} is increased, more miRNAs are sponged away by R_2 and a lower amount of R_1 is needed to cross the threshold, whose location moves towards a lower value of k_{R_1}. The width of the range of bimodality of $P(R_1)$ is controlled by the interaction strength g_2 of the second target with the miRNA, with respect to the fixed value of g_1. If $g_2 > g_1$, miRNAs are sponged away with high frequency by R_2 and the net effect is the reduction of the amount of miRNAs that can bind R_1. As a consequence, the range of bimodality shifts and increases its width when k_{R_2} is increased. Conversely, if $g_2 < g_1$, the second target R_2 interacts with lower frequency with the miRNA, R_1 is slightly derepressed and the net effect is a shrink of the range of transcription rates k_{R_1} for which bimodality is present, as the expression of R_2 is increased.

The emerging picture is that, for a given transcription rate k_{R_1}, the distribution of target R_1 can be tuned from unimodal to bimodal and from unrepressed to repressed and vice versa through the expression of target R_2. Figure 4.18 presents an example of this phenomenon and shows also the good agreement between the analytic approximation and numerical simulations. The presence of extrinsic noise is fundamental, because it allows to obtain this cross regulation effect even when the interaction strengths are small.

This interplay between targets in presence of extrinsic noise, gives rise to a heterogeneous scenario. Indeed, depending on the level of expression of the two targets, their marginal distributions can be both bimodal or unimodal, but the case in which one distribution is bimodal and the other one is unimodal is also possible, see figure 4.17 (B). As a consequence, this phenomenon should be taken into account when dealing with biological systems in which fine tuning or stabilization of specific phenotypes are required.

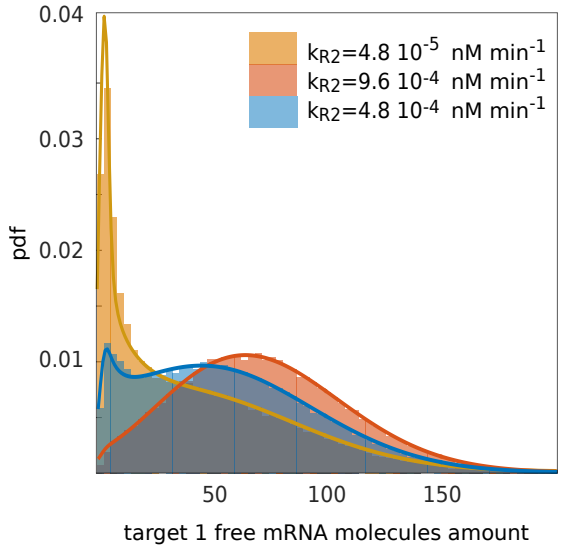

Fig. 4.18 Explanatory example of how it is possible to modulate target 1 distribution increasing the expression of target 2 for small interaction strength between miRNA and targets. The extrinsic noise is fixed. Adapted from [10].

4.4 Implications and conclusions

In this chapter we investigated the coupling induced by the competition between two targets of the same sequester. By considering a minimal model under analytic control, we studied the behavior of the correlation between the competitors in terms of Pearson correlation coefficient and mutual information. In case of pure intrinsic noise, we showed that the two quantities convey almost the same message, with some slight but non-trivial differences. The Pearson coefficient highlighted the fact that the competition between the targets induces a negative correlation between them. Furthermore, regions of optimal anticorrelation can be found in the space of the two dissociation constants. Extrinsic noise was then added to the system, in terms of a fluctuating total amount of sequester molecules. In this case, we showed that the fluctuations of the common resource induce a positive correlation between the competitors. This positive correlation tends to counterbalance the negative one induced by competition. As a consequence, depending on the level of extrinsic noise, regions of positive, negative or vanishing correlation can be identified. After the characterization of the minimal model, we turned to study a model of miRNA-mediated regulation with two genes competing for the same miRNA. The solution of the master equation associated to the model was obtained

thorough an analytic approximation and numerical simulations were performed. With this system we mainly focused on the appearance and tuning of bimodality in the marginal probabilities of the targets, in presence of extrinsic noise. We showed that the competition for the same miRNA allows to tune the profile of the marginal distribution of one target by changing the expression level of the other target. This interplay between extrinsic noise and competition leads to a level of cross regulation between the targets that goes beyond mean values and has to be taken into account together with the correlation counterpart. In particular, these aspects can be of fundamental importance in the framework of miRNA-mediated regulation and cell differentiation, when coupled to other downstream mechanisms able to stabilize the phenotypes.

Chapter 5

Bridging the minimal model to the miRNA-mediated regulation

In chapter 3 we introduced and studied a minimal model of molecular sequestration. We derived the analytic solution of the corresponding master equation and discussed the role played by the threshold response and the extrinsic noise in giving rise to bimodal distributions. Afterwards, we studied a more complex model, based on molecular sequestration, that described the process of miRNA-mediated gene regulation. With this model, we characterized the presence of bimodal distributions as a function of the parameters and in relation to some source of extrinsic noise, obtaining interesting results that could be in principle experimentally validated.

With respect to the minimal model, the one describing miRNA-mediated gene regulation took into account also the processes of production and spontaneous degradation of the species involved, relaxing then the constraint on the conservation of the total number of molecules. In this chapter we will briefly analyze some slightly modified variations of the minimal model, in order to build the bridge towards the model of miRNA-mediated gene regulation. The addition of some elements that make the minimal model more similar to the complete one, will lead to non-trivial results, like the disappearance of the threshold behavior.

5.1 Stochastic approach to the model with production and degradation of T

In order to better understand the role of fluctuations originated by the processes of production and degradation, we modify the minimal model allowing only one species (T) to be produced and degraded, respectively with rate k_T and g_T, while keeping constant the total amount of molecules of the other species. The reactions that define the model are:

$$T + S \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \overline{TS}\,, \tag{5.1}$$

$$\emptyset \underset{g_T}{\overset{k_T}{\rightleftharpoons}} T\,, \tag{5.2}$$

$$S_T = S + \overline{TS} = cost\,. \tag{5.3}$$

In this model only one conservation law is present, as a consequence the corresponding master equation has two variables and reads:

$$\begin{aligned}\frac{dP(T,S)}{dt} &= k_T P(T-1,S) + g_T(T+1)P(T+1,S) + \\ &+ k_+(T+1)(S+1)P(T+1,S+1) + k_-(\overline{TS}+1)P(T-1,S-1) + \\ &- [k_T + g_T T + k_+ TS + k_- \overline{TS}]P(T,S)\,.\end{aligned} \tag{5.4}$$

By exploiting the conservation law, eq. (5.3), the master equation can be written in terms of T and S only:

$$\begin{aligned}\frac{dP(T,S)}{dt} &= k_T P(T-1,S) + g_T(T+1)P(T+1,S) + \\ &+ k_+(T+1)(S+1)P(T+1,S+1) + k_-(S_T - S + 1)P(T-1,S-1) + \\ &- [k_T + g_T T + k_+ TS + k_-(S_T - S)]P(T,S)\,.\end{aligned} \tag{5.5}$$

In this network of interactions detailed balance holds. Indeed, let's consider a closed loop that links together four different states of the system: the clockwise product of the transition probabilities is equal to the counterclockwise one, then no probability flows are present at the steady state:

$$\frac{k_-(S_T-S+1)k_T k_+(T+1)Sg_T T}{k_T k_-(S_T-S+1)g_T(T+1)k_+ TS}=1\,.$$

Detailed balance implies that each reaction is balanced by its reverse one, then the following relations hold:

$$k_T P(T-1,S)=g_T T\,P(T,S)\,, \tag{5.6}$$
$$k_-(S_T-S+1)P(T-1,S-1)=k_+ TSP(T,S)\,. \tag{5.7}$$

By summing equation (5.6) over all possible values of S we obtain a relation that can be used to recursively write the marginal probabilities for T:

$$\sum_{S=0}^{S_T} k_T P(T-1,S)=\sum_{S=0}^{S_T} g_T T\,P(T,S)\,,$$

$$P(T)=\frac{k_T}{g_T}\frac{1}{T}P(T-1)\,. \tag{5.8}$$

Then:

$$\begin{aligned} P(1)&=\frac{k_T}{g_T}P(0)\,,\\ P(2)&=\frac{k_T}{g_T}\frac{1}{2}P(1)\,,\\ &\dots\\ P(T)&=\frac{k_T}{g_T}\frac{1}{T}P(T-1)\,. \end{aligned}$$

Which leads to:

$$P(T) = \left(\frac{k_T}{g_T}\right)^T \frac{1}{T!} P(0)\,. \tag{5.9}$$

$P(0)$ is given by the normalization of the probability distribution and exploiting the power series expansion of the exponential $e^x = \sum_{n=0}^{\infty} x^n/n!$ it can be written as:

$$P(0) = \frac{1}{\sum_{T=0}^{\infty} \left(\frac{k_T}{g_T}\right)^R \frac{1}{T!}} = e^{-\frac{k_T}{g_T}}\,.$$

The marginal probability of R assumes then the Poisson form:

$$P(T) = \left(\frac{k_T}{g_T}\right)^T \frac{1}{T!} e^{-\frac{k_T}{g_T}}\,. \tag{5.10}$$

At the steady state the dynamics of the species T is simply governed by the birth and death process, the binding and unbinding with S have no influence and the average value of T is simply $< T >= k_T/g_T$.

By using equation (5.6) to obtain the expression for $T\,P(T,S)$ and substituting it in equation (5.7) we obtain:

$$P(T-1,S) = \left(\frac{g_T k_-}{k_T k_+}\right) \frac{S_T - S + 1}{S} P(T-1,S-1)\,. \tag{5.11}$$

Through this relation, keeping R fixed, we can write:

$$\begin{aligned}
P(T,1) &= \left(\frac{g_T k_-}{k_T k_+}\right) S_T\, P(T,0)\,,\\
P(T,2) &= \left(\frac{g_T k_-}{k_T k_+}\right) \frac{S_T - 1}{2} P(T,1)\,.,\\
&\cdots\\
P(T,S) &= \left(\frac{g_T k_-}{k_T k_+}\right) \frac{S_T - S + 1}{S} P(T,S-1)\,.
\end{aligned}$$

Which, recursively on S, leads to:

$$P(T,S) = \left(\frac{g_T k_-}{k_T k_+}\right)^S \frac{S_T!}{S!(S_T-S)!} P(T,0) = \left(\frac{g_T k_-}{k_T k_+}\right)^S \binom{S_T}{S} P(T,0)\,. \tag{5.12}$$

By dividing equation (5.12) by the marginal probability $P(T)$, we obtain the conditional probability of S given T:

$$P(S|T) = \left(\frac{g_T k_-}{k_T k_+}\right)^S \binom{S_T}{S} P(0|T)\,. \tag{5.13}$$

As a conditional probability, $P(S|T)$ satisfies the normalization condition $\sum_{S=0}^{S_T} P(S|T) = 1$. By exploiting the binomial identity $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$, we can write the explicit expression for $P(0|T)$:

$$P(0|T) = \sum_{S=0}^{S_T} \binom{S_T}{S} 1^{S_T - S} \left(\frac{g_T k_-}{k_T k_+}\right)^S = \left(\frac{g_T k_-}{k_T k_+} + 1\right)^{S_T}\,. \tag{5.14}$$

The explicit expression of the conditional probability is then:

$$\begin{aligned} P(S|T) &= \left(\frac{g_T k_-}{k_T k_+}\right)^S \binom{S_T}{S} \left(\frac{g_T k_-}{k_T k_+} + 1\right)^{-S_T} = \\ &= \left(\frac{k_-}{\langle T\rangle k_+}\right)^S \binom{S_T}{S} \left(\frac{k_-}{\langle T\rangle k_+} + 1\right)^{-S_T}\,. \end{aligned} \tag{5.15}$$

The joint probability $P(R,S)$ can be recovered by multiplying the conditional probability $P(S|R)$ and the marginal probability $P(R)$, eventually obtaining:

$$\begin{aligned} P(T,S) = P(S|T)P(T) &= \left(\frac{g_T k_-}{k_T k_+}\right)^S \binom{S_T}{S} \left(\frac{g_T k_-}{k_T k_+} + 1\right)^{-S_T} \left(\frac{k_T}{g_T}\right)^T \frac{1}{T!} e^{-\frac{k_T}{g_T}} = \\ &= \left(\frac{k_-}{\langle T\rangle k_+}\right)^S \binom{S_T}{S} \left(\frac{k_-}{\langle T\rangle k_+} + 1\right)^{-S_T} \langle T\rangle^T \frac{1}{T!} e^{-\langle T\rangle}\,. \end{aligned} \tag{5.16}$$

This is the solution of the master equation, eq. (5.5). As can be seen, the conditional probability $P(S|T)$ depends on the average value of T only and then

coincides with the marginal probability $P(S)$. The joint probability is equal to the product of the marginal probabilities, T and S are therefore independent variables at the steady state. The coupling caused by the titrative interaction is found in the dependence on $\langle T \rangle$ of $P(S)$. In conclusion, for this system the threshold-like behavior of the mean of T is completely disrupted, as it increases linearly with k_T ($\langle T \rangle = k_T/g_T$). In conclusion, this minimal model extended with the addition of stochastic production and degradation of one species cannot be considered a step forward towards the complete model of miRNA-target interaction.

5.2 Mean field

The above result showed that the connection between the minimal model and the model of miRNA-mediated gene regulation is not trivial. In order to have a quick insight on the behavior of the mean of T, we will here perform a mean field analysis over a series of slightly modified variations of the minimal model.

5.2.1 Model with production and degradation of one species

We here report the steady state mean field solution of the minimal model in which one species is produced and degraded.

The rate equations for the concentrations of T and S are:

$$\frac{d[T]}{dt} = k_T - g_T[T] - k_+[T][S] + k_-[\overline{TS}]\,, \tag{5.17}$$

$$\frac{d[S]}{dt} = -k_+[T][S] + k_-[\overline{TS}]\,. \tag{5.18}$$

Exploiting the only conservation law that holds in this system ($[S_T] = [S] + [\overline{TS}]$), at the steady state, equation (5.17) gives:

$$[S] = \frac{[S_T]k_-}{k_- + [T]k_+}\,. \tag{5.19}$$

Substituting equation (5.19) in (5.18), after some algebra we obtain:

$$[T]^2 - \frac{k_+k_T - k_-g_T}{k_+g_T}[T] - \frac{k_-k_T}{k_+g_T} = 0\,. \tag{5.20}$$

The only acceptable solution of (5.20) is:

$$2[T] = \frac{k_+k_T - k_-g_T}{k_+g_T} + \sqrt{\left(\frac{k_+k_T - k_-g_T}{k_+g_T}\right)^2 + 4\frac{k_-k_T}{k_+g_T}}\,. \tag{5.21}$$

Which can be simplified, obtaining the result expected from the solution of the master equation, eq. (5.5):

$$[T] = \frac{k_T}{g_T}\,. \tag{5.22}$$

5.2.2 Model with production and degradation of both the species

We here check the effect of having both the species produced and degraded.

The rate equations for the concentrations of T, S and $\overline{TS}$ are:

$$\frac{d[T]}{dt} = k_T - g_T[T] - k_+[T][S] + k_-[\overline{TS}]\,, \tag{5.23}$$

$$\frac{d[S]}{dt} = k_S - g_S[S] - k_+[T][S] + k_-[\overline{TS}]\,, \tag{5.24}$$

$$\frac{d[\overline{TS}]}{dt} = k_+[T][S] - k_-[\overline{TS}]\,. \tag{5.25}$$

Equation (5.25) can be easily solved at the steady state as a function of $[T][S]$:

$$[\overline{TS}] = \frac{k_+}{k_-}[T][S]\,. \tag{5.26}$$

The substitution of equation (5.26) in (5.23) and (5.24) decouples T and S in both the equations, giving:

$$[T] = \frac{k_T}{g_T}\,, \tag{5.27}$$

$$[S] = \frac{k_S}{g_S}\,, \tag{5.28}$$

$$[\overline{TS}] = \frac{k_+ k_T k_S}{k_- g_T g_S}\,. \tag{5.29}$$

As a result, also in this case the threshold-like behavior is disrupted.

In Buchler and Louis [1] a threshold-like behavior was found by adding to this model the degradation of the complex $\overline{TS}$.

5.2.3 Model with production and degradation of both the species and instantaneous degradation of the complex

By assuming an instantaneous degradation of the complex $\overline{TS}$, the rate equations that describe the system are:

$$\frac{d[T]}{dt} = k_T - g_T[T] - k_+[T][S]\,, \tag{5.30}$$

$$\frac{d[S]}{dt} = k_S - g_S[S] - k_+[T][S]\,. \tag{5.31}$$

By imposing the steady state and solving the system of equations with respect to $[T]$ we obtain:

$$2[T] = \frac{k_+ k_T - g_T g_S - k_S k_+}{k_+ g_T} + \sqrt{\left(\frac{k_+ k_T - g_T g_S - k_S k_+}{k_+ g_T}\right)^2 + 4\frac{k_T g_S}{k_+ g_T}}\,. \tag{5.32}$$

In the limit of $k_+ \to \infty$ we have:

$$[T] = \frac{k_T - k_S}{g_T}\,, \tag{5.33}$$

which shows the presence of a threshold in correspondence of $k_T = k_S$, with $[T]$ increasing linearly with k_T after that point.

5.2.4 Model with production and degradation of both the species and instantaneous degradation of the molecule of T in the complex

In this variation of the model we assume that both the species are produced and degraded and that the molecule of T is instantaneously degraded after binding in the complex. Molecules of S are instead completely recycled (catalytic reaction). The rate equations are:

$$\frac{d[T]}{dt} = k_T - g_T[T] - k_+[T][S]\,, \tag{5.34}$$

$$\frac{d[S]}{dt} = k_S - g_S[S]\,. \tag{5.35}$$

By solving the system at the steady state we obtain:

$$[S] = \frac{k_S}{g_S}\,, \tag{5.36}$$

$$[T] = \frac{k_T g_S}{g_T g_S + k_+ k_S}\,. \tag{5.37}$$

Despite the degradation of the T molecule, the complete catalyticity of the reaction prevents the arising of a threshold-like behavior.

5.2.5 Model with production and spontaneous and binding-induced degradation of T

We here assume that the species T is produced and degraded, both spontaneously and induced by the binding with S. The total amount of molecules of S is fixed, giving the conservation law $[S_T] = [S] + [\overline{TS}]$.

If the degradation of T is instantaneous upon binding with S, the concentration $[S]$ is always equal to $[S_T]$ and the system is described by a single rate equation for $[T]$:

$$\frac{d[T]}{dt} = k_T - g_T[T] - k_+[T][S_T]\,. \tag{5.38}$$

Which at the steady state gives:

$$[T] = \frac{k_T}{g_T + k_+ S_T}\,. \tag{5.39}$$

Therefore, there is no threshold-like behavior in case of instantaneous degradation of T in the complex.

If the degradation of T in the complex is not instantaneous, but happens with a finite rate g, the system is described by the equations:

$$\frac{d[T]}{dt} = k_T - g_T[T] - k_+[T][S] + k_-[\overline{TS}]\,, \tag{5.40}$$

$$\frac{d[S]}{dt} = -k_+[T][S] + k_-[\overline{TS}] + g[\overline{TS}]\,. \tag{5.41}$$

By imposing the steady state and solving the system with respect to $[T]$ we obtain:

$$2[T] = \frac{k_+k_T - k_-g_T - gg_T - k_+S_Tg}{k_+g_T} + \sqrt{\left(\frac{k_+k_T - k_-g_T - gg_T - k_+S_Tg}{k_+g_T}\right)^2 + 4\frac{k_T(k_- + g)}{k_+g_T}}\,. \tag{5.42}$$

This system was already solved by Mukherj and coworkers [4] who demonstrated the presence of the threshold in $k_T = \frac{g}{g_T}S_T$.

5.2.6 Conclusions

In conclusion, if production and spontaneous degradation are present, the degradation of the complex is essential to obtain a threshold response. Indeed, the threshold-like behavior is restored either in case of degradation of the entire complex or in case of not instantaneous degradation of T in the complex with total recycle of S.

Chapter 6

Experimental analysis

So far, we presented the results of our theoretical study regarding the steady state of systems governed by the mechanism of molecular sequestration. The biological motivation of this analysis was a better understanding of the process of miRNA-mediated gene regulation. On this basis, we first performed a study aimed at characterizing the appearance of two differentially expressed phenotypes in relation to the presence of some extrinsic source of noise. We considered a minimal model under analytic control, then a more complex model of miRNA-target interaction, both with just one sequester species (miRNA) and one target species (mRNA). We demonstrated that bimodal distributions of the target are enhanced by a source of extrinsic noise on the sequester. In particular, we showed that the extrinsic noise can compensate for a low miRNA-mRNA coupling in order to obtain differentially expressed phenotypes across a population of cells.

In addition to this study, we investigated the crosstalk and correlations induced by the competition between two targets of the same sequester, both in presence and absence of extrinsic noise. We again performed this analysis first in a minimal model with two molecular species that bind to the same sequester, then in the model of miRNA-target interaction with two different mRNAs that are regulated by the same miRNA. The minimal model allowed us to compute analytically the steady-state correlation between the targets and characterize its behavior as a function of the parameters. We showed that the competition for the common resource induces a negative correlation between the targets, while fluctuations of the amount of sequester positively correlate them. Finally, in the model of miRNA-mediated regulation, we

described the role played by the crosstalk between the targets in stabilizing or preventing bimodal phenotypes.

Keeping in mind the biological motivation of the study, these theoretical results suggest some experimental strategies that can be useful to better understand how miRNAs influence cell-to-cell variability and how they can couple their target genes. In the following sections, we will present two experiments that we designed in order to investigate the subjects mentioned above. These experiments have been performed in a wet cell biology lab and the preliminary results obtained will be described in the following sections as well.

6.1 Flow cytometry experiments on bimodality and cell cycle with preliminary results

This first experiment is aimed at exploring in vitro the role played by the extrinsic noise in the rise of bimodal distributions of miRNA targets. It is based on the study of the expression distributions of a synthetic target of a miRNA that is in turn endogenous of the cell line used. The heterogeneity of the cells with respect to cell cycle is assumed as an extrinsic source of noise and measurements are performed at the cytofluorimeter. In the following sections, we will describe in detail the idea and motivation behind this experiment, we will illustrate the experimental setup and finally present preliminary data, together with some open questions.

6.1.1 State of the art

The theoretical study that we developed characterizes the presence of bimodal distributions in the free amount of a miRNA target. This study was inspired by the work by Bosia and coworkers [8], in which the presence of bimodal distributions was shown in vitro in a synthetic circuit made of two exogenous genes coding for fluorescent proteins, engineered to be targets of an endogenous miRNA. The experimental setup of this work was already presented in section 2.2.5 and we refer to it for a detailed description. Nevertheless, in the following we will briefly review it, in order to present more clearly the state of the art about miRNA targets bimodal distributions.

The main goal of the experiment was not to investigate the presence of differentially expressed phenotypes, but to characterize the behavior of crosstalk and correlations originated by the competition for the same miRNA. For this purpose, two bidirectional plasmids were built, each of which coding for two different fluorescent proteins. One of the two genes present on each plasmid was engineered in order to contain on its 3'UTR a fixed number (1, 4 or 7) of miRNA binding sites for the miRNA miR-20a. The other gene was left unchanged. The number of miRNA binding sites modulates the miRNA-target interaction strength. The 16 possible combinations of the two plasmids, including the case with 0 binding sites, were used to probe different crosstalk conditions between the two artificial targets. The plasmids were cotransfected in mammalian cells (HEK 293) which endogenously express the miRNA miR-20a. At the steady state, single-cell measurements of the expression of the two targets across the entire population were performed through flow cytometry. In addition, the constitutive expression of each target gene was monitored by measuring the expression of the second gene of each plasmid, the one not regulated by the miRNA. As a result of these measurements, the expression distributions of the two targets, for a given constitutive expression, were reconstructed. This technique allowed to discover the presence of bimodal distributions in certain ranges of constitutive expression. The presence of bimodality was characterized in relation to the intensity of the interaction strengths of the two targets. In particular, in a fixed range of constitutive expression above threshold, it was shown that an increase of the number of miRNA binding sites, i.e. of the interaction strength, induces the rise of a bimodal distribution. Moreover, it was observed that an indirect decrease of the coupling through a stronger competitor, disrupts bimodality.

These observations suggest a possible way to verify the role played by the extrinsic noise in this system. The point is to identify this source of extrinsic noise and find a way to modulate it. The cotransfection experiments described above were performed in a population of identical cells. Nonetheless, each cell during its life undergoes a series of different phases (named G0, G1, S, G2, M), namely the cell cycle, that lead to cell division. During the cell cylce, the cell increases its size and the number of organelles (like ribosomes), duplicates the chromosomes and the expression of many of its genes is modulated.

In the experiment described above, the cells were heterogeneous with respect to the cell cycle. Moreover, the miRNA miR-20a that binds the synthetic target, is known to be related to cell cycle [142–145] and, consequently, we expect its

expression to vary over time. We here propose that such heterogeneity of the population can represent a source of extrinsic noise for the system, since the level of transcription of the miRNA will vary from cell to cell across the population. In section 3.6.4, we showed that our theoretical results are consistent with extrinsic fluctuation with a time scale comparable to the one of the cell cycle. Following this idea, we need a way to modify the intensity of the extrinsic noise, in order to check the predictions of our model. Presumably, we cannot increase the level of extrinsic noise due to cell cycle, but we can decrease it in a rather simple way. Indeed, the different phases can be monitored experimentally using cell-cycle markers, which are fluorescent dyes that bind the DNA. The fluorescence of the dye indicates the amount of DNA present in the cell, which is in turn related to the cell cycle. Then, by measuring the fluorescence emission, the cell-cycle phase of each cell can be determined. Given the assumption that the extrinsic noise originates from the heterogeneity of the population, binning the cells according to their phase should result in a reduction of the noise intensity.

In our experiment we will focus on just one target of miR-20a and we will then use a single bidirectional plasmid. The population of cells will be transfected with the synthetic target and its expression will be measured in single cells, at the steady state, through flow cytometry, adding at that moment also the cell-cycle marker. The experiment will be repeated with different numbers of miRNA binding sites on the target. Finally, the distributions obtained from the heterogeneous population will be compared to the distributions of sub-populations composed of cells belonging to the same phase and the results will be compared to the predictions of the model.

In the following section we will describe in detail the experimental setup e present some preliminary data.

6.1.2 Experimental setup

The bidirectional plasmid used in these experiments is the one developed by Mukherji and coworkers in [4] and it is represented in figure 6.1. The plasmid hosts two genes coding for fluorescent proteins that can be transcribed with the same probability due to the bidirectional promoter. One gene codes for the fluorescent protein mCherry, the other one for the enhanced yellow fluorescent protein (eYFP). The 3'UTR of the gene coding for mCherry is engineered to host a fixed number (0, 1, 4, 7) of miRNA

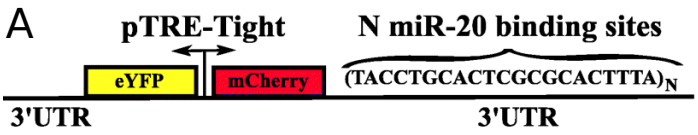

Fig. 6.1 Two-color fluorescent reporter system used in [8]. Bidirectional plasmid coding for mCherry and eYFP. The mCherry gene is engineered to contain a fixed number N of binding sites for the miRNA mir-20 on its 3'UTR. Adapted from [8].

binding sites for the miRNA miR-20a. The gene coding for eYFP is left without any miRNA binding site and is used as control to monitor the constitutive expression of the target, i.e. the expression in absence of regulation.

The bidirectional plasmid is transfected in epithelial cells from human kidney embryo (HEK 293), which endogenously express the miRNA miR-20a. Transfections are performed using the Effectene Transfection Reagent (QIAGEN).

Steady-state measurements of target expression across the population of cells are performed 48 hours after transfection at the flow cytometer. Through this technique the fluorescence of each single cell, both in the mCherry and eYFP channel, is measured. The fluorescence intensity in the mCherry channel is a proxy for the expression of the target, while the one in the eYFP channel is a proxy for the constitutive expression.

Constitutive expression indicates the transcriptional activity of the system and is proportional to the number of plasmids that enter each cell during transfection. Indeed, more plasmids will correspond to a higher number of mRNAs transcribed from the target and control gene.

At the moment of transfection we do not have any control on the number of plasmids that enter the cells. However, this fact is an advantage, since it allows us to investigate a wide range of values of transcriptional activity. The level of transcriptional activity can be considered as a proxy for the target transcription rate, k_R in the model.

Immediately before performing measurements at the flow cytometer, cells are stained with the DNA marker, in order to monitor their cell cycle phase. The marker used is Hoechst 33342, a blue fluorescent dye used to stain DNA in living cells. The emission spectrum of Hoechst does not interfere with the one of mCherry or eYFP and its fluorescence intensity can be measured simultaneously to the one of the two fluorescent proteins.

After the measurement at the flow cytometer, the data collected for each cell are then the fluorescence intensity in the mCherry, eYFP and Hoechst channels, together with quantities related to the cell size and granularity. The latter measurements are needed in order to discriminate alive cells from dead cells and debris.

In order to estimate the background fluorescence intensity in each channel, the measurements described above have to be performed also with non-transfected cells.

Raw data are processed with the Matlab software (MathWorks, MA, USA), subtracting background intensities and sorting cells according to their eYFP intensity level.

6.1.3 Preliminary data

We here present some preliminary results that show how cell cycle influences the expression distribution of the miRNA target.

We first show how different phases of the cell cycle can be discriminated by using the DNA marker. The distribution across the entire cell population of the fluorescence intensity (arbitrary units) in the Hoechst channel is reported in figure 6.2. Cells in phase G2 have approximately twice the amount of DNA of cells in phase G0/G1. Looking at the distribution of the Hoechst, we can indeed notice two peaks, one located at a value of intensity that is approximately two times bigger than the one of the other peak. These two maxima correspond to cells in phase G0/G1 and G2, while the region between them is populated by cells in phase S. The difference in the relative abundance of cells in the three regions is due to the different amount of time that the cells spend in each phase of the cycle. By sorting the cells according to their fluorescence intensity in the Hoechst channel, we can select sub-populations belonging to the desired phase. We will use this technique to study the effect of the variability induced by cell cycle in the regulation of the exogenous target of miR-20a.

As an additional observation, we mention that a rough distinction between cells in phase G0/G1 and cells in phase G2 can be obtained by discriminating them according to their size. Indeed, selecting cells with small sizes is approximately equivalent to select cells in G0/G1, while selecting bigger cells, is approximately equivalent to select cells in G2, see figure 6.3.

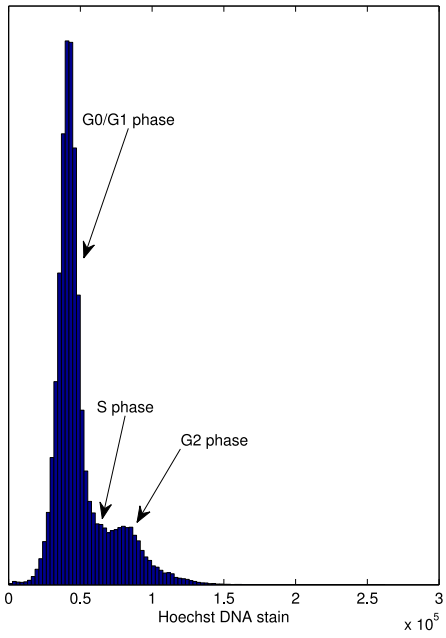

Fig. 6.2 Hoechst DNA stain to identify the cell cycle phases.

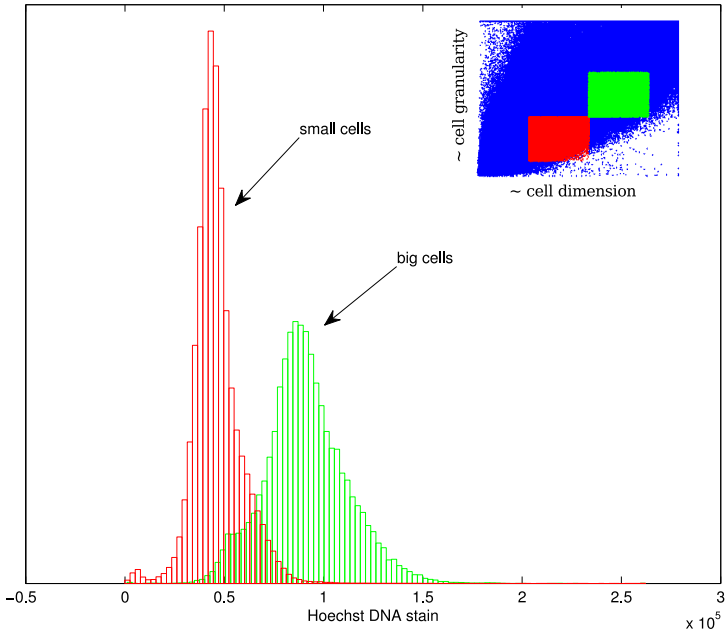

Fig. 6.3 Hoechst DNA stain histogram with small and big cells.

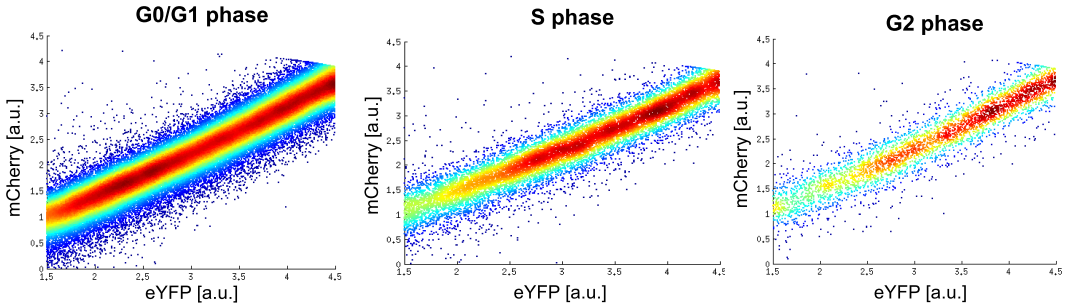

Fig. 6.4 No miRNA binding sites. Scatter plots of the fluorescence intensity in the eYFP and mChery channels, each dot is a cell. Cells are sorted according to their phase of the cycle.

The first experimental observation that can be made on the basis of our preliminary data, is that the cell cycle does not influence the transcriptional activity of the plasmid. The scatter plots in figure 6.4, where each dot is a cell, are the result of fluorescence measurements performed on cells transfected with the plasmid containing no miRNA binding sites on the mCherry gene. The y-axis indicates the fluorescence of the target mCherry, while the x-axis indicates the one of the control eYFP, proxy for the transcriptional activity of the construct. The left-most plot in figure is obtained from the entire population, while the others are the result of sub-populations homogeneous with respect to the phase of the cell cycle. As can be seen, there is a linear dependence between mCherry and eYFP. This is expected, since the miRNA does not regulate mCherry in this construct. In addition to that, we see that this linear dependence is the same, regardless of the phase of the cell cycle considered. This fact indicates that the expression of the construct is independent from the cell cycle. As a consequence, in case of mCherry regulated by the miRNA, differences among the phases should be attributed to the variability of the miRNA expression along the cell cycle.

The scatter plots corresponding to the experiments in which the transfected construct has 1, 4 or 7 miRNA binding sites on the mCherry gene, are shown in figure 6.5. From these results we observe that the expression of miR-20a changes with the cell cycle progression. For a given transcriptional activity (i.e. a given eYFP level) the intensity of mCherry changes both depending on the miRNA repression strenght (i.e. the number of miRNA binding sites present on the construct) and on the cell cycle phase. Since we showed that the transfected plasmids do not change expression with the cell cycle progression, the variability in mCherry expression

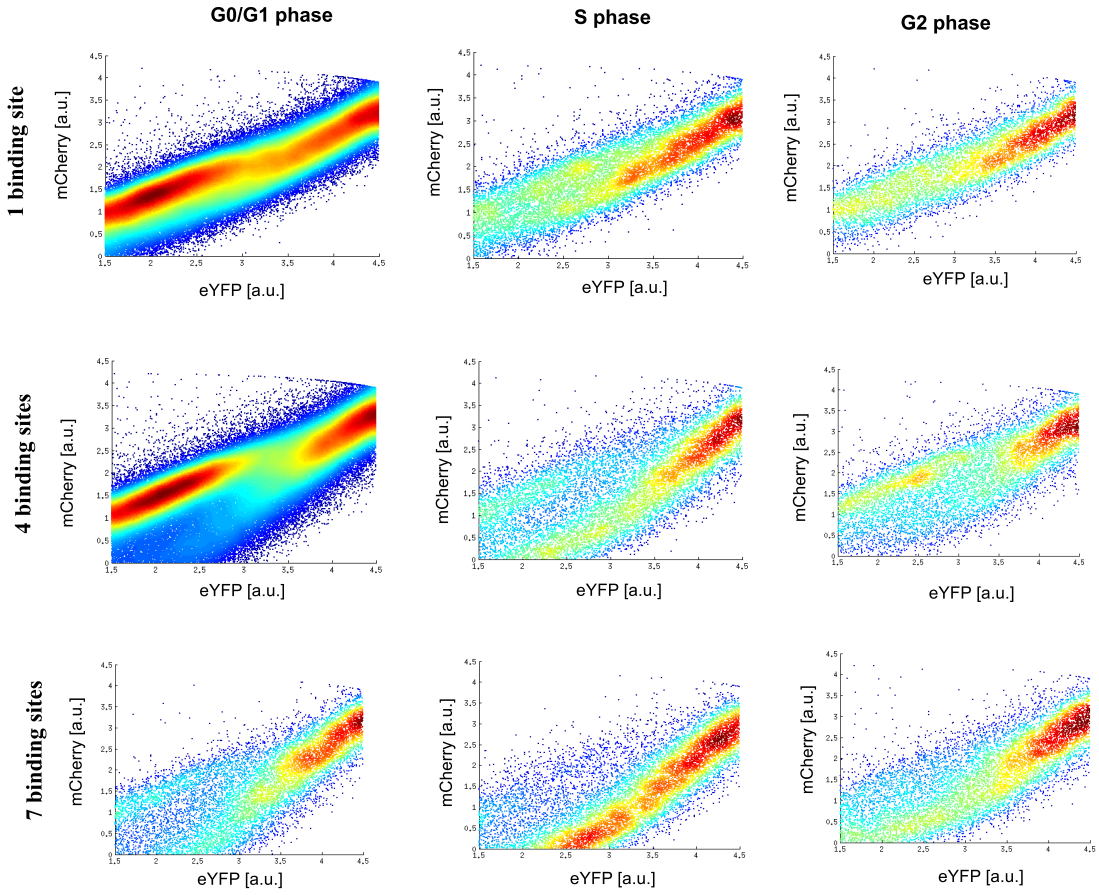

Fig. 6.5 1, 4 and 7 miRNA binding sites. Scatter plots of the fluorescence intensity in the eYFP and mCherry channels, each dot is a cell. Cells are sorted according to their phase of the cycle.

observed for the different cell cycle phases is the fingerprint of a variation in the miR-20a expression.

As can be seen in the scatter plots of the entire population, the regulation operated by the miRNA can induce bimodal expression distributions of the target, for a wide range of transcriptional activity. As expected, this effect is more evident in case of higher repression strength, i.e. 4 or 7 binding sites. In these conditions we can discriminate a repressed peak, whose mean value seems to follow a threshold-like behavior as a function of the transcriptional activity, and an unrepressed peak, whose mean value is essentially linear as a function of the transcriptional activity. The cell cycle influences the shape of the expression distributions. Indeed, in the scatter plots corresponding to sub-populations homogeneous with respect to the phase of the

cycle, the distribution is clearly different from the one of the whole population. In particular, we see that by sorting according to the phase of the cycle, we can roughly select cells belonging mostly to one of the peaks of the population distribution. This seems more evident for cells in phase G0/G1.

The distribution observed in the sub-populations, could be the result of the noise reduction due to the homogeneity with respect to the phase of the cell cycle. Our model predicts the shrinking of the range of bimodality as extrinsic noise is decreased, eventually leading to the disappearance of the bimodal shape. Nonetheless, these results require further analysis and some open questions still remain. In the following section we will discuss these aspects.

6.1.4 Open questions

In the previous section we presented our preliminary results and proposed that they can be in agreement with our model of noise-induced bimodality. The experiments should be repeated and a more accurate analysis has to be performed. This will involve in particular the characterization of the range of bimodality as a function of the noise level (i.e. in the entire population or in sub-populations) and of the repression strength, in terms of the number of miRNA binding sites.

The most important open question regards the shape of the distribution of the miRNA miR-20a. We already showed that the expression of miR-20a changes with the progression of cell cycle and that the range of bimodality of the target can be controlled by sorting cells accordingly. In order to state that this is due to a noise reduction in the expression of the miRNA, we should show that the distribution of the miRNA transcription rates does not correlate with that of the target. If this were the case, the selection of a specific phase of the cycle would correspond to the selection of a specific level of expression of the miRNA. Consequently, the modulation of bimodality would be caused not by a change in the extrinsic noise, but simply by a change in the mean miRNA expression level.

MiRNA quantification in single cells is not an easy task, nonetheless in the last few years, some assays have been developed to detect miRNAs in single cells through flow cytometry [146–148]. These assays are based on Fluorescence in situ Hybridization (FISH), a technique that allows to label with fluorescent probes specific DNA or RNA sequences. FISH is combined to the technology called branched-DNA

amplification, which exploits a Pre-amplifier molecule to achieve thousands-fold amplification of the detection signal, allowing it to be detected with a flow cytometer.

In principle, we could use this kind of assay to indirectly measure the quantity of miR-20a molecules in each single cell, reconstructing the expression distribution across the population. Nevertheless, the compatibility of this technique with the already used fluorescent proteins and in particular with the DNA marker, is still to be evaluated.

As a further development of this work, we could introduce the study of the role played by the mCherry protein stability in shaping the expression distribution. In our model we showed that the time scale of the protein dynamics, controlled by the degradation rate g_P, can have a strong influence in the final shape of the protein distribution. Indeed, we presented some examples in which a bimodal protein distribution corresponded to a monomodal distribution for the mRNA and vice versa. The fluorescent proteins that we use in our experiments are extremely stable, their half-life is of the order of days. This means that the corresponding g_P is very small and that mRNA fluctuations are very likely buffered away (the protein dynamics "sees" only the average mRNA amount). In order to investigate regimes in which protein turnover is faster, destabilized forms of the fluorescent proteins have to be used. These proteins have a much shorted half-life, which makes mRNA fluctuations not negligible. The comparison of the results obtained with this proteins to the ones obtained with their stable forms, could be an important element for the validation of the model.

6.2 Single-cell tracking in time with fluorescence microscopy

The goal of the second experiment is to investigate the behavior of the correlation between two miRNA targets beyond the steady state. Experiments performed at the flow cytometer return a snapshot of the population of cells at a given time instant, which is assumed to be at the steady state. This technique allows to simultaneously study hundreds of thousands of cells, but it does not provide any information on the time behavior of the system. In order to investigate this aspect, different experimental methods should be used, at the cost of a smaller number of cells that can be measured.

In the following sections we will describe an experimental approach based on fluorescence time-lapse microscopy that will allow us to study temporal correlations between two exogenous miRNA targets in single cells.

6.2.1 State of the art and idea of experiment

Most of the dynamics of a system of genes competing for binding to a common miRNA has not been investigated yet, neither experimentally, nor theoretically. Some theoretical results about response times to an external perturbation were obtained by Bosia et al. in [7]. In this work, response times are defined as the time needed by the system to reach half of the way between the initial and final steady state. Through the numerical integration of the deterministic set of equations associated to a system of N targets competing for the same miRNA, the authors studied the response time of one target of the pool (T_1), in reaction to the sudden over-expression or knock-out of another target (T_2). They found that, as a function of the transcription rate of T_1, the response time to the over-expression presented a maximum in proximity to the threshold, while the response time to the knock-out presented a minimum. These are interesting results and an experimental validation would be important.

Another point regarding the dynamics of the system, consists in the possible synchronization of the fluctuations in proximity to the threshold. Indeed, it was shown theoretically in [7] and experimentally in [8], that a system of two targets competing for one miRNA presents a maximum both of noise and correlation in proximity to the threshold at the steady state. This fact suggests that, in these conditions, the stochastic fluctuations of the amount of mRNA molecules of the two targets could be synchronized in time. This scenario is extremely interesting, since it could shed a different light on the role that miRNAs can have in physiological conditions. For instance, synchronized fluctuations could be very helpful in case of two target genes coding for proteins that have to interact making a dimer.

Both these challenging topics, related to the dynamics of the system, can be studied in vitro with the help of synthetic miRNA targets, by following their expression over time in single cells. For this purpose, we will use two bidirectional plasmids, similar to the ones used in [8], each of which coding for two fluorescent proteins. One gene of each plasmid will be engineered to contain 1, 4 or 7 binding sites for the miRNA miR-20a, the other one will be left as control. The two synthetic

targets will be transfected in HEK 293 cells, which endogenously express miR-20a. These are epithelial cells that live in adhesion to the Petri dish, forming tissue-like structures. Transfected cells will be then followed over time by using a fluorescence microscope able to discriminate the emission spectra of the four fluorescent proteins. Time-lapse movies of a fraction of the population of plated cells will be recorded in each fluorescence channel. By analyzing these movies in order to track each cell from one frame to the following one, single-cell fluorescence trajectories over time in the four channels can be reconstructed. These trajectories will be used to study both the synchronization of fluctuations and the response times, by performing ad hoc stimulation experiments.

As mentioned in section 3.6.3, in order to be able to follow the mRNA fluctuations, the half-life of the fluorescent proteins should be short, comparable to the one of the mRNAs, to avoid accumulation and noise buffering. For this reason, in these experiments the use of destabilized fluorescent proteins will be essential.

In the following section a detailed description of the experimental setup for the two kind of experiment will be provided.

6.2.2 Experimental setup

Two different bidirectional plasmids are used in this experiment. Their structure is the same as those described in section 2.2.5. Each plasmid hosts two genes coding for fluorescent proteins that can be transcribed with the same probability due to the bidirectional promoter. The 3'UTR of one gene in each plasmid is engineered to host a fixed number (0, 1, 4, 7) of miRNA binding sites for the miRNA miR-20a, while, as above, the other one is left as control, to monitor the transcriptional activity of the construct. A scheme of the plasmids is reported in figure 6.6.

The four fluorescent proteins have to be chosen according to two requirements: their half-life has to be short and they must have emission spectra that can be clearly discriminated. The first requirement is necessary in order to be able to follow fluctuations of the amount of target mRNA transcribed and to consequently make meaningful considerations about the synchronization of the correlation. The second requirement is essential to limit the phenomenon of spillover, which can be a significant issue, since we deal with four different fluorescent proteins. Fluorescence measurements are obtained by exciting fluorescent proteins with light, from a UV

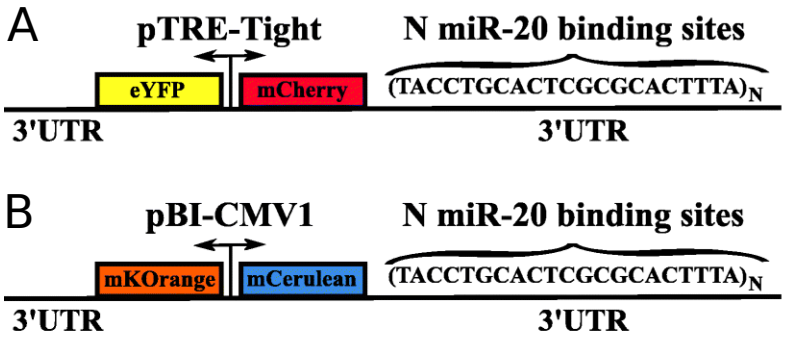

Fig. 6.6 Example of 4-color fluorescent reporter system. The plasmids in figure are the ones used in [8]. Adapted from [8].

lamp or laser, that has a wavelength close to the peak of the excitation spectrum of the protein. As a consequence of excitation, the fluorescent protein emits light according to an emission spectrum which is specific of each protein. Excitation and emission spectra of different fluorescent proteins can overlap. This means that, by exciting one protein, another one can result partially excited, consequently emitting light in its emission spectrum. If the emission spectra partially overlap, the light collected in one channel (i.e. in the range of wavelengths in which we assume to collect the fluorescent light of one protein) can be composed also by some light emitted by another protein. This phenomenon, the so called spillover, is extremely detrimental in our experiments. Indeed, due to the overlap of the spectra, fluctuations in the expression of the protein of interest can be completely masked by fluctuations of another fluorescent protein. Usually, a compensation matrix is computed in order to distinguish the real signal from the spilled one. Nonetheless, this approach is not efficient to distinguish intensity fluctuations. The only way to face this problem consists then in selecting fluorescent proteins with emission or excitation spectra with minimal overlap. In addition to this strategy, an accurate tuning of the width and position of the collection channels is crucial when allowed by the instrument used for the measurements.

As already described above, the two bidirectional plasmids are transfected in HEK 293 cells, which endogenously express miR-20a. Transfections are performed using the Effectene Transfection Reagent (QIAGEN). As described for previous experiments, at the moment of transfection we do not have any control on the number of plasmids that enter the cells and this allows us to investigate a wide range of values of transcriptional activity.

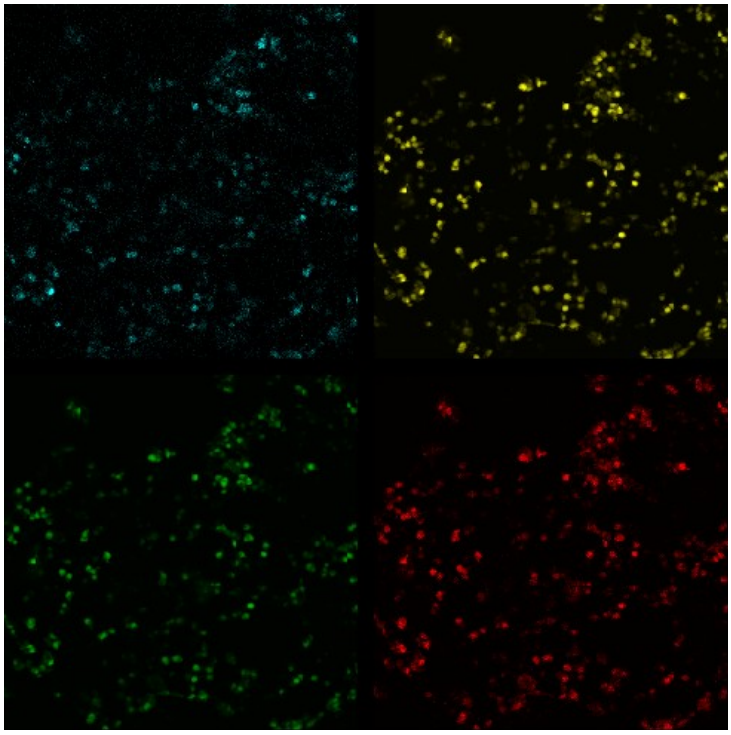

Fig. 6.7 HEK 293 cells transfected with the fluorescent reporter of figure 6.6.

Fluorescence measurements of the transfected cells are performed at a confocal microscope (Leica TCS SP5), leaving the pinhole full open, in order to collect most of the fluorescence light coming from the cell. This microscope has 5 different lasers and allows to tune the light collection channels. The setup of the microscope includes an incubator with controlled temperature and carbon dioxide, which allows to keep cells in physiological conditions while measurements are performed.

Measurements consist in pictures, collecting the fluorescent light in each of the four channels, performed on the same sub-population of the plated cells every 2-5 minutes for several hours. An example is reported in figure 6.7. The fluorescent protein used in this experiment have a nuclear tag, so that they are actively transported to the nucleus, where they accumulate, making easier the localization of the cells.

The number of cells that can be analyzed with this technique is orders of magnitude smaller than the one obtained from a flow cytometer. Indeed, by using the moving plate of the microscope, which automatically select different fields of the population of cells, we can follow up to order 10^3 cells. Nonetheless, this number is sufficient to obtain statistically significant results.

The time-lapse movies are processed with the Matlab software (MathWorks, MA, USA). First, cells are localized in each picture through a segmentation algorithm, then each cell is tracked from one frame to the following one by minimizing the distance from the last known position. Further details on the image processing analysis are reported in the following section. As a result of this analysis, single-cell fluorescence trajectories over time in the four channels are reconstructed. These trajectories can be used to study temporal correlations between the two exogenous targets of the miRNA. Moreover, thorough this approach we can study not only the steady state, but also the transient that, after transfection, leads to the steady state.

The technique described above is at the basis also of the experiments aimed at studying the response times of the system to external perturbations. In particular, we study the response times of the system as a consequence of a sudden switch on or off of one of the two exogenous targets. The switch can be obtained by using a so called TET-ON or TET-OFF system. In both the systems, a tetracycline-inducible promoter (like pTet) is inserted upstream the gene coding for one of the two targets. The second component of the TET system, different between ON and OFF, consists in a gene coding for a tetracycline-controlled transactivator (rtTA for the ON system, tTA for the OFF system). Tetracycline is an antibiotic used in this system as external stimulus to induce or block the expression of a gene controlled by the TET-inducible promoter. When the rtTA transactivator is bound to a tetracycline molecule, it can bind the inducible promoter, activating the expression of the gene. On the contrary, the tTA transactivator can bind the promoter and induce the expression of the gene, only if it not bound to a tetracycline molecule. As a result, in a TET-ON system, tetracycline activates the expression of the inducible gene, while in a TET-OFF system it blocks the expression. In our experiments we use this system to switch on or off the expression of one of the targets. In particular, for this purpose, we exploit two different HEK 293 cell lines, namely HEK 293 TET-ON and HEK 293 TET-OFF, that are engineered in order to endogenously express the corresponding transactivator gene.

Finally, in the experiments described above, the influence of cell cycle can be monitored by staining cells with the Hoechst DNA marker. In this case, the choice of the fluorescent proteins should be made taking into account also the excitation and emission spectra of the markers.

6.2.3 Data analysis: image processing and preliminary results

As described in the previous section, the raw data of these experiments are time-lapse movies of a 2-dimensional layer of cells in four fluorescence channels. The analysis of these data, whose aim is to obtain single-cell fluorescence trajectories over time, consists in two main steps: image segmentation and cell tracking. Image segmentation allows to detect each single cell in one frame, while tracking is necessary to recognize and follow the cells from one frame to the other. The analysis that we will describe is performed with the Matlab software (MathWorks, MA, USA), in addition with the Image Processing Toolbox and the Bio-formats Toolbox [149].

Image segmentation, i.e. the recognition of objects of interest inside an image, is an extremely well studied topic and segmentation algorithms are embedded in many types of commercial and open source software. Nonetheless, the structure of a segmentation algorithm, as well as the values of the parameters used, are extremely sensitive to the specific problem under study. For this reason, we decided to develop a custom segmentation algorithm optimized for our data, exploiting the library of image processing functions already contained in the Matlab Image Processing Toolbox.

Raw data are saved in the $.lif$ Leica proprietary format, in order to open them in Matlab, the "bfopen" function of the Bio-formats Toolbox is used.

The fluorescence intensity of the target proteins can be very low, due to the repression operated by the miRNA. For this reason, the idea is to perform the segmentation on the image obtained by combining the two control channels, reconstructing a mask that can be then used to analyze the fluorescence in the channels of the targets.

As a first step, the contrast of the images in the control channels is adjusted according to fixed parameters, then thresholding is performed, in order to eliminate pixels belonging to the background. The result of thresholding is a binary image in which background pixels have value 0 and foreground pixels have value 1. The threshold value for this operation is estimated by measuring the average intensity of a region of the image in which no cells are present, this value is defined for the first frame and is kept constant throughout the analysis of the movie.

The two binary images obtained from the two control channels are then summed using an AND gate. The resulting binary image has then pixels with value 1 only

where both the control channels presented a signal above threshold. This image will be used to create the mask to identify each cell in the frame.

In order to eliminate noisy pixels and to smooth the borders of the objects, a Gaussian smoothing is performed on the combined image. The resulting image is not binary anymore, another thresholding is then needed to restore its binary form.

In order to move towards the most important step of this segmentation process, the Euclidean distance transform of the last image is computed. Result of this operation is a new image in which to each pixel with value 1 in the input image is assigned a number that is the distance between that pixel and the nearest pixel with value 0.

The distance transform is used as input image of the Watershed transform. This operation treats the input image as a surface with high and low elevations and identifies and labels catchment basins. In this way, pixels of the image belonging to the different cells are correctly labeled, but some artifacts, like very small objects, still remain.

In order to eliminate very small unwanted objects and to regularize the borders, a morphological opening is performed on the image resulting from the Watershed transform. Morphological opening consists in an erosion of the foreground objects of an image, followed by a dilation. Both erosion and dilation are performed with the same structuring element. In this way, small objects that disappear after the erosion step cannot be recovered by the dilation step and are eliminated from the image.

Finally, slightly bigger objects that remain after the opening operation but are not cells, are eliminated by performing a thresholding on the size of the identified objects. The value of the threshold is manually tuned.

The objects of the final image are labeled again, obtaining the mask that identifies pixels belonging to each cell of the original image, see Fig. 6.8.

The fluorescence intensity of each object identified by the mask is computed, in each channel, by taking the average of the values of the pixels that compose the object. Background intensity is also monitored and its average value is subtracted to the value computed for each object.

So far, we described the workflow to identify cells and measure their average fluorescence intensity in each frame of the movie. To reconstruct single-cell fluorescence trajectories, we must be able to recognize cells from one frame to the other. Since

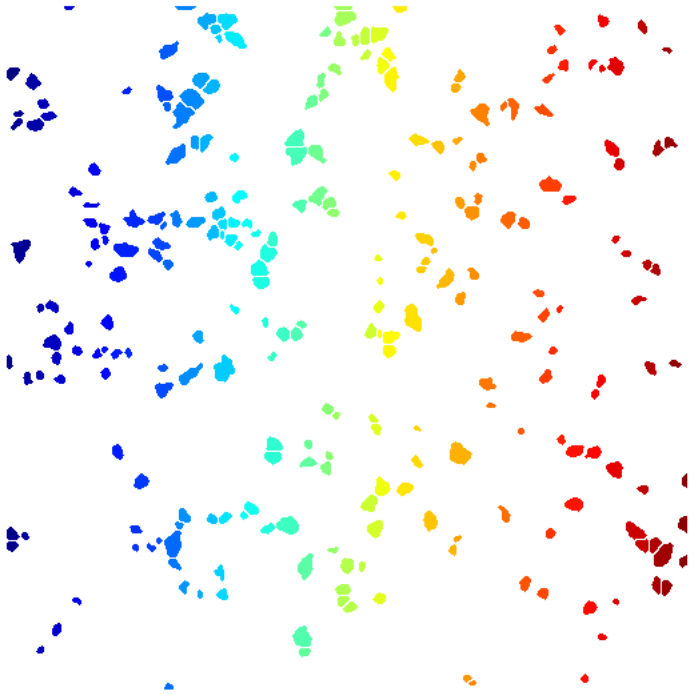

Fig. 6.8 Example of cell segmentation. Different colors have no specific meaning.

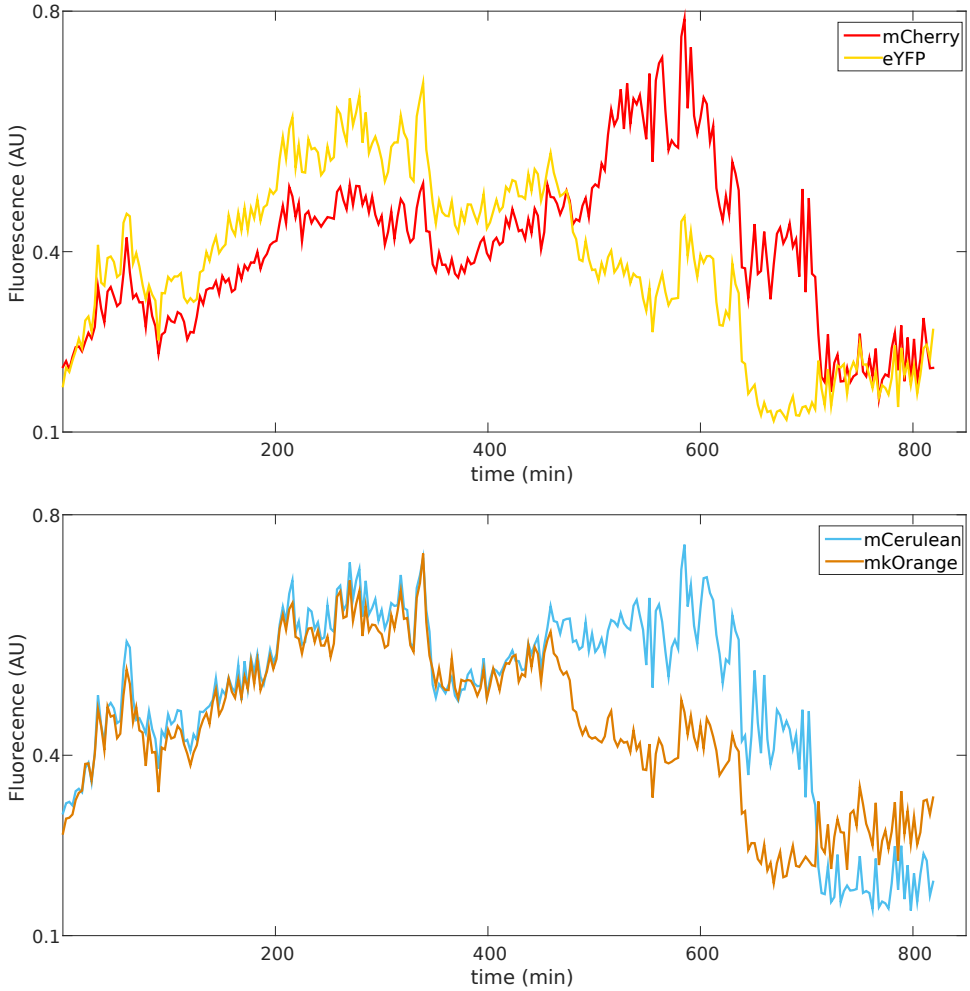

Fig. 6.9 Example of single-cell fluorescence trajectory over time in the four channels.

the duration of the time-lapse movies can be up to 48 hours, it is very likely that some cells divide into two daughter cells during the time of the measurement. For this reason, the tracking procedure is performed backward in time. This trick makes it easier to deal with cell divisions, in fact, when such event occurs, the trajectory of one of the two cells simply stops.

Tracking of the identified cells from one frame to the other is performed by minimizing the distance of the centroids. The procedure to track cells from frame A to frame B is the following.

1. The position of the centroid of each cell identified in frame A is computed.

2. The position of the centroid of each cell identified in frame B is computed.

3. The size of the area in which the change of position of the cell is searched, is defined on the basis of a trial-and-error analysis performed in phase of optimization of the algorithm.

4. Each cell identified in frame B is associated to the cell of frame A that minimizes the distance of their centroids, in the limits of the search area. Once a match is found, the cell of frame B receives the same label of the corresponding cell of frame A.

5. The position of cells of frame A that do not have any correspondence with cells of frame B is recorded and stored in a vector, together with their label.

6. A second research step is performed for cells of frame B for which a correspondence with frame A was not found. For each of these cells, the search is performed on the position of the cells that disappeared in previous frames, which is recorded in the vector of step 5. If a match is found, the cell receives the label of the previously disappeared cell. Otherwise, a new label is assigned to the cell.

Following this procedure, single-cell fluorescence trajectories can be reconstructed. Some examples are reported in figure 6.9. These trajectories are obtained from experiments in which the plasmids are the ones used by Bosia et al. in [8].

The core of these single-cell tracking experiments is still to be performed, but the analysis tools are ready to be employed.

6.2.4 Spin-off: tracking vesicle trafficking and counting cells

The analysis setup developed for these experiments was successfully applied, with appropriate modifications, in two different experimental projects, in which image analysis played an important role. In the following, we will briefly describe the contribute given by our analysis in these two contexts.

The first project [56] regards the study of vesicle trafficking in mammalian cells. In particular, the aim of the experiment was to unravel how the protein Rab11 controls transport of recycling cargo from early endosomes to the endocytic recycling

compartment. The analysis was performed by using a novel genetically-encoded FRET biosensor for Rab11. FRET stands for Fluorescence Resonance Energy Transfer or Förster Resonance Energy Transfer. This phenomenon involves two light-sensitive molecules (two fluorescent proteins fused with Rab11, in case of the biosensor): a donor and an acceptor, with the emission spectrum of the donor sufficiently overlapping with the excitation one of the acceptor. If the two molecules are sufficiently close one to the other, when the donor is in its excited state, it can transfer energy to the acceptor, exciting it and making it emit fluorescence light. This effect is extremely sensitive to the distance between donor and acceptor, indeed, its efficiency is inversely proportional to the sixth power of this distance. For this reason, FRET is used in biosensors to detect small changes of the configuration of molecules due to structural modifications. In the biosensor designed for these experiments, when the Rab11 protein is in its inactive form, the two fluorescent proteins are relatively far one from the other, this leading to a low FRET emission. When Rab11 is activated, the two fluorescent proteins get closer and FRET emission increases. This property is used to monitor Rab11 activity during vesicle trafficking, by performing time-lapse movies measuring the fluorescence emission of the two proteins alone, together with the FRET emission. In order to analyze these data, vesicles were segmented and tracked from frame to frame, obtaining FRET emission trajectories over time. These trajectories were used to analyze the temporal behavior of Rab11 activation at the single-vesicle level. This analysis was performed by adapting to this context the segmentation and tracking algorithm described in the previous section.

The second project [57] concerns the quantitative physiology of cancer cells. In particular, it's aim is the identification of connections between cancer cell physiology, cell growth and internal partitioning of resources inside the cell. In this framework, experiments were conducted in order to obtain complete growth curves of populations of human leukemia cells (Jurkat). The idea behind these experiments is that a variation of the cells' growth rate in the exponential phase could correspond to a reorganization of the cell proteome. In order to obtain the growth curves and estimate the growth rate, the population of cells should be counted every day for several days, until the growth medium is saturated and the system reaches its carrying capacity. Cell counting was performed by taking a homogeneous sample of the cell population and inserting it in a micro-chamber with fixed, known volume. Then, bright-field pictures of portions of the chamber were taken at the microscope. By counting the

number of cells present inside a known area of each picture, the concentration of the population of cells can be estimated. This counting was performed by adapting the segmentation algorithm described above, in order to work with bright-field images instead of fluorescence images. This procedure resulted quite efficient, leading to an error comparable to the one obtained counting cells by hand.

References

[1] N. E. Buchler and M. Louis. Molecular titration and ultrasensitivity in regulatory networks. *Journal of molecular biology*, 384:1106–1119, 2008.

[2] F. Fazi and C. Nervi. Microrna: Basic mechanisms and transcriptional regulatory networks for cell fate determination. *Cardiovascular research*, 79:553–561, 2008.

[3] D. Bartel. MicroRNAs: target recognition and regulatory functions. *Cell*, 136:215 – 233, 2009.

[4] S. Mukherji *et al.* MicroRNAs can generate thresholds in target gene expression. *Nature Genetics*, 43(9):854–859, 2011.

[5] J. M. Schmiedel *et al.* MicroRNA control of protein expression noise. *Science*, 348:128–132, 2015.

[6] X. Lai and J. Vera. MicroRNA Regulation, Feed-Forward Loops. In W. Dubitzky, O. Wolkenhauer, K. H. Cho, and H. Yokota, editors, *Encyclopedia of Systems Biology*, pages 1324–1328. Springer, 2013.

[7] C. Bosia, A. Pagnani, and R. Zecchina. Modelling competing endogenous RNA networks. *PLoS One*, 8(6):e66609, 2013.

[8] C. Bosia *et al.* Rnas competing for microRNAs mutually influence their fluctuations in a highly non-linear microRNA-dependent manner in single cells. *Genome Biology*, 18:37, 2017.

[9] **M. Del Giudice**, C. Bosia, S. Grigolon, and S. Bo. Stochastic sequestration dynamics: a minimal model with extrinsic noise for bimodal distributions and competitors correlation. *Scientific Reports*, 8(1):10387, 2018.

[10] **M. Del Giudice**, S. Bo, S. Grigolon, and C. Bosia. On the role of extrinsic noise in microRNA-mediated bimodal gene expression. *PLOS Computational Biology*, 14(4):1–26, 04 2018.

[11] R. C. Lee, R. L. Feinbaum, and V. Ambros. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 75(5):843–854, 1993.

[12] B. J. Reinhart *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, 403:901–906, 2000.

[13] M. Lagos-Quintana, R. Rauhut, W. Lendeckel, and T. Tuschl. Identification of novel genes coding for small expressed RNAs. *Science*, 294:853–858, 2001.

[14] N. C. Lau, L. P. Lim, E. G. Weinstein, and D. P. Bartel. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science*, 294:858–862, 2001.

[15] R. C. Lee and V. Ambros. An extensive class of small RNAs in Caenorhabditis elegans. *Science*, 294:862–864, 2001.

[16] D. P. Bartel. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116(2):281–297, 2004.

[17] N. Bushati and S. M. Cohen. MicroRNA Functions. *Annual Review of Cell and Developmental Biology*, 23(1):175–205, 2007. PMID: 17506695.

[18] I. Alvarez-Garcia and E. A. Miska. MicroRNA functions in animal development and human disease. *Development*, 132(21):4653–4662, 2005.

[19] A. Esquela-Kerscher and F. J. Slack. Oncomirs — microRNAs with a role in cancer. *Nature Reviews Cancer*, 6(4):259–269, 2006.

[20] U. Ala *et al.* Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments. *Proceedings of the National Academy of Sciences*, 110(18):7154–7159, 2013.

[21] M. Figliuzzi, E. Marinari, and A. De Martino. MicroRNAs as a selective channel of communication between competing RNAs: a steady-state theory. *Biophysical Journal*, 104(5):1203–1213, 2013.

[22] X. Lai, O. Wolkenhauer, and J. Vera. Understanding microRNA-mediated gene regulatory networks through mathematical modelling. *Nucleic Acids Research*, 44:6019–6035, 06 2016.

[23] J. M. Luna *et al.* Hepatitis C virus RNA functionally sequesters miR-122. *Cell*, 160:1099–1110, 2015.

[24] C. M. Pickart and M. J. Eddins. Ubiquitin: structures, functions, mechanisms. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1695:55–72, 2004.

[25] D. E. Klein, V. M. Nappi, G. T. Reeves, and S. Y. Shvartsman. Argos inhibits epidermal growth factor receptor signalling by ligand sequestration. *Nature*, 430:1040–1044, 2004.

[26] M. Yanagita. BMP antagonists: Their roles in development and involvement in pathophysiology. *Cytokine and Growth Factor Reviews*, 16:309–317, 2005.

[27] M. Van Doren, H. M. Ellis, and J. W. Posakony. The Drosophila extramacrochaetae protein antagonizes sequence-specific DNA binding by daughterless/achaete-scute protein complexes. *Development*, 113:245–255, 1991.

[28] R. Benezra, R. L. Davis, D. Lockshon, D. L. Turner, and H. Weintraub. The protein id: A negative regulator of helix-loop-helix dna binding proteins. *Cell*, 61:49–59, 1990.

[29] I. L. Grigorova, N. J. Phleger, V. K. Mutalik, and C. A. Gross. Insights into transcriptional regulation and σ competition from an equilibrium model of rna polymerase binding to dna. *Proceedings of the National Academy of Sciences*, 103:5332–5337, 2006.

[30] E. Rotem *et al.* Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. *Proceedings of the National Academy of Sciences*, 107:12541–12546, 2010.

[31] R. C. Friedman, K. K. Farh, C. B. Burge, and D. P. Bartel. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, 19(1):92–105, 2009.

[32] M. Thomas, J. Lieberman, and A. Lal. Desperately seeking microRNA targets. *Nature Structural and Molecular Biology*, 17:1169–1174, 2010.

[33] A. Krek *et al.* Combinatorial microRNA target predictions. *Nature Genetics*, 37:495–500, 2005.

[34] H. Seitz. Redefining microRNA targets. *Current Biology*, 19:870–873, 2009.

[35] F. Karreth *et al.* In vivo identification of tumor-suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell*, 147:382–395, 2011.

[36] A. Jacobsen, J. Wen, D. S. Marks, and A. Krogh. Signatures of rna binding proteins globally coupled to effective microrna target sites. *Genome Research*, 20:1010–1019, 2010.

[37] L. Salmena, L. Poliseno, Y. Tay, L. Kats, and P. P. Pandolfi. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*, 146(3):353–358, 2011.

[38] Y. Tay *et al.* Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell*, 147(2):344–357, 2011.

[39] L. Poliseno *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*, 465:1033–1038, 2010.

[40] M. Cesana *et al.* A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*, 147(2):358–369, 2011.

[41] A. Raj and A. van Oudenaarden. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135:216–226, 2008.

[42] M. Kaern, T. Elston, W. Blake, and J. Collins. Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics*, 6(6):451–464, 2005.

[43] P. Swain, M. Elowitz, and E. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, 99(20):12795–12800, 2002.

[44] M. Elowitz, A. Levine, E. Siggia, and P. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

[45] L. S. Tsimring. Noise in biology. *Reports on Progress in Physics*, 77:026601, 2014.

[46] B. Munsky, G. Neuert, and A. Van Oudenaarden. Using gene expression noise to understand gene regulation. *Science*, 336:183–187, 2012.

[47] V. Shahrezaei, J. Ollivier, and P. Swain. Colored extrinsic fluctuations and stochastic gene expression. *Molecular Systems Biology*, 4:196, 2008.

[48] D. Volfson, J. Marciniak, W. Blake, N. Ostroff, L. Tsimring, and J. Hasty. Origins of extrinsic variability in eukaryotic gene expression. *Nature*, 439:861–864, 2006.

[49] M. Bessarabova *et al.* Bimodal gene expression patterns in breast cancer. *BMC Genomics*, 11(Suppl I):S8, 2010.

[50] A. K. Shalek *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*, 498:236–240, 2013.

[51] Y. Nikolsky *et al.* Sequential clustering of breast cancers using bimodal gene expression. *Proceed AACR Ann Meeting*, page 141, 2007.

[52] I. Dozmorov *et al.* Hypervariable genes–experimental error or hidden dynamics. *Nucleic Acids Research*, 32(19):e147–10, 2004.

[53] H. Zhao, P. Yue, and K. Fang. Identification of differentially expressed genes with multivariate outlier analysis. *J Biopharm Stat.*, 14(3):629–646, 2004.

[54] A. Teschendorff *et al.* PACK: Profile analysis using clustering and kurtosis to find molecular classifiers in cancer. *Bioinformatics*, 22(18):2269–2275, 2006.

[55] A. Martirosyan, **M. Del Giudice**, C. Enrico Bena, A. Pagnani, C. Bosia, and A. De Martino. Kinetic modelling of competition and depletion of shared miRNAs by competing endogenous RNAs. *Methods in Molecular Biology - Under revision*, 2018.

[56] C. C. Campa, J. P. Margaria, A. Derle, **M. Del Giudice**, M. C. De Santis, L. Gozzelino, F. Copperi, C. Bosia, and E. Hirsch. Rab11 activity and PtdIns(3)P turnover removes recycling cargo from endosomes. *Nature Chemical Biology*, 2018.

[57] C. Enrico Bena, **M. Del Giudice**, T. Gueudré, M. Miotto, E. Turco, A. De Martino, and C. Bosia. Inoculum-density dependent growth reveals inherent cooperative effects and stochasticity in cancer cell cultures. *arXiv:1710.10978*, 2017.

[58] A. V. Hill. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J Physiol*, 40:iv–vii, 1910.

[59] G. S. Adair. The hemoglobin system. IV. The oxygen dissociation curve of hemoglobin. *J Biol Chem*, 63:529–545, 1910.

[60] H. Abeliovich. An empirical extremum principle for the Hill coefficient in ligand-protein interactions showing negative cooperativity. *Biophysical Journal*, 89(1):76–79, 1910.

[61] M. I. Stefan and N. Le Novère. Cooperative binding. *PLOS Computational Biology*, 9(6):e1003106, 2013.

[62] A. Flynt and E. Lai. Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nature Reviews Genetics*, 9:831– 842, 2008.

[63] D. P. Bartel. Metazoan MicroRNAs. *Cell*, 173(1):20 – 51, 2018.

[64] L. He and G. J. Hannon. MicroRNAs: small RNAs with a big role in gene regulation. *Nature Reviews Genetics*, 5(7):522–531, 2004.

[65] D. Cora', A. Re, M. Caselle, and F. Bussolino. MicroRNA-mediated regulatory circuits: outlook and perspectives. *Physical Biology*, 14(4):045001, 2017.

[66] A. Kozomara and S. Griffiths-Jones. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, 42(Database issue):D68–D73, 2014.

[67] N. S. Asli, M. E. Pitulescu, and M. Kessel. MicroRNAs in organogenesis and disease. *Current Molecular Medicine*, 8(8):698–710, 2008.

[68] A. Shrestha *et al.* MicroRNA-142 is a multifaceted regulator in organogenesis, homeostasis, and disease. *Developmental Dynamics*, 246:285–290, 2017.

[69] Y. Zhao and D. Srivastava. A developmental view of microRNA function. *Trends in Biochemical Sciences*, 32(4):189–197, 2007.

[70] A. Cimmino *et al.* miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proceedings of the National Academy of Sciences*, 102:13944–13949, 2005.

[71] Z. Su *et al.* MicroRNAs in apoptosis, autophagy and necroptosis. *Oncotarget*, 6(11):8474–8490, 2015.

[72] R. Brosh *et al.* p53-repressed miRNAs are involved with E2F in a feed-forward loop promoting proliferation. *Molecular Systems Biology*, 4:229, 2008.

[73] D. Lenkala *et al.* The impact of microRNA expression on cellular proliferation. *Human Genetics*, 133(7):931–938, 2014.

[74] L. Bai *et al.* Microrna-532 and microrna-3064 inhibit cell proliferation and invasion by acting as direct regulators of human telomerase reverse transcriptase in ovarian cancer. *PLOS ONE*, 12(3):1–13, 2017.

[75] P. Bu *et al.* A miR-34a-Numb feedforward loop triggered by inflammation regulates asymmetric stem cell division in intestine and colon cancer. *Cell Stem Cell*, 18(2):189 – 202, 2016.

[76] H. W. Hwang and J. T. Mendell. MicroRNAs in cell proliferation, cell death, and tumorigenesis. *British Journal of Cancer*, 94(6):776–780, 2006.

[77] M. D. Jansson and A. H. Lund. MicroRNA and cancer. *Molecular Oncology*, 6(6):590–610, 2012.

[78] G. Di Leva, M. Garofalo, and C. M. Croce. MicroRNAs in cancer. *Annual Review of Pathology: Mechanisms of Disease*, 9(1):287–314, 2014. PMID: 24079833.

[79] P. Ramalingam *et al.* Biogenesis of intronic miRNAs located in clusters by independent transcription and alternative splicing. *RNA*, 20(1):76–87, 2014.

[80] C. H. Chien *et al.* Identifying transcriptional start sites of human microRNAs based on high-throughput sequencing data. *Nucleic Acids Research*, 39(21):9345–9356, 2011.

[81] Y. Lee *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO Journal*, 23(20):4051–4060, 2004.

[82] A. Rodriguez, S. Griffiths-Jones, J. L. Ashurst, and A. Bradley. Identification of Mammalian microRNA Host Genes and Transcription Units. *Genome Research*, 14(10a):1902–1910, 2004.

[83] Y-K Kim and V. N. Kim. Processing of intronic microRNAs. *EMBO Journal*, 26:775–783, 2007.

[84] J. G. Ruby, C. H. Jan, and D. P. Bartel. Intronic microRNA precursors that bypass Drosha processing. *Nature*, 448:83–86, 2007.

[85] H. R. Chiang *et al.* Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes & Development*, 24:992–1009, 2010.

[86] I. Slezak-Prochazka *et al.* Cellular localization and processing of primary transcripts of exonic microRNAs. *PLOS ONE*, 8(9):e76647, 09 2013.

[87] V. N. Kim and J. W. Nam. Genomics of microrna. *Trends in Genetics*, 22:165–173, 2006.

[88] T. Tam. Identification and characterization of human BIC, a gene on chromosome 21 that encodes a noncoding RNA. *Gene*, 274(1):157 – 167, 2001.

[89] Z. Melamed *et al.* Alternative splicing regulates biogenesis of miRNAs located across exon-intron junctions. *Molecular Cell*, 50(6):869–881, 2013.

[90] V. Boivin, G. Deschamps-Francoeur, and M. S. Scott. Protein coding genes as hosts for noncoding RNA expression. *Seminars in Cell & Developmental Biology*, 75:3 – 12, 2018.

[91] J. Wang *et al.* Systematic study of cis-antisense miRNAs in animal species reveals miR-3661 to target PPP2CA in human cells. *RNA*, 22(1):87–95, 2016.

[92] V. Narry Kim. MicroRNA biogenesis: coordinated cropping and dicing. *Nature Reviews Molecular Cell Biology*, 6:376–385, 2005.

[93] H. Zhou and I. Rigoutsos. MiR-103a-3p targets the 5' UTR of GPRC5A in pancreatic cells. *RNA*, 20:1431–1439, 2014.

[94] J. .J Forman, A. Legesse-Miller, and H. A. Coller. A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proceedings of the National Academy of Sciences*, 105:14879–14884, 2008.

[95] Y. Tay, J. Zhang, A. M. Thomson, B. Lim, and I. Rigoutsos. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*, 455:1124–1128, 2008.

[96] M. Leclercq, A. B. Diallo, and M. Blanchette. Prediction of human mirna target genes using computationally reconstructed ancestral mammalian sequences. *Nucleic Acids Research*, 45(2):556–566, 2017.

[97] R. Zhang *et al.* Long non-coding RNA PTENP1 functions as a ceRNA to modulate PTEN level by decoying miR-106b and miR-93 in gastric cancer. *Oncotarget*, 8(16):26079–26089, 2017.

[98] D. P. Bartel and C. Z. Chen. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nature Reviews Genetics*, 5(5):396–400, 2004.

[99] M. S. Ebert and P. A. Sharp. Roles for MicroRNAs in Conferring Robustness to Biological Processes. *Cell*, 149(3):515 – 524, 2012.

[100] A. Stark, J. Brennecke, N. Bushati, R. B. Russell, and S. M. Cohen. Animal MicroRNAs Confer Robustness to Gene Expression and Have a Significant Impact on 3'UTR Evolution. *Cell*, 123(6):1133 – 1146, 2005.

[101] A. Shkumatava, A. Stark, H. Sive, and D. P. Bartel. Coherent but overlapping expression of microRNAs and their targets during vertebrate development. *Genes & Development*, 23(4):466–481, 2009.

[102] K. K. Farh *et al.* The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science*, 310:1817–1821, 2005.

[103] D. Baek, J. Villén, C. Shin, F. D. Camargo, S. P. Gygi, and D. P. Bartel. The impact of microRNAs on protein output. *Nature*, 455:64–71, 2008.

[104] M. Selbach *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature*, 455:58–63, 2008.

[105] P. S. Swain. Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *Journal of Molecular Biology*, 344(4):965 – 976, 2004.

[106] A. Re, D. Cora', D. Taverna, and M. Caselle. Genome-wide survey of microRNA-transcription factor feed-forward regulatory circuits in human. *Molecular Biosystems*, 5(8):854–867, 2009.

[107] J. Tsang, J. Zhu, and A. Van Oudenaarden. MicroRNA-Mediated Feedback and Feedforward Loops Are Recurrent Network Motifs in Mammals. *Molecular Cell*, 26(5):753–767, 2007.

[108] M. Osella, C. Bosia, D. Cora', and M. Caselle. The Role of Incoherent MicroRNA-Mediated Feedforward Loops in Noise Buffering. *PLOS Computational Biology*, 7(3):1–16, 2011.

[109] E. S. Balakirev and F. J. Ayala. Pseudogenes: Are They "Junk" or Functional DNA? *Annual Review of Genetics*, 37(1):123–151, 2003. PMID: 14616058.

[110] R. C. Pink *et al.* Pseudogenes: Pseudo-functional or key regulators in health and disease? *RNA*, 17(5):792–798, 2011.

[111] Z. Kai and A. Pasquinelli. Microrna assassins: factors that regulate the disappearance of mirnas. *Nature Structural and Molecular Biology*, 17:5–10, 2010.

[112] G. Hutvágner and P. Zamore. A microRNA in a Multiple-Turnover RNAi Enzyme Complex. *Science*, 297:2056–2060, 2002.

[113] B. Haley and P. Zamore. Kinetic analysis of the RNAi enzyme complex. *Nature Structural and Molecular Biology*, 11(7):599–606, 2004.

[114] J. Liu *et al.* A role for the P-body component GW182 in microRNA function. *Nature Cell Biology*, 7(12):1261–1266, 2005.

[115] R. Pillai *et al.* Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science*, 309:1573–1576, 2005.

[116] A. Baccarini *et al.* Kinetic analysis reveals the fate of a microRNA following target regulation in mammalian cells. *Current Biology*, 21(5):369–376, 2011.

[117] Y. Yuan *et al.* Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. *Proceedings of the National Academy of Sciences*, 112(10):3158–3163, 2015.

[118] D. Del Vecchio, A. J. Ninfa, and E. D. Sontag. Modular cell biology: retroactivity and insulation. *Molecular Systems Biology*, 4:161, 2008.

[119] L. Pantoja-Hernández and J. C. Martínez-García. Retroactivity in the context of modularly structured biomolecular systems. *Frontiers in Bioengineering and Biotechnology*, 3:85, 2015.

[120] S. Tanase-Nicola, P. B. Warren, and P. R. ten Wolde. Signal detection, modularity, and the correlation between extrinsic and intrinsic noise in biochemical networks. *Physical Review Letters*, 97:068102, 2006.

[121] N. G. Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.

[122] D. T. Gillespie. A rigorous derivation of the chemical master equation. *Phys. A: Stat. Mech. its Appl.*, 188:404–425, 1992.

[123] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover Books on Mathematics, 1965.

[124] I. Gradshteyn and I. Ryzhik. *Table of Integrals, Series and Products*. Academic Press, 1994.

[125] D. A. McQuarrie. Stochastic approach to chemical kinetics. *J. applied probability*, 4:413–478, 1967.

[126] T. Lindberg. Scale-space for discrete signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3):234–254, 1990.

[127] P. Thomas and R. Popovic, N. ans Grima. Phenotypic switching in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 111(19):6994–6999, 2014.

[128] J. Elf and M. Ehrenberg. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Research*, 13:2475–2484, 2003.

[129] C. Gardiner. *Handbook of stochastic methods*. Springer Verlag, Berlin, 1985.

[130] H. Risken. *The Fokker-Planck equation*. Springer Verlag, Berlin, 1984.

[131] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

[132] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[133] D. T. Gillespie. Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*, 58(1):35–55, 2007.

[134] D. T. Gillespie. *Markov Processes: An Introduction for Physical Scientists*. San Diego: Academic, 1992.

[135] W. P. Kloosterman and R. H. A. Plasterk. The diverse functions of microRNAs in animal development and disease. *Developmental Cell*, 11(4):441–450, 2006.

[136] C. Alberti and L. Cochella. A framework for understanding the roles of miRNAs in animal development. *Development*, 144:2548–2559, 2017.

[137] D. Sayed and M. Abdellatif. MicroRNAs in development and disease. *Development*, 91(3):827–887, 2011.

[138] B. Schwanh ausser *et al.* Global quantification of mammalian gene expression control. *Nature*, 473:337–342, 2011.

[139] G. Marinov *et al.* From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Research*, 24:496–510, 2014.

[140] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[141] D. Foster and P. Grassberger. Lower bounds on mutual information. *Physical Review E*, 83:010101, 2011.

[142] M. T. Pickering, B. M. Stadler, and T. F. Kowalik. miR-17 and miR-20a temper an E2F1-induced G1 checkpoint to regulate cell cycle progression. *Oncogene*, 28:140–145, 2009.

[143] X. Fan *et al.* miR-20a promotes proliferation and invasion by targeting APP in human ovarian cancer cells. *Acta Biochimica et Biophysica Sinica*, 42(5):318–324, 2010.

[144] S. Zhao *et al.* MiR-20a promotes cervical cancer proliferation and metastasis in vitro and in vivo. *PLOS ONE*, 10(3):1–11, 03 2015.

[145] F. Ai *et al.* miR-20a regulates proliferation, differentiation and apoptosis in P19 cell model of cardiac differentiation by targeting Smoothened. *Biology Open*, 5(9):1260–1265, 2016.

[146] F. Porichis *et al.* High-throughput detection of miRNAs and gene-specific mRNA at the single-cell level by flow cytometry. *Nature Communications*, 5:5641, 2014.

[147] M. Wu, M. E. Piccini, and A. K. Singh. MiRNA detection at single-cell resolution using microfluidic LNA flow-FISH. *Methods in Molecular Biology*, 1211:245–260, 2014.

[148] C. Lai, D. Stepniak, L. Sias, and C. Funatake. A novel flow cytometric assay for single-cell analysis of microRNA, messenger RNA and protein expression during cell differentiation and maturation. *Journal of Immunology*, 198(1 Supplement):157.12, 2017.

[149] BIO-FORMATS, https://www.openmicroscopy.org/bio-formats/.

[150] C. Zhang, B. E. Mapes, and B. J. Soden. Bimodality in tropical water vapour. *Journal of the Royal Meteorological Society*, 129:2847–2866, 2003.

Appendix A

Bimodality amplitude

Throughout this work, we addressed bimodality as a binary quantity, i.e. presence of two distinct peaks in the distribution. However, how probability distributions change shape during the unimodal-bimodal transition is also an interesting point. To give an insight into this aspect, we computed the bimodality amplitude [150], a coefficient that allows to quantify the level of bimodality of the distribution. This quantity is defined as follows:

$$A_B = \frac{A_1 - A_{an}}{A_1} \tag{A.1}$$

where A_1 is the amplitude of the smallest peak, while A_{an} is amplitude of the local minimum between the two peaks. The bimodality amplitude is by definition smaller than 1: the higher the values, the more separated the maxima. To study how this coefficient varies as a function of the extrinsic noise in the system and the target transcription rate, we first sampled several different targets' probability distributions. By interpolating these distributions with built-in functions in Matlab, we extracted the positions of the maxima and minima as well as their frequency which allowed us to compute the bimodality amplitude. The resulting phase diagram is shown is Fig. A.1. Bimodal distributions appear for a very well-defined range of parameters in agreement with what found in Fig. 3.16. The transition from unimodal to bimodal appears to be peaked around some particular values of the noise coefficient of variation and the target transcription rate, suggesting the existence of optimal values for these parameters. However, it would be even more interesting to study in the future how smoothly bimodal distributions appear in this parameters'

space and whether this observed peak could really be an optimal point for cell variability.

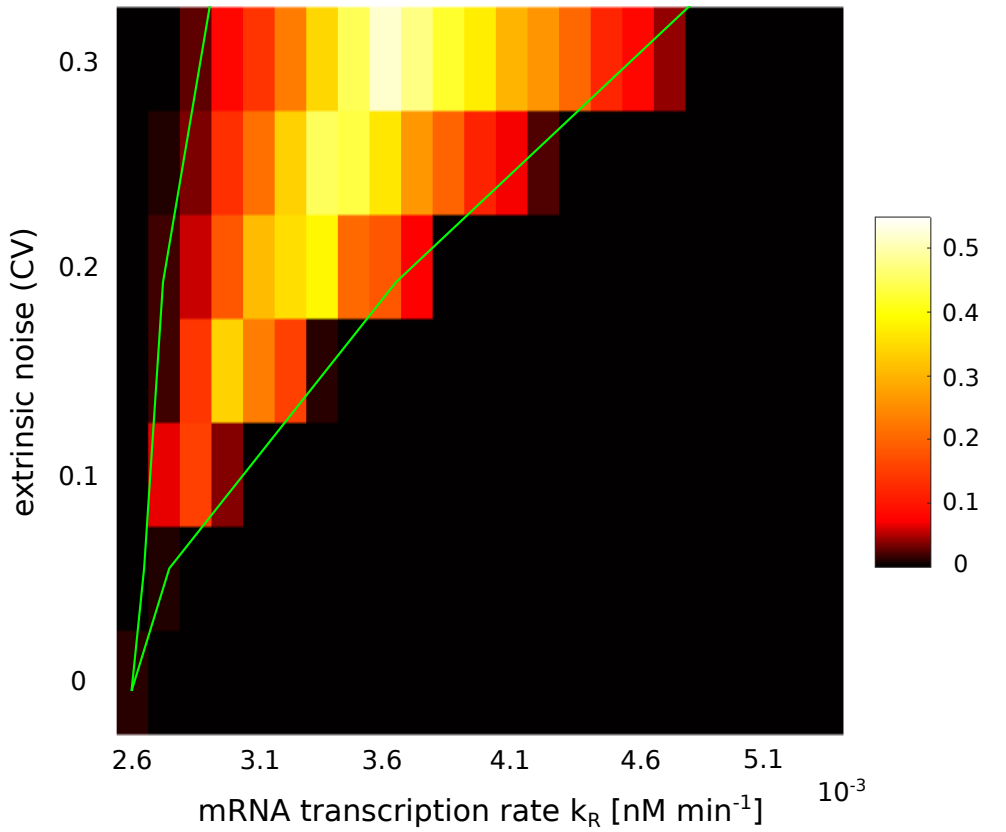

Fig. A.1 Bimodality amplitude phase diagram. Phase diagram of the bimodality amplitude of the mRNA distribution as a function of the mRNA transcription rate k_R and of the extrinsic noise level. The parameters here used coincide with the ones of figure 3.16. Target mRNA transcription rate is one of the control parameters and ranges from $k_R = 2.6 \times 10^{-3}$ nM min^{-1} to $k_R = 5.1 \times 10^{-3}$ nM min^{-1}. Extrinsic noise is tuned by varying the standard deviation of the distribution with mean $\bar{k}_S = 1.2 \times 10^{-3}$ nM min^{-1} from which miRNA transcription rates are picked. The standard deviation ranges from $\sigma = 0$ nM min^{-1} (no extrinsic noise) to $\sigma = 3.6 \times 10^{-4}$ nM min^{-1}. This phase diagram was obtained by interpolating the single distributions obtained from numerical simulations. The green line represents the separation between bimodal and unimodal regions as shown in Fig. 3.16. Adapted from [10].

Appendix B

Role of extrinsic noise in competitive inhibition kinetics

In previous chapters we introduced and discussed a minimal model of molecular titration and studied its connections to miRNA regulation in relation to the presence of some extrinsic noise. In this chapter we will briefly analyze the implications of the titration mechanism in a different context. Indeed, we will study a simple model of enzymatic kinetics, called competitive inhibition, in which enzyme activation is added to the sequestration mechanism. As a result, we will show that all the predictions derived from the minimal model still hold also for this system. This suggests that our results could be significant for different biological processes.

B.1 Bimodal distributions in competitive inhibition kinetics

Competitive inhibition kinetics is a process based on the interaction between an enzyme and its inhibitor. The inhibitor plays the role of the sequester and the enzyme is the target. A free molecule of target, T_F , can bind to a substrate (which is assumed to be at fixed concentration) and form the active enzyme T_A from which the product is made. The activation of the target occurs with a rate that depends on the intrinsic activation rate k_f and on the concentration of substrate c_S . The active target molecule, T_A , can be deactivated with rate k_r becoming T_F. The free molecule of

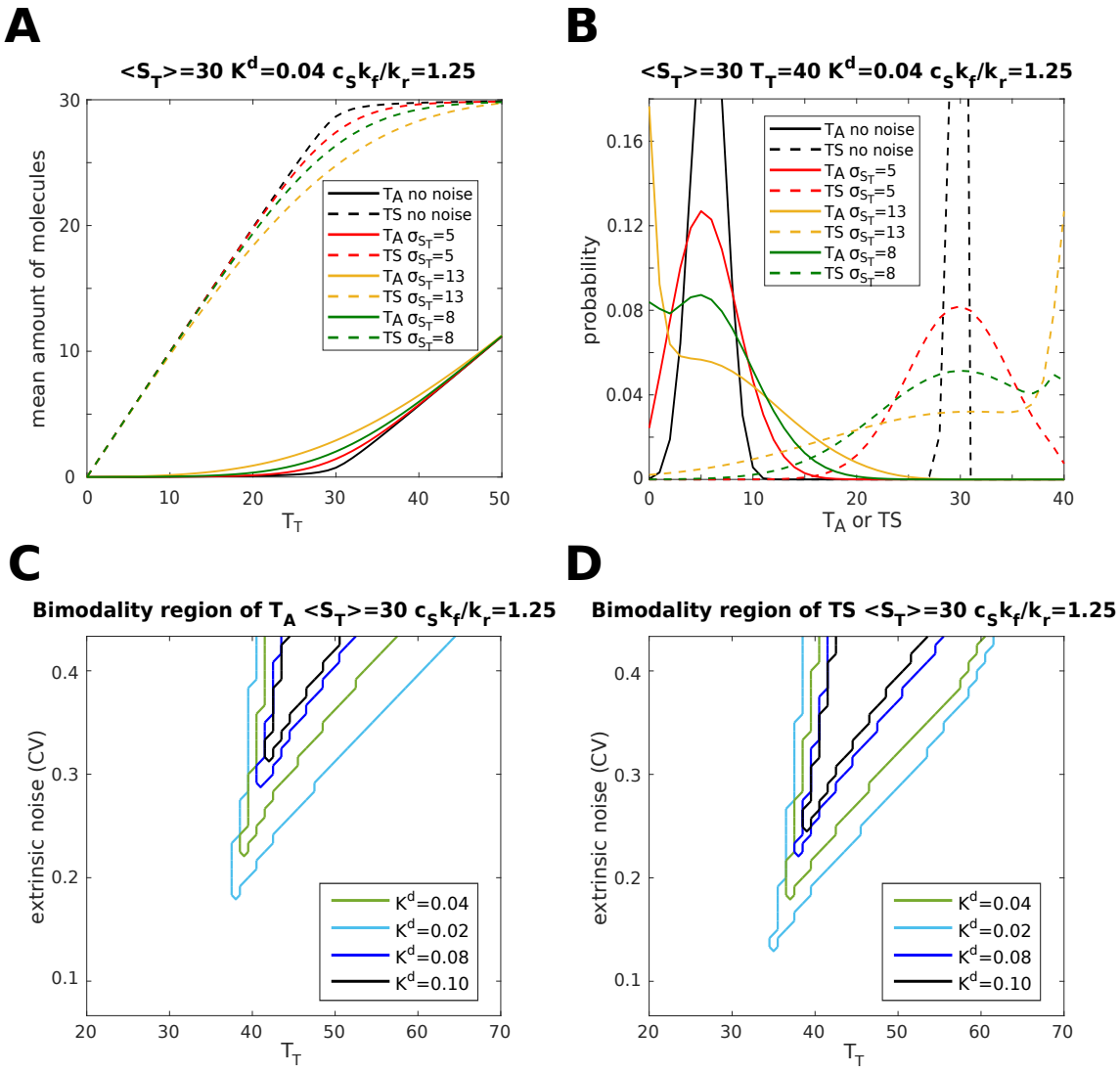

Fig. B.1 Extrinsic noise effects on competitive inhibition kinetics. (A) $\langle T_A \rangle$ (solid lines) and $\langle \overline{TS} \rangle$ (dashed lines) vs T_T for different levels of extrinsic noise. The black curve is the pure intrinsic noise case, while the other three correspond to a standard deviation of the Gaussian distribution of S_T equal to 5 (red), 8 (green) and 13 (yellow). (B) Examples of probability distribution of T_A (solid lines) and $\overline{TS}$ (dashed lines) in presence of extrinsic noise. (C,D) Plots of the bimodality region of the marginal distributions of T_A (c) and $\overline{TS}$ (d) for different values of K^d as a function of T_T and extrinsic noise. Bimodal distributions are present for parameters inside the areas delimited by the plotted lines. K^d assumes the values: 0.02 (light blue), 0.04 (green), 0.08 (blue) and 0.10 (black). The size of the step along T_T is $\Delta T_T = 1$, while the size of the step for the extrinsic noise level is $\Delta\sigma_T = 0.25$ ($\Delta CV = 8 \cdot 10^{-3}$). For each point defined by these steps, the distribution $P(T)$ was computed analytically and the number of its maxima was evaluated. Note that the step-like features in the plot are due to the discreteness of T_T. Adapted from [9]

target can also be bound by a molecule of sequester, S, forming the inactive complex $\overline{TS}$ with rate k_+ . The complex can in turn dissociate with rate k_- , returning a free molecule of target and one of sequester. The reaction network is then:

$$T_F + S \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \overline{TS} \tag{B.1}$$

$$\text{Substrate} + T_F \underset{k_r}{\overset{k_f c_s}{\rightleftharpoons}} T_A \rightarrow \text{Product} + T_F \tag{B.2}$$

We consider the case in which both substrate and product concentrations are large so that their fluctuations are negligible and focus on the stochastic dynamics of the enzyme (target) and the inhibitor (sequester). In addition to the reactions above, we assume that the total amounts of target and sequester molecules are conserved, defining then the following conservation laws:

$$T_T = T_F + T_A + \overline{TS} = const \quad , \tag{B.3}$$

$$S_T = S + \overline{TS} = const \quad . \tag{B.4}$$

As a consequence of the conservation laws, the number of free stochastic variables for this system is reduced to 2. Here, we focus on the active enzyme T_A from which the product is formed and the inactive complex $\overline{TS}$. For the sake of clarity we consider the case of quasi-equilibrium dynamics in which the product formation reaction is much slower than the others. The master equation describing the time evolution of the probability observing T_A and $\overline{TS}$ molecules is:

$$\begin{aligned}\frac{dP(T_A,\overline{TS},t)}{dt} &= k_f c_s(T_T - T_A + 1 - \overline{TS})P(T_A - 1,\overline{TS},t) + k_r(T_A+1)P(T_A+1,\overline{TS},t) + \\ &+ k_+(T_T - T_A - \overline{TS} + 1)(S_T - \overline{TS} + 1)P(T_A,\overline{TS} - 1,t) + k_-(\overline{TS}+1)P(T_A,\overline{TS}+1,t) + \\ &- [k_f c_s(T_T - T_A - \overline{TS}) + k_r T_A + k_+(T_T - T_A - \overline{TS})(S_T - \overline{TS}) + k_-\overline{TS}]P(T_A,\overline{TS},t) \quad .\end{aligned} \tag{B.5}$$

The analytic steady-state solution of the master equation can be written recalling the grand canonical distribution for ideal particle mixtures [121] and reads:

$$P(T_A, \overline{TS}) = \frac{1}{N}(K^d)^{-\overline{TS}} \left(\frac{k_f c_s}{k_r}\right)^{T_A} \frac{1}{\overline{TS}!(S_T - \overline{TS})!T_A!(T_T - T_A - \overline{TS})!} \quad , \quad \text{(B.6)}$$

where $K^d = k_-/k_+$ is the dissociation constant and

$$N = \sum_{\overline{TS}=0}^{min(T_T, S_T)} \sum_{T_A=0}^{T_T - \overline{TS}} \frac{1}{N}(K^d)^{-\overline{TS}} \left(\frac{k_f c_s}{k_r}\right)^{T_A} \frac{1}{\overline{TS}!(S_T - \overline{TS})!T_A!(T_T - T_A - \overline{TS})!} \quad , \quad \text{(B.7)}$$

is the normalization factor.

Eq. B.6 corresponds to the steady-state solution of the model for the case with pure intrinsic noise. However, also in this case we are interested in investigating the effects of some extrinsic noise on the sequester. To do that, we again assume that the total amount of sequester molecules is a fluctuating quantity described by a discretised Gaussian distribution $P(S_T)$. The analytic equilibrium solution in presence of extrinsic noise is obtained as a weighted superposition of conditional probabilities:

$$P(T_A, \overline{TS}) = \sum_{S_T=0}^{\infty} P(T_A, \overline{TS}|S_T)P(S_T) \quad , \quad \text{(B.8)}$$

where $P(T_A, \overline{TS}|S_T)$ are solutions of the master equation (B.5) with a given value of S_T.

Having access to the analytic equilibrium solution for this system, we can first study the average behaviour of T_A and $\overline{TS}$. In Fig. B.1A the means of T_A and $\overline{TS}$ are plotted as a function of T_T for different levels of extrinsic noise. Both quantities present the threshold-like profile typical of molecular sequestration. The mean of T_A displays a profile similar to the one of the minimal model, with a linear increase after the threshold, where T_T equals S_T. Differently, the mean of $\overline{TS}$ increases linearly until the threshold and then converges to a finite value which is set by the total amount of sequester molecules S_T. For both the quantities, an increase of the level of extrinsic noise leads to a smoothing of the threshold behaviour, exactly as observed in the minimal model. One of the main results of the analysis of the minimal model

is the possibility of observing bimodal distributions induced by extrinsic noise. As can be observed in Fig. B.1B, this property is conserved also in this model of enzymatic kinetic. Indeed, we find that both the inhibited complex (sequester-target) and the target bound to the substrate can display bimodal distributions. However, it is important to note that, due to the presence of an additional chemical reaction, with its inherent stochasticity, the probability distributions are smoother with respect to the minimal model. As a consequence, stronger affinities (lower dissociation constants) are required to ensure robust bimodal phenotypes, especially for the active enzyme. In analogy with the minimal model, the distributions can be tuned from being bimodal to unimodal and vice versa by modulating the level of extrinsic noise. Finally, the effects of the extrinsic noise level and on the dissociation constant ($K_d = k_-/k_+$) on bimodality are in qualitative agreement with the minimal model. Indeed, for both T_A and $\overline{TS}$, as extrinsic noise is increased, the range of bimodality over the values of T_T becomes wider. Similarly, a low value of the dissociation constant, i.e. a steeper threshold, favours the presence of bimodality (Fig. B.1C-D). Altogether, these results confirm the validity of the predictions based on the analysis of the minimal model and show that extrinsic-noise-induced bimodal distributions can play a role also in these enzymatic reactions governed by competitive inhibition.

B.2 Correlations

In previous chapters, by extending the minimal model with the addition of a second target of the same sequester, we showed how extrinsic noise can induce positive correlation between the two targets, contrasting the negative one induced by competition. In order to verify this prediction in the model of competitive inhibition, we add a second enzyme that can be activated by binding to a substrate and that is sequestered by the same inhibitor of the previous one. For simplicity, we assume that the two enzymes bind to the same substrate. The reaction network for this extended model is:

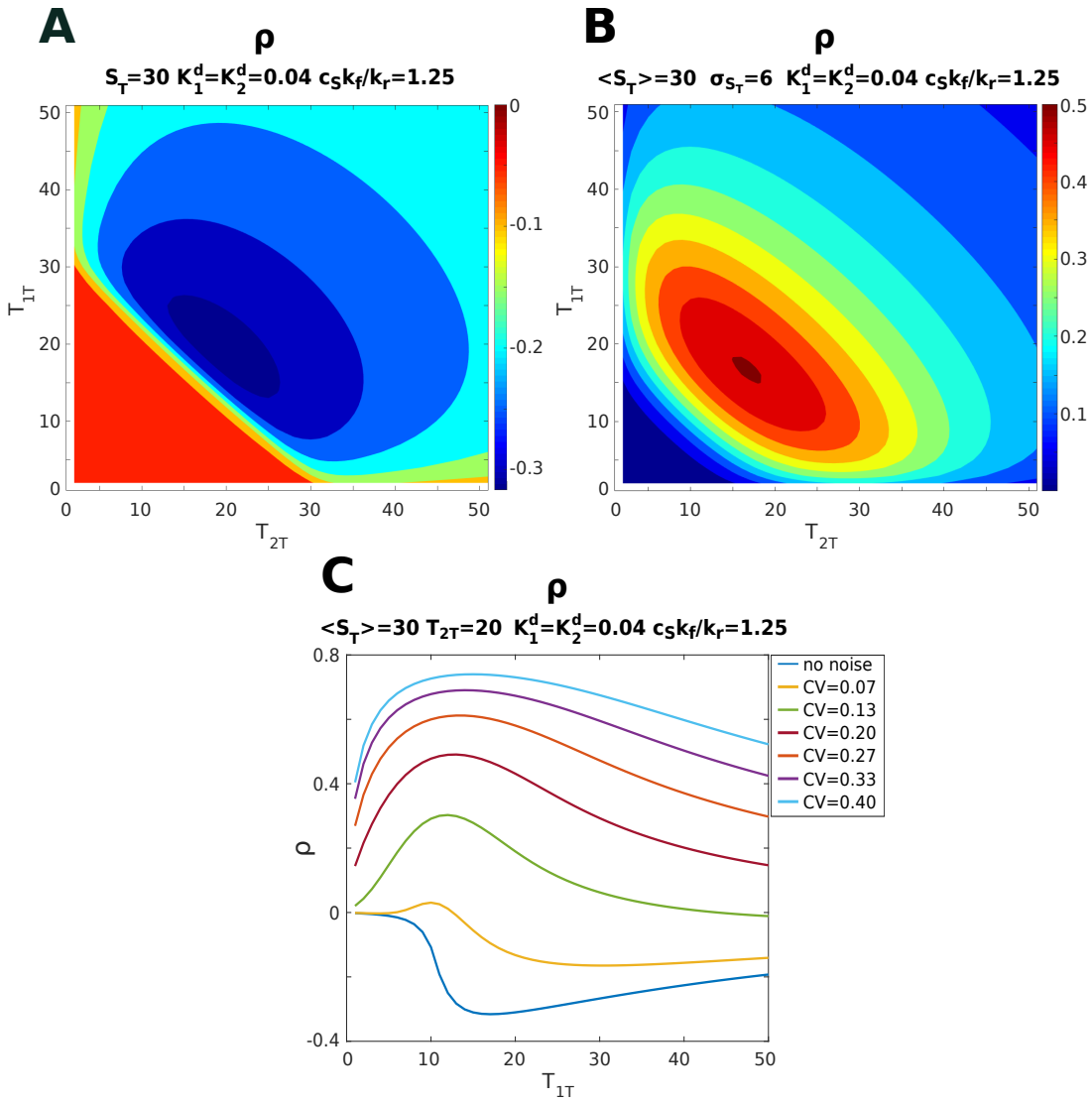

Fig. B.2 Correlations for competitive inhibition. (A) Pure intrinsic noise case: contour plot of the Pearson correlation between the active enzymes T_{1A} and T_{2A} as a function of the total number of enzymes T_{1T} and T_{2T}. (B) Extrinsic noise: contour plot of the Pearson correlation between T_{1A} and T_{2A} as a function of T_{1T} and T_{2T}. (C) Correlation between the active targets as a function of T_{1T} for different levels of extrinsic noise. The blue line on the bottom corresponds to the pure intrinsic noise case, for the other lines σ_{S_T} assumes the values: 2, 4, 6, 8, 10, 12. Adapted from [9].

$$T_{1F} + S \underset{k_-^1}{\overset{k_+^1}{\rightleftharpoons}} \overline{T_1 S} \quad ,$$

$$\text{Substrate} + T_{1F} \underset{k_r^1}{\overset{k_f^1 c_s}{\rightleftharpoons}} T_{1A} \rightarrow \text{Product} + T_{1F} \quad ,$$

$$T_{2F} + S \underset{k_-^2}{\overset{k_+^2}{\rightleftharpoons}} \overline{T_2 S} \quad ,$$

$$\text{Substrate} + T_{2F} \underset{k_r^2}{\overset{k_f^2 c_s}{\rightleftharpoons}} T_{2A} \rightarrow \text{Product} + T_{2F} \quad .$$

As in the other models, we assume that the total amounts of sequester and targets molecules are conserved, obtaining the following conservation laws:

$$T_{1T} = T_{1F} + T_{1A} + \overline{T_1 S} = const \quad , \tag{B.9}$$

$$T_{2T} = T_{2F} + T_{2A} + \overline{T_2 S} = const \quad , \tag{B.10}$$

$$S_T = S + \overline{T_1 S} + \overline{T_2 S} = const \quad . \tag{B.11}$$

These conservation laws reduce the number of free variables to 4: from now on we will focus on active targets and complexes. To use a lighter notation, we define the probability of observing $\overline{n} = (n_{T_{1A}}, n_{T_{2A}}, n_{\overline{T_1 S}}, n_{\overline{T_2 S}}) \equiv (T_{1A}, T_{2A}, \overline{T_1 S}, \overline{T_2 S})$ molecules at time t as: $P(\overline{n}, t) \equiv P(T_{1A}, T_{2A}, \overline{T_1 S}, \overline{T_2 S}, t)$. The master equation describing the time evolution of this probability distribution reads:

$$\begin{aligned}\frac{dP(\overline{n},t)}{dt} &= k_f^1 c_s(T_{1T}-T_{1A}+1-\overline{T_1S})P(\overline{n}_{T_{1A}}-1,t)+k_r^1(T_{1A}+1)P(\overline{n}_{T_{1A}}+1,t)+\\ &+k_+^1(T_{1T}-T_{1A}-\overline{T_1S}+1)(S_T-\overline{T_1S}-\overline{T_2S}+1)P(n_{\overline{T_1S}}-1,t)+k_-^1(\overline{T_1S}+1)P(n_{\overline{T_1S}}+1,t)+\\ &-[k_f^1 c_s(T_{1T}-T_{1A}-\overline{T_1S})+k_r^1 T_{1A}+k_+^1(T_{1T}-T_{1A}-\overline{T_1S})(S_T-\overline{T_1S}-\overline{T_2S})+k_-^1\overline{T_1S}]P(\overline{n},t)+\\ &+k_f^2 c_s(T_{2T}-T_{2A}+1-\overline{T_2S})P(\overline{n}_{T_{2A}}-1,t)+k_r^2(T_{2A}+1)P(\overline{n}_{T_{2A}}+1,t)+\\ &+k_+^2(T_{2T}-T_{2A}-\overline{T_2S}+1)(S_T-\overline{T_1S}-\overline{T_2S}+1)P(n_{\overline{T_2S}}-1,t)+k_-^2(\overline{T_2S}+1)P(n_{\overline{T_2S}}+1,t)+\\ &-[k_f^2 c_s(T_{2T}-T_{2A}-\overline{T_2S})+k_r^2 T_{2A}+k_+^2(T_{2T}-T_{2A}-\overline{T_2S})(S_T-\overline{T_1S}-\overline{T_2S})+k_-^2\overline{T_2S}]P(\overline{n},t)\quad .\end{aligned} \tag{B.12}$$

The analytic steady-state solution of the master equation above is [121]:

$$P(T_{1A},T_{2A},\overline{T_1S},\overline{T_2S}) = \frac{1}{N}\frac{\left(\frac{k_f^1 c_s}{k_r^1}\right)^{T_{1A}}\left(\frac{k_f^2 c_s}{k_r^2}\right)^{T_{2A}}(K_1^d)^{-\overline{T_1S}}(K_2^d)^{-\overline{T_2S}}}{\overline{T_1S}!\overline{T_2S}!(S_T-\overline{T_1S}-\overline{T_2S})!T_{1A}!(T_{1T}-T_{1A}-\overline{T_1S})!T_{2A}!(T_{2T}-T_{2A}-\overline{T_2S})!}\quad , \tag{B.13}$$

with $K_1^d = k_-^1/k_+^1$, $K_2^d = k_-^2/k_+^2$ and

$$N = \sum_{\overline{T_1S}=0}^{min(T_{1T},S_T)}\sum_{\overline{T_2S}=0}^{min(T_{2T},S_T-\overline{T_1S})}\sum_{T_{1A}=0}^{T_{1T}-\overline{T_1S}}\sum_{T_{2A}=0}^{T_{2T}-\overline{T_2S}}\left(\frac{k_f^1 c_s}{k_r^1}\right)^{T_{1A}}\left(\frac{k_f^2 c_s}{k_r^2}\right)^{T_{2A}}\cdot \tag{B.14}$$

$$\cdot\frac{(K_1^d)^{-\overline{T_1S}}(K_2^d)^{-\overline{T_2S}}}{\overline{T_1S}!\overline{T_2S}!(S_T-\overline{T_1S}-\overline{T_2S})!T_{1A}!(T_{1T}-T_{1A}-\overline{T_1S})!T_{2A}!(T_{2T}-T_{2A}-\overline{T_2S})!}\quad . \tag{B.15}$$

Finally, given the analytic solution of the master equation for the case with pure intrinsic noise, the full probability distribution in presence of extrinsic noise can be obtained in the usual way as a weighted superposition of conditional probabilities:

$$P(T_{1A},T_{2A},\overline{T_1S},\overline{T_2S}) = \sum_{S_T=0}^{\infty}P(T_{1A},T_{2A},\overline{T_1S},\overline{T_2S}|S_T)P(S_T)\quad . \tag{B.16}$$

Given these analytic solutions, we can first focus on the correlations that are induced on the active enzymes by the competition for binding to the inhibitor. In the absence of extrinsic noise, the two enzymes result to be negatively correlated, especially in proximity to the threshold (see Fig. B.2A). The activation reaction introduces an additional source of stochasticity. As a consequence, the correlations between the active enzymes (T_{1a} and T_{2a}) are generally lower than the ones between the sequestered, inactive enzymes ($\overline{T_1 S}$ and $\overline{T_2 S}$) and of the ones of the minimal model. This additional layer of stochasticity also affects the profile of the Pearson correlation coefficient, which now displays a minimum around the threshold instead of the plateau shown for the minimal model. This can be intuitively understood considering the fact that, in the minimal model, the plateau arose because the standard deviation of the targets roughly followed the behaviour of the covariance. The additional stochasticity related to the activation dynamics is not affected much by the sequestration dynamics, thus the standard deviation keeps growing with the total number of targets not compensating the slow-down of the covariance, giving rise to the minimum.

We can now focus on the effect on correlation of extrinsic fluctuations of the number of sequester molecules. As in the minimal model, such fluctuations induce positive correlations between the active target, especially around the threshold (see Fig. B.2B,C) and the considerations made for the minimal model hold also in this case. Again, since the interaction with the sequester is diluted by the presence of an additional stochastic reaction, the correlation tends to be generally lower.