The role of microRNA-mediated circuits in the control of gene expression fluctuations.

Matteo Osella

Department of Theoretical Physics
University of Torino

A thesis submitted for the degree of
Doctor of Philosophy

Doctoral school in Complex Systems in Medicine and Life Sciences

Ph.D programme in complexity in post-genomic biology

programme coordinator: Prof. Michele Caselle
XXIII Cicle - 2008/2010 - SSD: FIS/02

Advisor
Prof. Michele Caselle

1. Reviewer: Prof. Mario Nicodemi
Department of Physics and Complexity Science, University of Warwick
CV4 7AL United Kingdom Coventry, United Kingdom

Dipartimento di Scienze Fisiche, Universita' di Napoli *Federico II*
80126 Napoli, Italy.

2. Reviewer: Marco Cosentino Lagomarsino, PhD
Genophysique / Genomic Physics Group, CNRS Genomique des Microorganismes
Universite Pierre et Marie Curie, Centre de recherche des Cordeliers
75006 Paris, France

This thesis is partially based on the following papers:

Osella M, Bosia C, Cora' D and Caselle M
The role of incoherent microRNA-mediated feedforward loops in noise buffering (submitted)
preprint: arXiv:1004.0336v1

El Baroudi M, Cora' D, Bosia C, Osella M and M. Caselle
A curated database of miRNA-mediated feedforward loops involving MYC as master regulator (submitted)

Osella M, Bosia C, El Baroudi M, Cora' D and Caselle M
Gene autoregulation via intronic microRNAs and its functions (in preparation)

Osella M and Caselle M
Entropic contributions to the splicing process
Phys. Biol. 6 (2009) 046018

“E senti allora,
se pure ti ripetono che puoi
fermarti a mezza via o in alto mare,
che non c’è sosta per noi,
ma strada, ancora strada,
e che il cammino è sempre da ricominciare. ”
Eugenio Montale

Contents

List of Figures

LIST OF FIGURES

1

Introduction

1.1 Outline

Many complex biological behaviours are the result of the dynamics of underlying cellular networks. However, a deep understanding of the design principles of natural networks and of the functions they can perform is a challenging task, given their complexity and the small fraction of existing networks that are well characterized. Nonetheless, growing evidence suggests the existence of core modules with peculiar regulative tasks that represent simple building-blocks of biological networks. For example, it has been shown that there are recurrent wiring patterns (motifs), linked to specific functions as temporal expression programs, reliable cell decisions or tunable oscillations (see (1) for a review). These findings suggest an intriguing working hypothesis: mapping and dissecting the functions that can be robustly executed by simple regulative topologies would in principle allow a progressive scale-up to the whole network functioning. Following this branch of research, this thesis is focused on the functional properties of two simple regulative motifs involving microRNA regulation: microRNA-mediated feedforward loops (*Chapter 2*) and autoregulation via intronic microRNAs (*Chapter 3*). A special attention will be dedicated to the circuit functioning in the presence of the molecular noise which inherently affects gene expression.
Using analytical stochastic models and Monte Carlo simulations we will show how these two circuits can perform biologically important functions while dealing with noisy signals. In particular, microRNA-mediated incoherent feed-forward loops can fine tune the level of target expression while keeping the desired steady state robust with respect to fluctuations propagating from the upstream network, thus conferring precision and stability to the overall gene expression program. As a nontrivial prediction our analysis points out that the optimal noise-buffering efficiency coincides with a modest repression of the microRNA target expression, in agreement with experimental observations of the actual impact of a wide class of microRNAs on the protein output of their targets (*Chapter 2*).
With analogous modeling strategies we will address the possible functions of an

autoregulatory circuit that exploit an intronic microRNA to negatively regulate the expression of an host gene. We will prove that this circuit, despite its simple topology, can speed up the response time of the host gene to external stimuli, control fluctuations in steady-state gene expression and implement fold change detection (sensitivity on relative changes of input signals and not on their absolute values). These different functions concur to make the expression of the host gene robust to noise in different conditions. Furthermore, a deep exploration of the parameter space will depict a detailed map of the circuit behaviours, unraveling the potential role of endogenous miRNA-mediated self loops and suggesting possible applications in the growing field of synthetic biology (*Chapter 3*)
Finally, during my PhD I also worked on a different and unrelated topic: the splicing process. The results of this work will be presented in the Appendix. More specifically, it will be shown that the depletion attraction, resulting from molecular crowding, can play an important role in early steps of spliceosome assembly on a messenger RNA that have to be processed. Using modeling tools from soft-matter physics, a mathematical representation of the essential features of the mRNAs as polymers and of the splicing machinery assembly in a crowded environment will be introduced. Despite its simplicity, the model can explain qualitatively and quantitatively some general properties of the splicing mechanism such as the importance of the intron length in the choice of the splice-site recognition modality. On top of that, some qualitative features of the genome architecture of higher eukaryotes can find an evolutionary realistic motivation in the light of the proposed model (*Appendix A*).
The rest of this introductory chapter will be devoted to present some necessary background in stochastic gene expression and in microRNA regulation (*Chapter 1*).

1.2 Stochastic gene expression

A certain degree of randomness pervades the biochemical processes that make up life at the cellular level: any phenotypic characteristics measured from a population of cells or from a single cell at different times will not have a unique value, but a collection of values. In particular, gene expression in both prokaryotes and eukaryotes is inherently a stochastic process as chemical reactions are probabilistic and many genes, RNAs and proteins are present in low numbers per cell (2; 3). A long succession of probabilistic events is involved in gene expression, all contributing to the final level of stochasticity: conformational changes of the DNA chain, random binding and unbinding of transcription factors, and the intricate interplay between the molecular complexes that promote initiation, elongation and termination. Therefore, even in an homogeneous environment, this complexity produces the random fluctuations in protein and mRNA levels that have been observed between individual cells in an isogenic population or looking at a single cell in time (4; 5; 6; 7). The experimental approaches to study noise properties of gene expression commonly start with the insertion of a reporter gene (e.g. green fluorescent protein driven by a promoter of interest) into the genome and subsequently the fluorescence of individual cells in a population is measured with microscopy or flow cytometry (7; 8; 9; 10) (see Figure 1). Alternatively single

cells can be followed over time, yielding important information on the dynamics of stochastic gene expression (10; 11; 12).

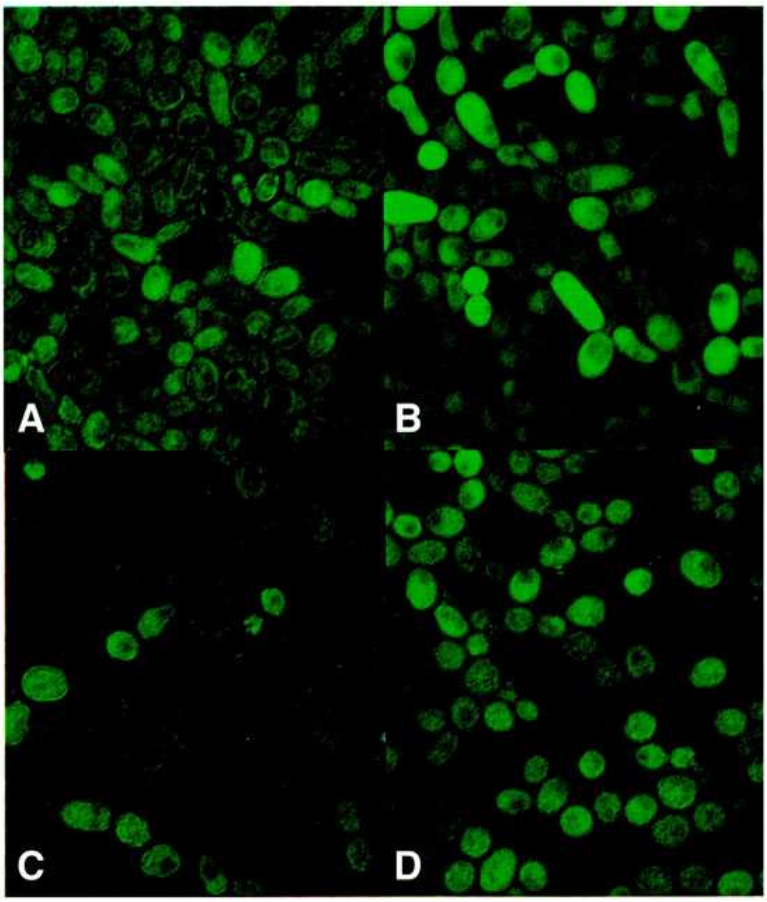

Figure 1.1: Cell-to-cell variability - (A-C) Four different snapshot of an isogenic population of eukaryotic cells obtained with fluorescence microscope imaging in different conditions. The variability in the expression of the fluorescent protein is clearly evident. (Adapted from reference (16))

1.2.1 Biological significance of stochasticity

The stochasticity in gene expression constitutes a "noise" that the cells have to live with. In many cases this stochasticity can be detrimental to cell physiology,

as fluctuations in protein levels can corrupt the quality of intracellular signals, making the proper cellular functioning dependent on noise counteracting mechanisms. For example, the architecture and the topology of the wiring patterns in the network of gene regulation play an important role in the control of noise propagation and can thus be crucial for the stability and reliability of the expression program. On the other hand, a possible benefit of stochasticity is to provide a mechanism for phenotypic and cell-type diversification. This is expected to be particularly beneficial to microbial cells that need to adapt fastly and efficiently to sudden environmental changes. Stochastic gene expression naturally implements a sampling of different phenotypes (bet-hedging) that can increase the probability of survival of the population in adverse conditions, without the need of genetic mutations (13; 14). Moreover, recent work suggests that a noise-triggered heterogeneity can be at the basis of many stages of cellular differentiation in eukaryotes (see (3; 15) for a review). In conclusion, two distinct roles can be imagined for noise in cellular functioning: one is a nuisance that serves as impediment to reliable and robust behaviour, and one is a source of variability that the cell can exploit.

1.2.2 Intrinsic versus extrinsic noise

Considering a single gene of interest, the fluctuations in the protein amount it produces are originated in two ways. First, even if cells are in precisely the same state, the biochemical events leading to transcription and translation still occur at different random time and in a different order in different cells, determining the level of cell-to-cell heterogeneity. Such stochastic effects are only due to the inherent stochasticity of a single gene expression, determined locally by the peculiar properties of the gene in analysis, and will be referred to as "intrinsic noise" (17). In addition, a single gene cannot be considered as an isolated system, therefore the stochasticity of its expression is affected by the interactions with other stochastic systems in the cell or its environment. These additional sources of external noise will cause fluctuations in the expression of the gene of interest that are commonly classified as "extrinsic noise" (5; 17). The sources of extrinsic noise are not yet fully characterized. However, several contributions have been identified. As a first example, genes are regulated by other gene products, like transcription factors or small non-coding RNAs that can equally fluctuate substantially in their copy number, exerting an influence on the degree of fluctuations in the protein product of the gene of interest through noise propagation (19; 20). Similarly, RNA polymerases and ribosomes are themselves gene products and their copy number will vary over time and among cells. Moreover, cell cycle effects and noisy growth rate can also contribute to the population heterogeneity (10; 20; 21). Extrinsic fluctuations seem the dominant source of cellular variation in both prokaryotes (5) and eukaryotes (7). At the "phenomenological" level, extrinsic fluctuations can have a lifetime comparable to the cell cycle (10; 18), are non specific, potentially affecting equally the expression of many genes (19), and they can alter the simple proportional scaling of intrinsic noise with the copy number of molecules (18).
The relative contribution of intrinsic over extrinsic noise can be experimentally

estimated with two-reporter assays (2; 5). Intrinsic noise can be operationally defined as the difference in the expression of two identical reporters, such as cyan and yellow fluorescent proteins, expressed from identical promoters located in equivalent chromosome positions. In fact, two identical genes will be equally affected by fluctuations in cell-specific factors (extrinsic noise) resulting in a perfectly correlated expression. However, the fluctuations generated by the biochemical reaction steps that are intrinsic to the process of gene expression are specific for each gene dynamics, generating the uncorrelated component of fluctuations (see Figure 1.2).

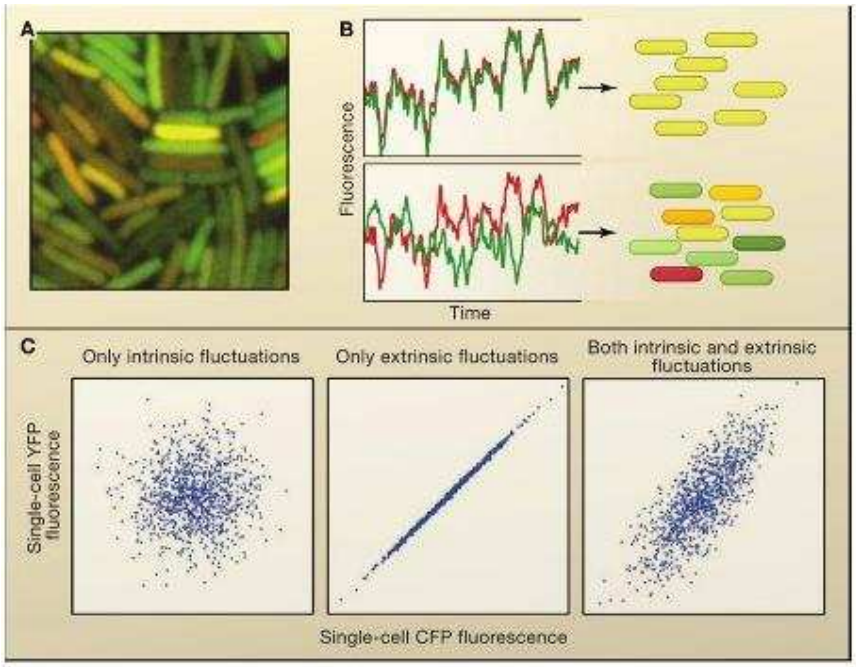

Figure 1.2: Intrinsic versus extrinsic noise - (A) Fluorescence image from a two-reporter experiment in *E. coli*.(B) Temporal behaviour of extrinsic fluctuations (upper) and intrinsic fluctuations (lower).(C) Cell-to-cell variability when noise is intrinsic, extrinsic or both. (Adapted from reference (3))

1.2.3 Noise in regulatory networks

Among many possible molecular origins of extrinsic fluctuations, in eukaryotes one of the major sources is noise propagation from upstream regulating factors

(20). In fact, genes and proteins are organized in an extensive complex network of regulations that allows signal propagation from one gene to the next. This process may be hindered by stochastic fluctuations arising from gene expression and propagating through the network. However, the network architecture may enable cells to deal with (or take advantage of) noise-driven fluctuations. Following this idea, several theoretical and experimental papers have focused on the relation between the topology of simple networks and noise susceptibility. For example, negative feedbacks have been shown to provide a noise-reduction mechanism (18; 22; 23; 24). In particular, they are effective in minimizing the effects of fluctuations on downstream processes and shift the fluctuation frequency to higher values (25), although at the expense of a reduced signal sensitivity (26). On the other hand, positive feedbacks generally amplify fluctuations and cell-to-cell variability (16; 27). Positive feedbacks are also suitable to implement bistability and the stochasticity in gene expression can cause random transitions between stable states, leading to bimodality in the population (28; 29).
In conclusion, the emerging idea is that circuit architectures can encode distinct noise properties that can be critical to the physiological process they implement. Even noise-related properties are themselves an evolvable trait (30), eventually driving the positive evolutionary selection of specific circuit topologies (31).

1.2.4 Models of gene expression

The simplest model of gene expression, based on the Central Dogma of molecular biology, can be summarized in terms of the following chemical reactions:

$$\begin{aligned} D &\xrightarrow{k_r} D + r \\ r &\xrightarrow{g_r} \emptyset \\ r &\xrightarrow{k_p} r + p \\ p &\xrightarrow{g_p} \emptyset, \end{aligned} \tag{1.1}$$

where r is the number of messenger RNAs (mRNAs) transcribed from the gene D (which is assumed not to undergo any dynamics) with rate k_r; p is the number of proteins, translated with rate k_p. Protein and mRNA degradation events are modeled as elementary reactions with rates g_p and g_r.
It is straigthforward to write down the corresponding rate-kinetic equations, describing the mRNA and protein dynamics:

$$\begin{aligned} \frac{dr(t)}{dt} &= k_r - g_r r(t) \\ \frac{dp(t)}{dt} &= k_p r(t) - g_p p(t). \end{aligned} \tag{1.2}$$

The equations above can be easily solved to obtain the dynamics of $r(t)$ and $p(t)$ (with the implicit assumption that they are continuous variables) for the initial conditions $r(0) = 0$ and $p(0) = 0$:

$$\begin{aligned} r(t) &= r_{ss}(1 - e^{-g_r t}) \\ p(t) &= p_{ss}\left(\frac{g_p(1 - e^{-g_r t}) + g_r(1 - e^{-g_p t})}{g_p - g_r}\right), \end{aligned} \tag{1.3}$$

where $r_{ss} = k_r/g_r$ and $p_{ss} = k_p k_r/g_r g_p$ are the steady state values of the number of mRNAs and proteins. However, this solution does not provide a satisfactory description of the gene expression dynamics when fluctuation effects are non-negligible. As discussed in the previous sections, biochemical reactions are probabilistic and the number of key molecules involved can be very small. As a consequence, in real cells the mean of a protein level may often be well comparable with its variance. Thus, the statistics of large numbers is typically not applicable to mRNAs and proteins in the context of gene expression and the above deterministic equations have to be considered as a mean field description of the real stochastic process.
The chemical reactions in Equation 1.1 can be represented more realistically with the Master Equation approach, fully describing the inherent stochasticity of reactions and the discrete nature of the molecular species involved. The state of the system is assumed to be specified at any time t by the total number of mRNAs r and proteins p, and the joint probability distribution that the system is in a given state $\{r, p\}$ at time t can be defined as $P(r, p; t)$. This distribution evolves according to the Master Equation:

$$\begin{aligned} \partial_t P(r,p;t) &= k_r(P(r-1,p;t) - P(r,p;t)) + \\ & k_p r(P(r,p-1;t) - P(r,p;t)) + g_r\Big[(r+1)P(r+1,p;t) - rP(r,p;t)\Big] \\ & + g_p\Big[(p+1)P(r,p+1;t) - pP(r,p;t)\Big]. \end{aligned} \tag{1.4}$$

In this framework, a reaction rate gives actually the probability per unit time that the reaction may occur. We firstly focus on steady state solutions ($\partial_t F = 0$) and in particular on the first two moments of the distribution, i.e. mean values ($< r >$ and $< p >$) and standard deviations (σ_r and σ_p). These are the system properties most clear to interpret and accessible experimentally. However, a quantification of fluctuations that allows an unbiased comparison of different systems requires a precise definition of noise. There is not a unique definition of the noise strenght. Typically, the relative amplitude of fluctuations of the molecular species x_i is measured with the coefficient of variation $CV_{x_i} = \sigma_{x_i}/ < x_i >$ which has a clear experimental interpretation and it will be the measure adopted in the following chapters of this thesis. Another common measure is the Fano factor $\sigma^2_{x_i}/ < x_i >$ which equals one for Poisson distributions. While the Fano factor can be useful in the context of univariate discrete random processes to distinguish between different sources of noise (2), the comparison with a Poisson

process can often be misleading, forcing an unnatural dependence of the noise measure on the average (33).
The Equation 1.4 can be solved using the moment generating function approach (23; 32). The generating function can be defined as:

$$F(z_1, z_2) = \sum_{r,p} z_1^r \; z_2^p \; P(r, p; t). \tag{1.5}$$

This function has the following properties: $F|_1 = 1$; $\partial_{z_i} F =< x_i >$; $\partial^2_{z_i} F =< x_i^2 > - < x_i >$; where $|_1$ means evaluation of F at $x_i = 1$ for each molecular species $\{r, p\}$. Equation 1.4 can now be converted into a first-order partial differential equation:

$$\begin{aligned} \partial_t F &= k_r(z_1 F - F) + k_p z_1(z_2 \partial_{z_1} F - \partial_{z_1} F) \\ &\quad + g_r(\partial_{z_1} F - z_1 \partial_{z_1} F) + g_p(\partial_{z_2} F - z_2 \partial_{z_2} F). \end{aligned} \tag{1.6}$$

The first two moments of the probability distribution can be easily obtained with successive differentiations of Equation 1.6 (23). In particular, the number of mRNA molecules follows a Poisson distribution with $< r >= k_r/g_r$ and $\sigma_r^2/ < r >= 1$, as mRNAs are generated by a birth-death process. Since the production of proteins is dependent on the mRNA concentration, their distribution at the steady state is broader than Poissonian:

$$\begin{aligned} < p > &= \frac{k_r b}{g p} \\ \frac{\sigma_p^2}{< p >} &= (\frac{b}{1+\eta}) + 1, \end{aligned} \tag{1.7}$$

where $\eta = g_p/g_r$ is the ratio of mRNA to protein lifetimes and $b = k_p/g_r$ is the average number of proteins produced per transcript. The emerging picture is that mRNAs are transcribed stochastically, then each transcript, in the time between its birth and death, gives rise to a protein production burst of random size (with average b). This "burstiness" leads to fluctuations of the protein number larger than Poissonian (23; 33; 34). According to this mechanism, for two genes expressed at the same average abundance, the one with the higher translational efficiency (k_p) and lower mRNA abundance is expected to show greater fluctuations in protein concentration and a broader population distribution than the gene with a higher mRNA concentration but a lower rate of translation (see Figure 1.3).
In the situation of mRNA lifetime much shorter than protein one, a complete time dependent solution can be achieved using the method of characteristics (35).
However, the analytic solution of the Master Equation describing more complicated models, for example involving gene regulations, can be unfeasible, even at the steady state. In many cases even the extraction of the cumulants (mean, variance) of the distribution may be precluded. For this reason several approximation techniques, as Linear Noise Approximation or Langevin approaches, have been

applied to models of stochastic gene expression (see (27; 33; 37) for a review).
So far we have considered only the effect of intrinsic fluctuations. In the contest of mathematical models, extrinsic noise can be defined as fluctuations in the rate constants associated with the various reactions (2). For example, fluctuations in the number of available ribosomes make the rate of translation k_p a random variable. Similarly, the rate of transcription fluctuates as the number of free polymerases or transcription factors changes. Extrinsic fluctuations are difficult to model, their effects hard to predict and their interplay with intrinsic noise may be non trivial (17; 18; 38). Since the real nature of extrinsic noise is so far elusive, it has often been introduced in gene expression models assuming an arbitrary distribution of extrinsic fluctuations or an arbitrary stochastic process that generates their distribution. In this thesis, extrinsic fluctuations will be assumed to be mainly generated by the propagation of the intrinsic noise of regulators (transcription factors or non-coding RNAs).
An interesting development of the simple model of gene expression proposed above is represented by the inclusion of the promoter dynamics. Both in bacteria and in eukaryotes, bursts of transcription have been observed, in which the gene itself randomly moves between states of transcriptional activity and inactivity (39; 40; 41). In eukaryotes, transcriptional bursts seems to be one of the major causes of the population variabilty and chromatin remodelling is the proposed responsible for promoter transitions (see (3) and references therein). In this thesis, possible promoter transitions due to changes in chromosome organization will be neglected, excluding a binary on/off promoter dynamics. Nonetheless, the action of transcription factors on the level of promoter activity will be considered.

1.2.5 The Gillespie algorithm

When the analytical solution of the Master Equation cannot be obtained or when the analytical approximations used must be tested (as it will be the case in this thesis), the Gillespie Algorithm can be used to run simulations. The Gillespie Algorithm belongs to the class of Monte Carlo methods and it allows the generation of trajectories (succession of biochemical events) that exactly represent solutions of the Master Equation. In this sense the algorithm is exact: while the Master Equation describes the probability distribution of each feasible trajectory, each Gillespie simulation provides a single trajectory compatible with the Master Equation. Performing a sufficient number of simulations we can have a significant sample of the probability distribution.
Gillespie developed two alternative (although equivalent) formulations: the Direct Method and the First Reaction Method. We will focus on the Next Reaction Method since it is the method that has been used to obtain the results presented in the following chapters. While the detailed mathematical description can be found in reference (42), the steps of the algorithm can be summarized as:

- Initialization: initialize the number of molecules for each species $x_1, ...x_n$, the reaction rates $k_1, ...k_m$ and the generator of random numbers uniformly distribuited in the interval $[0, 1]$;

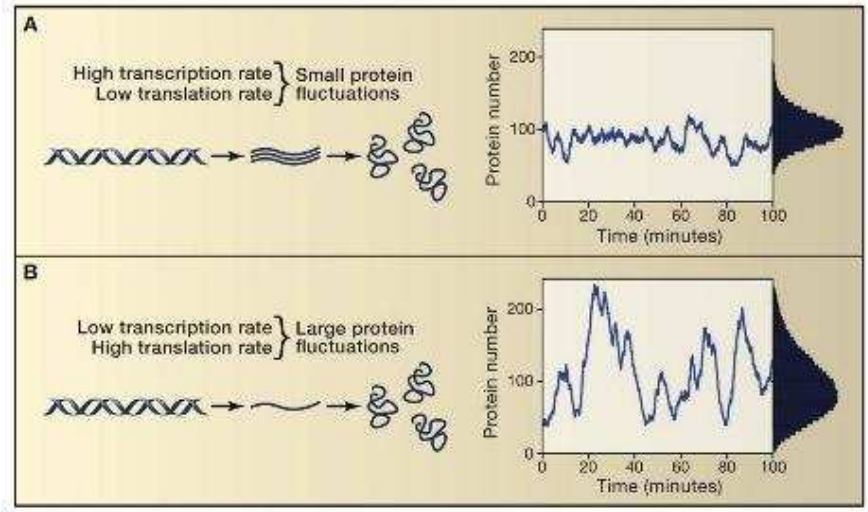

Figure 1.3: Burstiness in gene expression - The same mean value of proteins obtained with high transcription and low translation rate (A) or low transcription and high translation rate (B). In the latter case gene expression is far noisier. As Equation 1.7 shows, the noise level crucially depends on the amplitude of translation bursts which is larger in the B case, where from each transcript more proteins are produced in average. (Adapted from reference (3)).

- For each reaction μ the propensity function $a_\mu = k_\mu h_\mu$ must be calculated. k_μ is the reaction rate while h_μ is the product of the number of reactants;
- m random numbers j_μ are generated and for each reaction the putative time τ_μ is calculated following the exponential distribution $\tau_\mu = \frac{1}{a_\mu} ln(\frac{1}{j_\mu})$. τ_μ represents the putative amount it will take for the reaction μ to occur given the actual state of the system;
- The reaction corresponding to the smallest putative time τ is selected to occur: $\tau = min_\mu(\tau_\mu)$;
- The state of the system is updated according to the stoichiometry of the reaction that have been selected and the current time is updated $t \rightarrow t + \tau$.

1.3 MicroRNA regulation

MicroRNAs (miRNAs) are endogenous small non-coding RNAs which negatively regulate the protein production of their targets in metazoans and plants. They are expected to target a substantial portion of the human genome (43) and have been shown to play key roles in several biological processes ranging from development and metabolism to apoptosis and signaling pathways (44; 45; 46; 47; 48). Moreover, their profiles are altered in several human diseases (49; 50), making miRNAs a major focus of research in nowadays molecular biology.

1.3.1 MicroRNA biogenesis and action

MicroRNA genes can be found in genomes as distinct transcriptional units as well as in clusters of polycistronic units that include several microRNAs (45). Their expression is regulated by the same molecular mechanisms that control the protein-coding gene expression. About 37% of the known mammalian miRNAs are located within the introns of protein coding genes (so- called host genes) but this percentage is strikingly larger in human (53%) (51). The majority of intronic microRNAs is transcriptionally linked to their host gene expression (52; 53) and processed from the same primary transcript, leading to correlated expression with the host. On the other hand, intergenic microRNAs present their own independent promoter.
The process of miRNA biogenesis in vertebrates involves different steps (Figure 1.4). DNA polymerase II transcribes miRNA genes, generating long primary transcripts (pri-miRNAs) that are usually several kilobases long and that contain a local hairpin structure (54). In the nucleus, the RNase IIItype enzyme Drosha processes the pri-miRNA, producing a hairpin precursor (pre-miRNA) consisting of approximately 70 nucleotides (nt). Following nuclear processing by Drosha, the pre-miRNA is exported to the cytoplasm by the nuclear export factor Exportin-5. Once there, it is subjected to the second processing step by Dicer (another RNase III enzyme) to generate the final product which is approximately 22 nt long (mature miRNA). Lastly, the mature miRNA is incorporated into a ribonuclear particle, forming the RNA-induced gene silencing complex (RISC) that can downregulate the expression of target genes. The target recognition is based on the binding of the miRNA to sites present in 3'-untranslated region (3'UTR) of the target mRNA. The binding usually involves a sequence of 6 nt of the miRNA (miRNA seed). In the case of intronic miRNAs, it is believed that splicing precedes pri-miRNA processing and that the resulting intron lariat gets processed to release the pre-miRNA that can subsequently follow the processing steps described above.
MicroRNAs can direct the RISC to downregulate gene expression by either of two post-transcriptional mechanisms: promoting mRNA degradation or repressing mRNA translation (47; 55). According to the prevailing model, the choice is determined by the identity of the target: once incorporated into a cytoplasmic RISC, the miRNA will specify cleavage if the mRNA has sufficient complementarity to the miRNA, or it will repress productive translation if the mRNA does not have sufficient complementarity to be cleaved but does have a suitable constellation of miRNA complementary sites.

1.3.2 Interplay between transcription factors and microRNAs

The actions of miRNAs and transcription factors (TFs) are often highly coordinated (58), suggesting that the transcriptional and post-transcriptional layers of regulation are strongly correlated and that miRNA functions can be fully understood only by addressing TF and miRNA regulatory interactions together in a single "mixed" network. As it was found in the case of transcriptional regulation (57; 59), in this mixed network several recurrent wiring patterns can be detected, called network motifs (58; 60; 61; 62; 63). In particular, feedback and feedforward loop topologies have been identified as motifs in the mixed network. The common lore is that network motifs were selected by evolution (and are thus overrepresented in the network) to perform elementary regulatory functions. Therefore, understanding the functions associated with this basic regulative circuits can give important insights on the cellular information processing. Moreover, besides their statistical overrepresentation, feedback and feedforward loops involving miRNA regulations have been shown to play a key role in several biological processes, from differentiation to metastasis formation (64), making the study of the functions of simple mixed circuits (which is the topic of this thesis) a relevant biological problem.

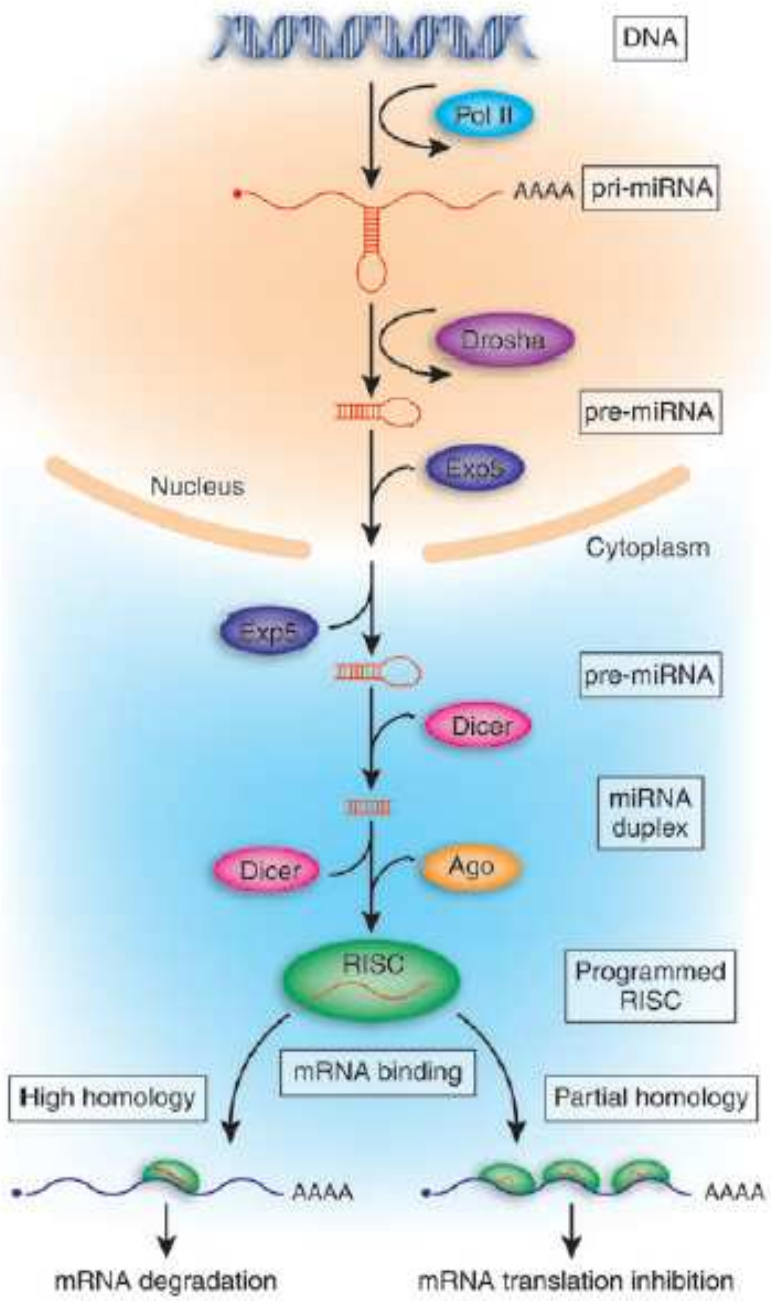

Figure 1.4: MicroRNA biogenesis and mode of action - (Adapted from reference (56))

2

MicroRNA-mediated feedforward loops

2.1 Introduction to the problem

As discussed in section 1.3.2, the mixed network including transcriptional and miRNA regulations presents several recurrent motifs. Among these motifs one of the most interesting is the miRNA-mediated feedforward loop (FFL) in which a master TF regulates a miRNA and, together with it a set of target genes (see Figure 2.1). This motif, which shall be the main interest of this part of the thesis, was recently discussed in (60; 61; 63). In all these papers, despite the fact that the authors used very different computational approaches, it was shown to be remarkably overrepresented in the network, thus supporting the idea that it should play an important regulatory role. Depending on the sign of the transcriptional regulations, FFLs can be divided into two classes: coherent and incoherent (61; 63; 65). In the coherent FFLs both pathways from the TF to the target have the same effect (both repressing or activating target expression), while in the incoherent ones the two pathways have opposite effects. Correspondingly one finds different expression patterns in the two cases: coexpression of miRNA and its target for incoherent FFLs and mutually exclusive expression for the coherent ones (Figure 2.1). This dual picture allows to better understand the complex patterns of correlated expression of miRNAs and their targets observed in experiments (43; 61; 66). In many cases the targets show low expression in miRNA-expressing cells, suggesting coherent regulation. On the other hand, several other cases present an opposite trend, showing that miRNA repression can act in opposition to transcriptional regulation. In fact, different degrees of expression overlap, due to different regulatory circuitries, have been related to different miRNA functions in several recent papers (43; 45; 46; 65; 67). For example, in a coherent FFL as the one in Figure 2.1D, the miRNA expression is induced by an upstream TF that

at the same time represses the target transcription, with the effect of enforcing mutually exclusive domains of expression as the ones observed in the fruit fly (68) or for miR-196 and its target Hoxb8 in mouse (69) and chicken (70). In this cases the miRNA can help the transcriptional repression of a target protein that should not be expressed in a particular cell type, acting as a post-transcriptional failsafe control. Instead, an incoherent FFL (Figure 2.1) can promote high target expression in miRNA-expressing cells, suggesting that miRNAs may have in this case a fine-tuning function, keeping the protein level in the correct functional range. A typical example is the regulation of the atrophin gene by the miRNA miR-8 in *Drosophila*. It was shown (71) that both a too high and a too low level of expression of the atrophin gene could be detrimental for the organism and that miR-8 is mandatory to keep the expression level exactly in the correct range.
It is by now well understood that gene espression is inherently a stochastic process (2; 3; 37). This has particularly relevant effects when the number of proteins and/or messenger RNAs (mRNAs) involved is small and stochastic fluctuations may give sizeable deviations from the mean value of the final protein product. Thus the question that naturally arises is how the cell can reconcile the fine-tuning function described above with these fluctuations. If there is only a relatively narrow protein level which is optimal, the tuning regulation must also prevent protein fluctuations outside the functional range. In fact, it has been conjectured that the incoherent FFLs that enable tuning interaction, can have also a role in buffering noise in the target expression (61; 65; 72).

The main goal of this part of the thesis is to introduce and solve analytically a stochastic model describing these incoherent FFLs in order to give a proof to this conjecture. Our results show that with respect to the simple gene activation by a TF, the introduction of a miRNA-mediated repressing pathway can significantly dampen fluctuations in the target protein output for essentially all the choices of input parameters and initial conditions. As a test of our analysis we also performed extensive numerical simulations which nicely agree with our analytical results. It is important to stress (and we shall discuss this issue in detail in the following) that this noise buffering function is actually a precise consequence of the peculiar topolgy of the FFL. In fact, in order to fine-tune the level of a target protein one would not necessarily need a FFL topology. The same result could well be obtained with an independent miRNA (not under the control of the master TF which activates the target), but this choice would lead to strong fluctuations in the target expression. In the same theoretical framework we can show that the construction of an optimal noise filter does not necessarily imply a strong repression, in agreement with the observation that the miRNA down-regulation of a target is often modest (73; 74).

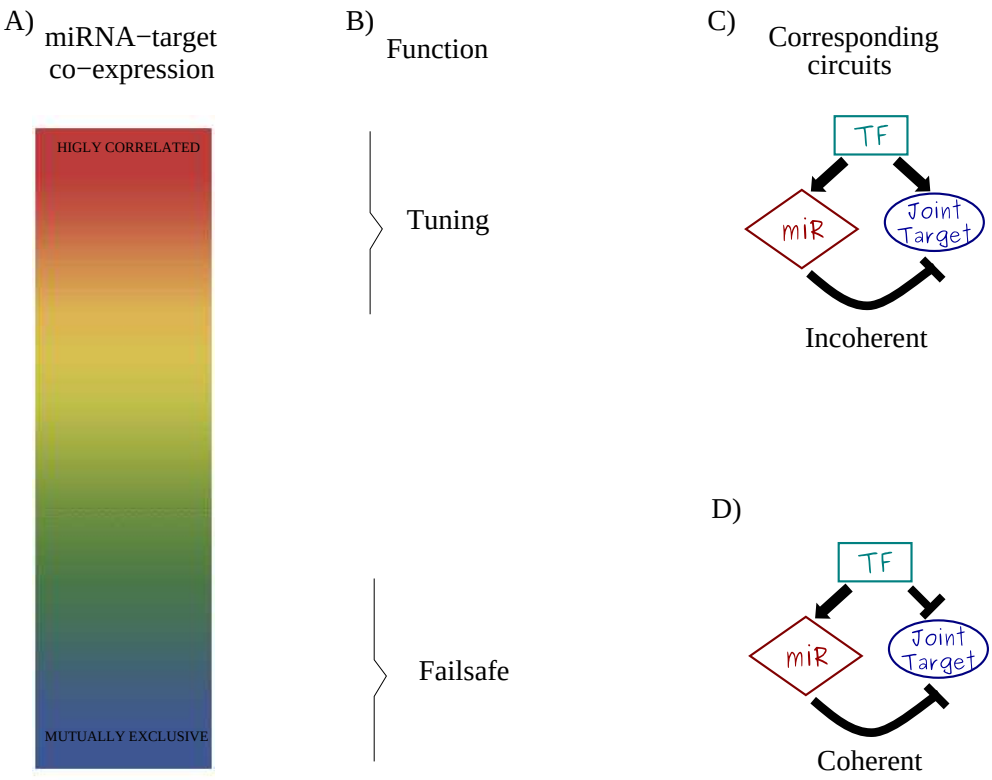

Figure 2.1: Overview of the connections between miRNA-target expression, miRNA function and regulatory circuitry. - (A) MiRNAs and corresponding targets can present different degrees of coexpression between the two extremes of concurrent expression (high correlation) and exclusive domains (high anticorrelation). These two opposite situations are expected to correspond to different functional roles (B) for the miRNA repression. A peculiar expression pattern, evidence of a functional aim, is a consequence of the network structure in which miRNAs are embedded. A high miRNA-target correlation can be achieved through the incoherent FFL (C) where the miRNA repression is opposite to the TF action. Whereas a failsafe control can be performed by a coherent FFL (D), in which the miRNA reinforces the TF action leading to mutually exclusive domains of miRNA-target expression.

2.2 The theoretical framework

2.2.1 Modeling the FFL

Here we focus on the incoherent FFL in Figure 2.2A to present our modeling strategy. For each gene in the circuit we take into account the essential features of transcription, translation, degradation and interactions between genes in the regulatory network (detailed scheme in Figure 2.2A'). Accordingly, the state of the system is described by five variables (w, q, s, r, p) representing: w the number of mRNAs transcribed from the TF gene, q the number of TF molecules, s the number of miRNAs, r the number of mRNAs transcribed from the target gene and p the number of target proteins. We want to explore the mean $(< x_i >)$ and the variance (σ_{x_i}) of each molecular species $x_i \in (w, q, s, r, p)$ and we will show that these quantities can be obtained analitically at the steady-state. The noise strength of the species x_i will be expressed by the coefficient of variation defined as $CV_{x_i} = \sigma_{x_i} / < x_i >$. As usual in this type of models, transcriptional activation is introduced by choosing the rate of transcription of the regulated gene $(k_s(q), k_r(q)$ in our case) as a nonlinear increasing function of the number of TFs (q) present in the cell (1; 23; 75; 76):

$$\begin{aligned} k_r(q) &= \frac{k_r q^c}{h_r^c + q^c} \\ k_s(q) &= \frac{k_s q^c}{h_s^c + q^c}, \end{aligned} \tag{2.1}$$

where h_r and h_s are dissociation constants, specifying the amount of TFs at which the transcription rate is half of its maximum value (k_r and k_s respectively). c is the Hill coefficient and fixes the steepness of the activation curve.
As discussed in section 1.3, the miRNA action can direct translational repression or destabilization of target mRNAs, i.e. it decreases the rate of translation or increases the rate of degradation of target mRNAs. We choose to model the effect of miRNA regulation by taking the translation rate of the target $(k_p(s))$ to be a repressive Hill function of the number of miRNAs (s):

$$k_p(s) = \frac{k_p}{1 + (\frac{s}{h})^c}. \tag{2.2}$$

The parameter h specifies the quantity of miRNAs that determines a rate of translation $k_p/2$, and c is again the Hill coefficient. For simplicity we use the same Hill coefficient c for each Hill function, but the analysis can be straigthforwardly generalized to the case of different steepnesses.
The alternative choice of a degradation rate of mRNAs as a function of miRNA concentration does not yield significantly different results, as reported in details in section 2.7.1. The use of Hill functions to model regulations by miRNAs is coherent with their established catalytic action in animals (77). A stoichiometric

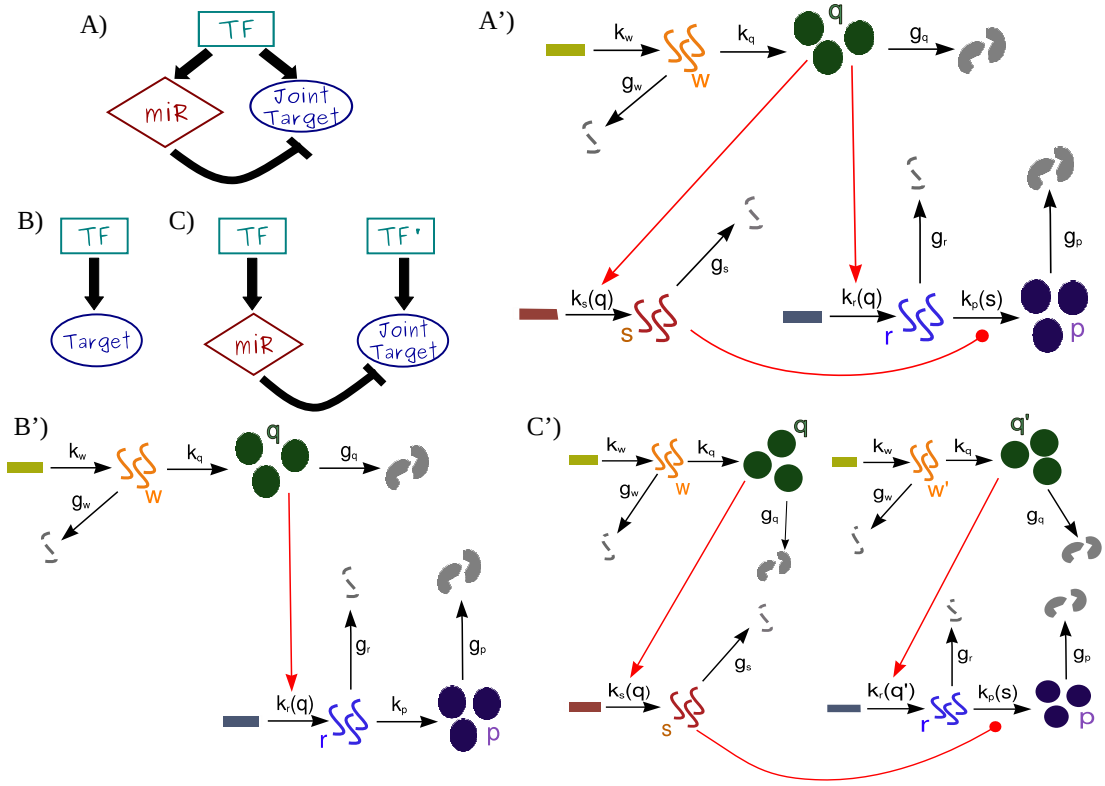

Figure 2.2: Representation of the incoherent FFL and the two circuits used for comparison. - (A) miRNA-mediated incoherent FFL that can be responsible of miRNA-target coexpression; (B) a gene activated by a TF; (C) an open circuit that leads to the same mean concentrations of the molecular species of the FFL in scheme A. (A')(B')(C') Detailed representation of the modelization of the corresponding circuits. Rectangles represent DNA-genes, from which RNAs (w, s, r) are transcribed and eventually degraded (broken lines). RNAs can be translated into proteins (q is the TF while p is the target protein) symbolized by circles, and proteins can be degraded (broken circles). Rates of each process (transcription, translation or degradation) are depicted along the corresponding black arrows. Regulation are represented in red, with arrows in the case of activation by TFs while rounded end lines in the case of miRNA repression. TF regulations act on rates of transcription that become functions of the amount of regulators. MiRNA regulation makes the rate of translation of the target a function of miRNA concentration.

model has been studied in the context of sRNA regulation in bacteria (78; 79; 80), in which each sRNA can pair with one messenger and drive its sequestration or degradation in an irreversible fashion. A comparison with a possible stoichiometric action is shown in section 2.7.2.
The probability of finding in our cell exactly (w, q, s, r, p) molecules at time t satisfies the master equation:

$$\begin{aligned}\partial_t P_{w,q,s,r,p} = \quad & k_w(P_{w-1,q,s,r,p} - P_{w,q,s,r,p}) + k_q w(P_{w,q-1,s,r,p} - P_{w,q,s,r,p}) \\ & + k_r(q)(P_{w,q,s,r-1,p} - P_{w,q,s,r,p}) + k_s(q)(P_{w,q,s-1,r,p} - P_{w,q,s,r,p}) \\ & + k_p(s) r (P_{w,q,s,r,p-1} - P_{w,q,s,r,p}) + g_w\Big[(w+1)P_{w+1,q,s,r,p} - wP_{w,q,s,r,p}\Big] \\ & + g_q\Big[(q+1)P_{w,q+1,s,r,p} - qP_{w,q,s,r,p}\Big] + g_r\Big[(r+1)P_{w,q,s,r+1,p} - rP_{w,q,s,r,p}\Big] \\ & + g_s\Big[(s+1)P_{w,q,s+1,r,p} - sP_{w,q,s,r,p}\Big] + g_p\Big[(p+1)P_{w,q,s,r,p+1} - pP_{w,q,s,r,p}\Big],\end{aligned} \tag{2.3}$$

where $k_w, k_r(q), k_s(q)$ are transcription rates, $k_q, k_p(s)$ are translation rates, and g_{x_i} represents the degradation rate of the species x_i.
In order to solve the master equation for $< x_i >$ and σ_{x_i} for all $x_i \in (w, q, s, r, p)$ at the steady state, we have to linearize Hill functions. This is by now a standard procedure (23; 75). The idea is that at the steady state the distributions of regulators (TFs or miRNAs) have a finite width and sample only small regions of the domains of the corresponding Hill functions. We may therefore approximate Hill functions by their linearizations around the mean values of the regulators q or s:

$$\begin{aligned} k_r(q) &\sim k_r(q)|_{<q>} + \partial_q k_r(q)|_{<q>}(q - < q >) \\ k_s(q) &\sim k_s(q)|_{<q>} + \partial_q k_s(q)|_{<q>}(q - < q >) \\ k_p(s) &\sim k_p(s)|_{<s>} + \partial_s k_p(s)|_{<s>}(s - < s >). \end{aligned} \tag{2.4}$$

Defining:

$$\begin{aligned} k_r^0 &= k_r(q)|_{<q>} - \partial_q k_r(q)|_{<q>} < q > \\ k_r^1 &= \partial_q k_r(q)|_{<q>} \\ k_s^0 &= k_s(q)|_{<q>} - \partial_q k_s(q)|_{<q>} < q > \\ k_s^1 &= \partial_q k_s(q)|_{<q>} \\ k_p^0 &= k_p(s)|_{<s>} - \partial_s k_p(s)|_{<s>} < s > \\ k_p^1 &= -\partial_s k_p(s)|_{<s>}, \end{aligned} \tag{2.5}$$

and substituting in Equations 2.4 we obtain:

$$\begin{aligned} k_r(q) &\sim k_r^0 + k_r^1 q \\ k_s(q) &\sim k_s^0 + k_s^1 q \\ k_p(s) &\sim k_p^0 - k_p^1 s. \end{aligned} \tag{2.6}$$

We would like to emphasize that linearizing the Hill functions does not mean to linearize the model. In fact, even with a linearized dependence on the miRNA copy number, our model keeps a nonlinear contribution in the term encoding the target translation (due to the fact that it depends on both the number of miRNAs and mRNAs). As we will see later this nonlinearity leads to non trivial consequences.
Despite this nonlinearity, the moment generating function approach (23; 75; 81) can be succesfully used. By defining the generating function:

$$F(z_1, z_2, z_3, z_4, z_5) = \sum_{w,q,s,r,p} z_1^w \; z_2^q \; z_3^s \; z_4^r \; z_5^p \; P_{w,q,s,r,p}, \tag{2.7}$$

and using the linearization in equation 2.6 we can convert equation 2.3 into a second-order partial differential equation:

$$\begin{aligned} \partial_t F \;=\; & k_w(z_1 F - F) + k_q z_1(z_2 \partial_{z_1} F - \partial_{z_1} F) + k_r^0(z_4 F - F) \\ & + k_r^1 z_2(z_4 \partial_{z_2} F - \partial_{z_2} F) + k_s^0(z_3 F - F) + k_s^1 z_2(z_3 \partial_{z_2} F - \partial_{z_2} F) \\ & + k_p^0 z_4(z_5 \partial_{z_4} F - \partial_{z_4} F) - k_p^1 z_3 z_4(z_5 \partial_{z_3,z_4} F - \partial_{z_3,z_4} F) \\ & + g_w(\partial_{z_1} F - z_1 \partial_{z_1} F) + g_q(\partial_{z_2} F - z_2 \partial_{z_2} F) + g_s(\partial_{z_3} F - z_3 \partial_{z_3} F) \\ & + g_r(\partial_{z_4} F - z_4 \partial_{z_4} F) + g_p(\partial_{z_5} F - z_5 \partial_{z_5} F). \end{aligned} \tag{2.8}$$

We now use the following properties of the moment generating function: $F|_1 = 1$; $\partial_{z_i} F =< x_i >$; $\partial_{z_i}^2 F =< x_i^2 > - < x_i >$, where $|_1$ means evaluation of F at $x_i = 1$ for all i. At the steady state ($\partial_t F = 0$) differentiation of equation 2.8 generates equations for successively higher moments. In particular we are interested in $< p >$ and σ_p and differentiating up to the fourth moments leads to their analytical expressions.
Noise in gene expression is originated by a combination of two types of fluctuations: intrinsic and extrinsic ones. Intrinsic fluctuations are essentially due to the stochasticity of the gene expression process. Extrinsic ones, instead, are due to the environment. In the latter case a prominent role is played by the noise transmitted by upstream genes (19; 20).
As a matter of fact there is a certain degree of arbitrariness in the definition of extrinsic and intrinsic noise (33). Since we focus on the target production we define as "intrinsic" the noise derived from the stochastic steps of its expression (transcription, translation and degradation) and as "extrinsic" the noise propagating from its regulators (s, q) that makes the parameters $(k_r(q), k_p(s))$ fluctuate through the Hill functions. Therefore, in our model we do not have to include extrinsic noise with an arbitrary distribution as it naturally arises from the stochastic steps of production of regulators and propagates to the target gene.

2.3 Comparison with null models

2.3.1 Comparison with a transcription factor control

To show the noise buffering role of the miRNA-mediated incoherent FFL (Figure 2.2A) we first compare it to a simpler process: a gene activated by a TF (Figure 2.2B), without any post-transcriptional regulation. The strategy used to model this linear network is equivalent to the one explained in the previous section for the FFL. The master equation that describes the circuit in the scheme of Figure 2.2B' is:

$$\begin{aligned}\frac{\partial P_{w,q,r,p}}{\partial t} = \quad & k_w(P_{w-1,q,r,p} - P_{w,q,r,p}) + k_q w(P_{w,q-1,r,p} - P_{w,q,r,p}) \\ & + k_r(q)(P_{w,q,r-1,p} - P_{w,q,r,p}) + k_p r(P_{w,q,r,p-1} - P_{w,q,s,r,p}) \\ & + g_w\Big[(w+1)P_{w+1,q,r,p} - wP_{w,q,r,p}\Big] + g_q\Big[(q+1)P_{w,q+1,r,p} - qP_{w,q,r,p}\Big] \\ & + g_r\Big[(r+1)P_{w,q,r+1,p} - rP_{w,q,r,p}\Big] + g_p\Big[(p+1)P_{w,q,s,r,p+1} - pP_{w,q,s,r,p}\Big].\end{aligned} \tag{2.9}$$

Introducing the moment generating function:

$$F(z_1, z_2, z_3, z_4) = \sum_{w,q,r,p} z_1^w \; z_2^q \; z_3^r \; z_4^p \; P_{w,q,r,p}, \tag{2.10}$$

and using the linearized form of Hill functions in Equations 2.6, Equation 2.9 can be converted into a first order partial differential equation (PDE):

$$\begin{aligned}\partial_t F \quad = \quad & k_w(z_1F - F) + k_q z_1(z_2\partial_{z_1}F - \partial_{z_1}F) + k_r^0(z_3F - F) \\ & + k_r^1 z_2(z_3\partial_{z_2}F - \partial_{z_2}F) + k_p z_3(z_4\partial_{z_3}F - \partial_{z_3}F) \\ & + g_w(\partial_{z_1}F - z_1\partial_{z_1}F) + g_q(\partial_{z_2}F - z_2\partial_{z_2}F) \\ & + g_r(\partial_{z_3}F - z_3\partial_{z_3}F) + g_p(\partial_{z_4}F - z_4\partial_{z_4}F).\end{aligned} \tag{2.11}$$

It is now easy to extract the first two moments of the probability distribution $P_{w,q,r,s}$ at the steady state, and thus a close expression for p and $CV_p = \sigma_p / < p >$.

Starting from a gene activated by a TF, in principle the gain of a new regulator implies also a new source of extrinsic noise for the target, given that the fluctuations in the number of regulators propagate to downstream genes and, as discussed in (18), the addition of extrinsic fluctuations generally increases the noise of a system. However, the peculiar structure of the FFL can overcome this problem, actually reducing noise, as was previously shown in the case of negative transcriptional auto-regulation (22). Given that the two circuits lead to different mean values, the comparison of noise strengths in target protein will be done in

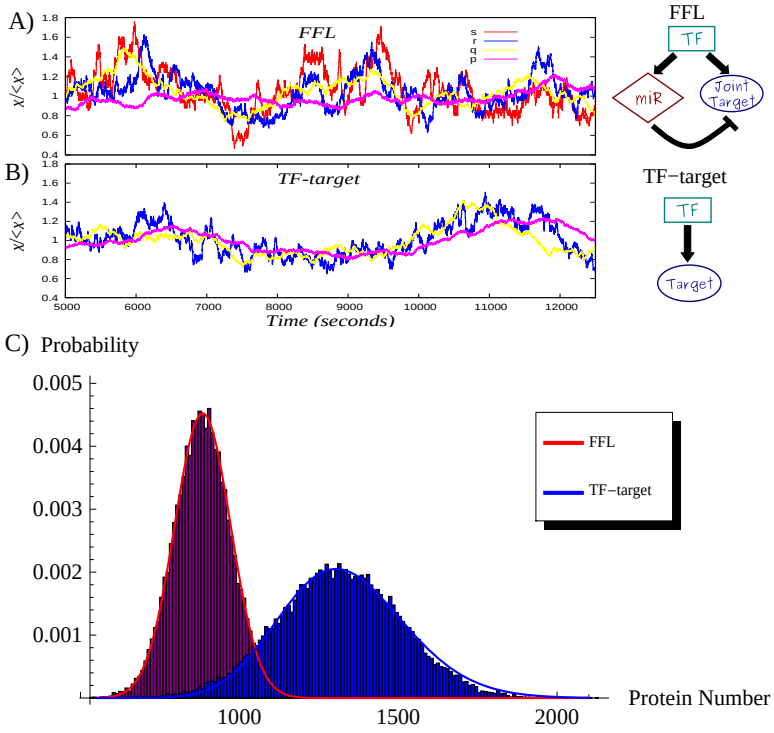

Figure 2.3: Noise properties of the FFL compared with a TF-gene linear circuit. - (A) An example of simulation results for the FFL (scheme on the right or more detailed in Figure 2.2A'). The normalized trajectory of each molecular species is shown as a function of time after reaching the steady state. The rate of transcription of the TF is $k_w = 0.06s^{-1}$ and of translation $k_q = 0.04s^{-1}$. Proteins degrade with rate $g_q = g_p = 0.002s^{-1}$ while mRNAs and miRNAs with rate $g_w = g_r = g_s = 0.006s^{-1}$. The parameters in the Hill functions of regulation (equations 2.1,2.2) are the following: the maximum rate of transcription for mRNAs is $k_r = 0.8s^{-1}$, while for miRNAs is $k_s = 0.5s^{-1}$; the maximum rate of translation of the target is $k_p = 0.04s^{-1}$; dissociation constants are $h_s = 200, h_r = 200, h = 60$; Hill coefficients are all $c = 2$, as typical biological values range from 1 (hyperbolic control) to 30 (sharp switching)(23). (B) Time evolution in a simulation for the molecular players in the linear TF-gene cascade (scheme on the right or more detailed in Figure 2.2B'). Compared to the FFL case, the p evolution is more sensitive to TF fluctuations. (C) The probability distribution of protein number for the two circuits.

terms of the coefficient of variation ($CV_p = \sigma_p / < p >$). The results of the noise properties of the two circuits are reported in Figure 2.3, where histograms are the result of Gillespie simulations while continuous lines are empirical distributions (gaussian for the FFL and gamma for the TF-gene) with mean and variance predicted by the analytical model. With the parameter choice explained in the caption of Figure 2.3, the predicted CV_p are 0.147 and 0.1 for the TF-gene cascade and the FFL respectively. Therefore, the introduction of the miRNA pathway not only controls the mean value but also reduces the relative fluctuations. This effect can be clearly seen looking at the shape of the probability distributions in Figure 2.3C. It is rather easy to understand the origin of this noise buffering effect: any fluctuation in the concentration of TFs affects the rate of mRNA transcription, driving consequently the target protein away from its steady state, but mRNA and miRNA concentrations tend to vary in the same direction in the FFL. In this way, miRNAs can always tune the protein production against TF fluctuations. As can be seen in Figure 2.3A and B , there is a certain degree of correlation in the time evolution of q, r, p due to noise propagation, despite the overimposed higher-frequency intrinsic noise of each molecular species, but in the case of the FFL the p trajectory is less sensitive to q fluctuations thanks to the action of miRNAs (s). It is important to stress that this result is not affected by the Hill function linearization discussed above. In fact, the predictions of the model are in good agreement with Gillespie simulations (which keep into account the full nonlinear repressing and activating Hill functions). Moreover our results turn out to be robust with respect to parameter choice, showing a rather stable noise reduction essentially for any choice of expression and degradation constants, as discussed in more detail in section 2.8.

2.3.2 Comparison with an open circuit

The same fine-tuning of the mean target concentration achieved with a FFL could be equally obtained with an open circuit like the one in Figure 2.2C, where the miRNA gene is controlled by an independent TF. If the two TFs, activating the miRNA and target gene expression, have the same rate of transcription, translation and degradation of the single master TF in the FFL -as well as the other model parameters as in Figure 2.2A' and C'- the stationary mean levels of the various molecular species are the same in both circuits. In particular, the mean concentration of the target protein can be fine-tuned to the same desired value by both circuits. In fact, a deterministic description of the open circuit dynamics is given by the equations:

$$\begin{aligned}
\frac{dw}{dt} &= k_w - g_w w \\
\frac{dq}{dt} &= k_q w - g_q q \\
\frac{dw'}{dt} &= k_w - g_w w' \\
\frac{dq'}{dt} &= k_q w' - g_q q' \\
\frac{ds}{dt} &= k_s(q) - g_s s \\
\frac{dr}{dt} &= k_r(q') - g_r r \\
\frac{dp}{dt} &= k_p(s) r - g_p p.
\end{aligned} \tag{2.12}$$

The presence of two independent TFs (copy numbers q and q') that regulate respectively the transcription of s and r does not change the steady-state expression p_{ss} that can be obtained in the FFL case. This is true as long as all the parameters are exactly the same. The solutions of Equations 2.12 at equilibrium (with $c = 2$) are:

$$\begin{aligned}
w_{ss} &= w'_{ss} = \frac{k_w}{g_w} \\
q_{ss} &= q'_{ss} = \frac{k_q k_w}{g_q g_w} \\
s_{ss} &= \frac{k_q^2 k_s k_w^2}{g_s(g_q^2 g_w^2 h_s^2 + k_q^2 k_w^2)} \\
r_{ss} &= \frac{k_q^2 k_r k_w^2}{g_r(g_q^2 g_w^2 h_r^2 + k_q^2 k_w^2)} \\
p_{ss} &= \frac{h^2 k_p k_q^2 k_r k_w^2}{g_p g_r (g_q^2 g_w^2 h_r^2 + k_q^2 k_w^2)\left(h^2 + \frac{k_q^4 k_s^2 k_w^4}{g_s^2 (g_q^2 g_w^2 h_s^2 + k_q^2 k_w^2)^2}\right)}.
\end{aligned} \tag{2.13}$$

It is easy to check that this is also the solution of a deterministic description of the FFL in Figure 2.2A'. Therefore, the open circuit allows the same setting of the target level that can be achieved with a FFL. This makes the open circuit so constructed a good null model for a comparison to the FFL: as the mean field description is the same, any difference between the two will be due to stochastic fluctuations and actually the behaviour of fluctuations is completely different.

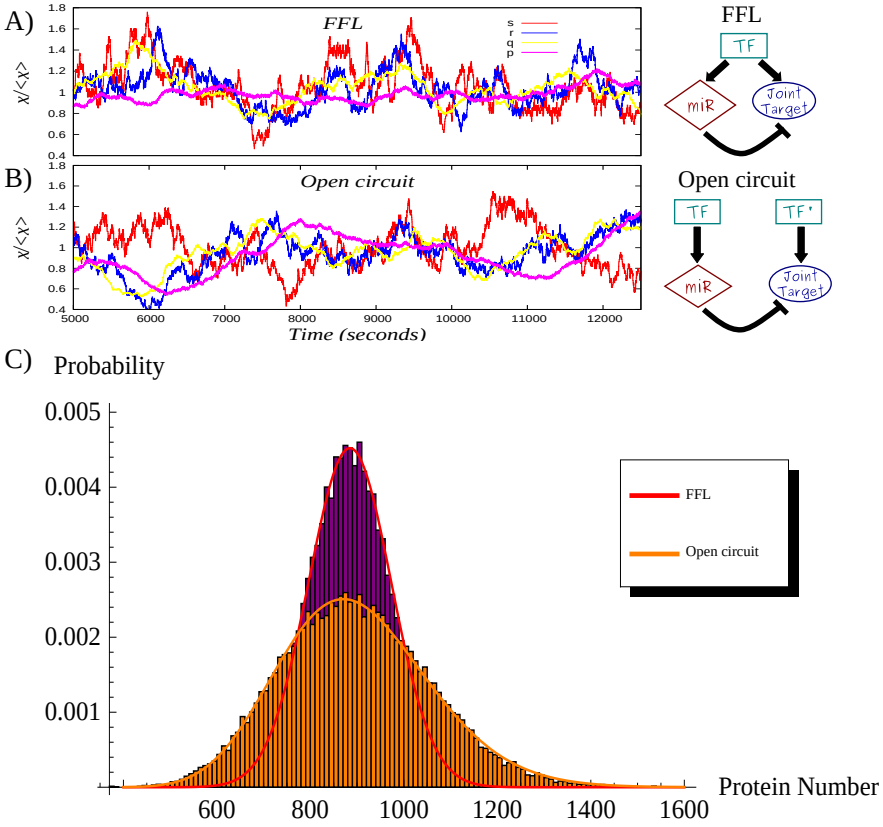

Figure 2.4: Noise properties of the FFL compared with an analogous open circuit. - (A) An example of simulation results for the FFL (scheme on the right or more detailed in Figure 2A'). The parameter values are the same of Figure 3.(B) Time evolution in a simulation for the molecular players in the open circuit (scheme on the right or more detailed in Figure 2C'). The correlation between the s and r trajectories that is present in the FFL (A) is completely lost in the open circuit. As a consequence while the mean value of p is approximately the same, its fluctuations are appreciably greater in the open circuit case.(C) The probability distribution of protein number for the two circuits. Histograms are the result of Gillespie simulations while continuous lines are empirical distributions (gaussian for the FFL and gamma for the open circuit) with mean and variance predicted by the analytical model.

As we shall see below, the open circuit leads to much larger fluctuations in the final product than the FFL. It is well possible that this is the reason for which FFLs have been positively selected by evolution and are presently over-represented in the mixed TF-miRNA regulatory network. In fact, fine-tuning can be implemented advantageously only together with fluctuations control: a precise setting of the mean value of a target protein is useless without a simultaneous damping of the stochastic fluctuations. To assess this result, we used the same strategy discussed in the previous sections: we solved analitically for both circuits the master equation and tested our results with a set of Gillespie simulations. Our results are shown in Figure 2.4: the lack of correlation between the miRNA and mRNA trajectories in the open circuit (Figure 2.4B) leads to much larger deviations from the mean number of proteins with respect to the FFL case. Using the same parameter values of Figure 2.3, the predicted CV_p for the open circuit is $CV_p = 0.175$, to be compared with the value $CV_p = 0.1$ of the FFL. Different cell-to-cell variability can be clearly seen comparing the distributions of the number of target proteins for the two circuits (Figure 2.4C). Note that a target embedded in an open circuit has an even more noisy expression than a gene simply regulated by a TF, for which $CV_p = 0.147$

2.3.3 The incoherent feedforward loop is effective in reducing extrinsic fluctuations

In the previous sections we compared different regulatory circuits keeping the same amount of input noise, i.e. the same amount of fluctuations in the copy number of master TFs. Since the topology of a regulatory motif can have stronger effects on extrinsic rather than intrinsic noise (18), it would be very interesting to study how the mixed incoherent FFL behaves as a function of such extrinsic noise. As a matter of fact, extrinsic and intrinsic fluctuations are generally coupled in a non-trivial way in biochemical networks (38), but we developed a strategy to control fluctuations in upstream TF expression, known to be one of the major sources of extrinsic noise in eukaryotes (20), without affecting the copy number of the molecular species in the circuit. From equation 2.3 we can calculate $< q >$ (which denotes the mean number of TFs) and its noise strength CV_q:

$$\begin{aligned} < q > &= \frac{k_q k_w}{g_q g_w} \\ CV_q &= \frac{1}{< q >}\sqrt{< q > \frac{g_q + g_w + k_q}{g_q + g_w}}, \end{aligned} \tag{2.14}$$

where, as above, the parameters k_w and k_q denote the rate of transcription and translation of the TF respectively and g_w and g_q the corresponding degradation constants. Setting the rates of degradation (k_q and k_w) and the product $k_w k_q$ to constant values, we end up with: $< q > \sim$ *constant* and $CV_q \sim \sqrt{k_q}$. This

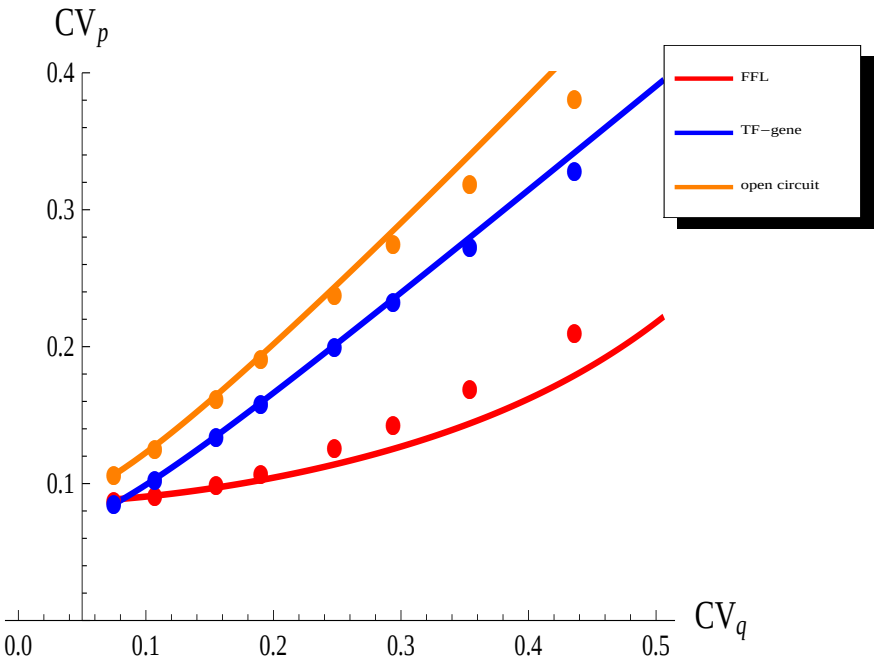

Figure 2.5: The effect of fluctuations in an upstream TF. - We maintain constant the number of TFs $< q >$, while we vary its relative fluctuations CV_q, tuning the relative contribution of transcription (rate k_w) and translation (rate k_q). All the other parameters have the values reported in caption of Figure 3. The incoherent FFL makes the target less sensitive to fluctuations in the upstream TF. The extent of the noise reduction, with respect to the other circuits, depends on the magnitude of the TF noise, pointing out that the FFL topology is particularly effective in filtering out extrinsic fluctuations. Dots are the result of Gillespie simulations with the full nonlinear dynamics while continuous lines are analytical predictions.

is indeed a well known result (presented in section 1.2.4): fluctuations in the protein level are stronger when the rate of translation is higher (3) and can be tuned (while keeping the mean value $< q >$ fixed) by changing the ratio k_w/k_q with $k_w k_q = constant$. This represents a perfect theoretical setting to test the dependence of the target noise CV_p on the input noise CV_q, while maintaining unchanged the mean value of all the molecular species involved in the circuit. We report in Figure 2.5 the results of such analysis for the three circuits discussed in the previous sections. While extrinsic fluctuations increase, so does the FFL's performance in filtering out noise, compared to the other circuits. Once again it is easy to understand the reason of this behaviour. The FFL architecture channels fluctuations of an upstream factor into parameters with opposite effect on the target protein, forcing them to combine destructively. Therefore, the specific FFL topology seems effective in the maintenance of gene expression robustness despite noisy upstream regulators. The introduction of a miRNA pathway, building a FFL from a TF-gene cascade, really makes the difference in situations of strong upstream noise. This feature can explain why miRNAs can often be deleted without observable consequences (72), since experiments usually do not measure fluctuations and are typically performed in controlled environments, where noise is kept at minimal levels.

2.4 Deviant effects

Stochastic equations are the natural formalism to describe a set of biochemical reactions when the number of molecules involved is small and thus fluctuations are important. As the number of molecules increases the stochastic description smoothly converges, at least for linear systems, toward a deterministic one and stochastic equations can be substituited by ordinary differential equations (ODE). It is usually expected that even in the regime in which fluctuations cannot be neglected one could use ODE if interested only in the time evolution of the mean values. This approximation can be thought as a sort of "mean field" approach (by analogy with statistical mechanics where the mean field approximation is implemented by neglecting fluctuations). However, similarly to what happens in statistical mechanics in the vicinity of a critical point , it may happen that, even at the level of mean values, the ODE description does not coincide with the (more rigorous) stochastic one. These breakdowns between the two descriptions are known as "deviant effects" (82) and are typically a consequence of nonlinear terms in the equations or strong extrinsic fluctuations (18; 35). In some cases these deviant effects can have relevant phenomenological consequences. This is exactly the case of our system: although the FFL (Figure 2.2A,A') and the open circuit (Figure 2.2C,C') have the same deterministic description at the steady state, the master equation approach gives a mean value of the target protein systematically lower in the FFL circuit, with respect to the same quantity in the open circuit. This is a non trivial consequence of the correlated fluctuations in the number of mRNAs and miRNAs in the FFL. This correlation between fluctuations obviously cannot be taken into account in the deterministic description and as a consequence the ODE analysis correctly describes the steady state mean number

of target proteins only for the open circuit. This result can be understood looking at the analytical expression of the mean value of p:

$$< p >= k_p^0 < r > -k_p^1 < rs > . \tag{2.15}$$

In a FFL, fluctuations of r and s are highly correlated (Figure 2.3A), because the transcription rates of messengers and miRNAs depend on a shared TF. The result is that in this case $< rs > > < r >< s >$. On the other hand, the production of s and r is independently regulated in an open circuit, implying that $< rs > \sim < r >< s >$. A deterministic description does not take into account fluctuations so correctly describes $< p >$ only when uncorrelated noise is averaged out without affecting mean values. In conclusion, the correlation in fluctuations introduced by the FFL topology affects the target protein mean value, establishing a systematic decrease with respect to the deterministic description. This deviant effect can be substantial and underlines the necessity of a stochastic nonlinear modeling. A fully linearized description, as for example the one proposed by (75) for post-transcriptional regulation, would not be able to show this type of effects.

2.5 Noise filtering optimization

The efficiency of the FFL in controlling the fluctuations of the target protein is a function of three main parameters: the number of master TFs (which in turn is a function of k_w and k_q), the number of miRNA copies (function of k_s and h_s) and the strength of miRNA repression (defined as $1/h$). In this section we shall study the efficiency of the FFL in buffering noise as a function of each one of these three quantities, changing a corresponding parameter while keeping fixed all others. As we shall see, in all the three cases the noise reduction with respect to a simple TF-target interaction (i.e. without the miRNA) shows a U-shaped profile with a well defined minimum which allows us to identify the values of the parameters which optimize the noise reduction properties of the FFL. This pattern is rather robust, and only marginally depends on the way the variable of interest is tuned (for instance, by changing k_s or h_s in the case of miRNA concentration). It is important to stress that in all three cases optimal noise filtering does not imply strong repression, a result which well agrees with the observation that miRNAs embedded in an incoherent FFL usually have a fine-tuning effect on the targets instead of switching them off completely. It is exactly in the intermediate region of the parameters, in which fine-tuning occurs, that we also have optimal noise reduction.

2.5.1 Optimal repression strength

The repression strength is defined as $1/h$ (inverse of the dissociation constant in the Hill function of equation 2.2). As it can be seen in Figure 2.6A, the FFL exhibits a noise profile with a typical U-shape and reaches an optimal value of noise

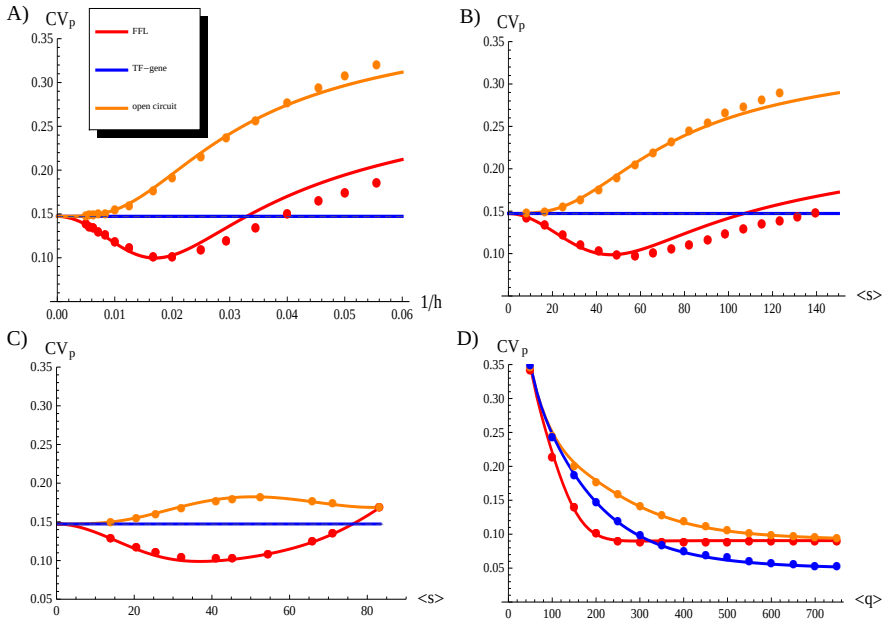

Figure 2.6: How an optimal noise filter can be built. - (A) The coefficient of variation of the target protein CV_p as a function of the repression strength $1/h$. The Figure shows the presence of an optimal repression strength for which the introduction of a miRNA regulation in a FFL leads the noise to a minimum. (B) CV_p as a function of the mean number of miRNAs $< s >$. In this case $< s >$ is changed through the maximum rate of transcription k_s (see equation 2.1). (C) CV_p as a function of $< s >$ by varying the dissociation constant h_s. In both cases (B and C) is evident an U-shaped profile, implying an optimal noise buffering for intermediate miRNA concentrations. (D) CV_p as a function of the mean number of TFs $< q >$. The number of TFs depends on the rate of their transcription k_w and of their translation k_q. The Figure is obtained manipulating k_q, but the alternative choice of k_w leads to equivalent results, as shown in section 2.8. For intermediate concentration of the TF, the noise control by the FFL outperforms the one of the other circuits. In each plot dots are the result of Gillespie simulations while continuous lines are analytical predictions. The values of parameters kept constant are the same of Figure 2.3, however the results are quite robust with respect to their choice.

reduction (measured as the difference in the noise strength CV_p with respect to the simple TF-gene circuit) for intermediate values of repression strength. The open circuit, constrained to give the same mean value $< p >$, always introduces larger target fluctuations. As mentioned above, optimal noise filtering does not require strong target repression. For instance, with the choice of parameter values of Figure 2.6, the optimal noise reduction is obtained for a mean value of the target protein which is about 66% of the value obtained without the miRNA, i.e. with a simple TF-target interaction. This prediction could be experimentally tested via manipulation of the repression strength, in analogy to the work of (83) on the transcriptional autoregulatory motif. It is instructive to notice the analogies between the behaviour of our FFLs and that of the negative transcriptional autoregulation loops. These purely transcriptional network motifs occur ubiquitously in transcriptional regulatory networks and were recently studied in great detail (18; 84). Also in this case optimal noise filtering is obtained for a well defined value of the repression strength. It is easy to understand the reason of this behaviour. In this motif the protein expressed from a gene inhibits its own transcription by increasing expression when protein numbers are low, while decreasing expression when protein numbers are high. Increasing the repression strength improves the circuit potential to reduce stochasticity, but at the same time decreases the copy number of mRNAs and proteins, with a rise in intrinsic fluctuations that can overcome any attenuation. Consistently, experiments show a well defined range of repression strength for which noise minimization is optimal (83).

2.5.2 Optimal miRNA concentration

Another variable which can be tuned in order to achieve optimal noise reduction is the miRNA concentration. If we keep the number of TFs constant then the miRNA concentration $< s >$ can only depend on the maximum rate of transcription of the miRNA gene (k_s) and on the affinity of its promoter to the TF (proportional to $1/h_s$, where h_s is the dissociation constant in equation 2.1). In Figure 2.6B and C we analyze the noise strength CV_p of the target protein in the FFL for different miRNA concentrations and compare it to the CV_p obtained with the simple TF-gene interaction and with the open circuit. Changing the miRNA concentration by varying k_s (Figure 2.6B) or h_s (Figure 2.6C), we find very similar U-shaped profiles for CV_p. As for the previous analysis, also in this case it is possible to find an optimal miRNA concentration, and again it is such that the effect on the protein target is only a modest reduction (in this case $\sim 60\%$ of the value obtained without the miRNA). Apart from the conserved U-shaped profile, there are rather deep differences in the noise behaviour depending on the choice of the tuning parameter. In fact, while an increase of k_s always induces an increase of $< s >$, this quantity becomes insensitive to h_s above a certain threshold. Since the number of TFs is constant in this analysis, it is clear that increasing $1/h_s$ (Figure 2.6C) the system can reach at best the value of $< s >$ consistent with the maximum rate of transcription. At the same time a large value of $1/h_s$ means that very few TFs are enough to support the maximum transcription rate for the miRNA gene, so fluctuations in the number of TFs become irrelevant despite the topology of the circuit. As a consequence, the CV_p curves for the FFL and

the open circuit converge to a commom value (Figure 2.6C). A refined experimental control of miRNA concentration through graded miRNA overexpression or silencing would permit a test of the U-shaped profile of CV_p in a FFL.

2.5.3 Optimal TF concentration

The last case that we consider in this section is the effect of different TF concentrations on the noise buffering properties of the FFL. It is expected that for high TF concentrations (i.e. high values of $< q >$) the activation functions in Equations 2.1 reach the saturation point, making the system insensitive to variations in TF concentration. As long as the number of TFs does not fluctuates below the saturation point, the miRNA and the target gene are maximally transcribed, with no reference to the exact number of TFs. Accordingly, CV_p becomes asymptotically constant for large $< q >$ for each circuit topology (Figure 2.6D). The gap between the asymptotic values of the direct TF regulation and the two other circuits is due to the fact that the former does not suffer for the additional external noise due to the fluctuations in the miRNA number. On the other hand, for small values of $< q >$ also the number of target proteins is very small as its expression is hardly activated regardless of the circuit type, with a consequent increase of the noise strength (Figure 2.6D). The central region is the most interesting one: this is the region in which the system is maximally sensitive to changes in TF concentration. In this regime the FFL outperforms both the simple direct regulation and the open circuit in buffering noise. Also in this case the optimal TF concentration is placed in a region corresponding to a modest reduction of $< p >$, despite the miRNA repression.

2.5.4 Exploring the parameter space

To give a more comprehensive insight into the relation between noise control and target repression, we finally evaluate the buffering of fluctuations (CV_p/CV_{p_0}) for each average number of TFs $< q >$ and each degree of target suppression ($< p > / < p_0 >$), where $< p_0 >$ and CV_{p_0} represent here the constitutive mean expression and fluctuations in absence of miRNA regulation. Results of this analysis are reported in Figure 2.7A. As discussed above, noise reduction can be implemented successfully in the parameter region where the target activation function (in Figure 2.7B) is not saturated, since this is the region where the target is sensitive to changes in TF concentrations and therefore also to its fluctuations, regardless of the presence or absence of miRNA regulation. It is exactly in this region that noise buffering can be observed (compare Figure 2.7A and B). In particular, for each TF concentration the best noise reduction appears for a target level again around the 60% of its constitutive expression. In the optimal setting, noise can be remarkably reduced to about one half of its constitutive value, but the reduction remains substantial also for weaker repressions, further confirming that a strong miRNA repression is not required for noise control.

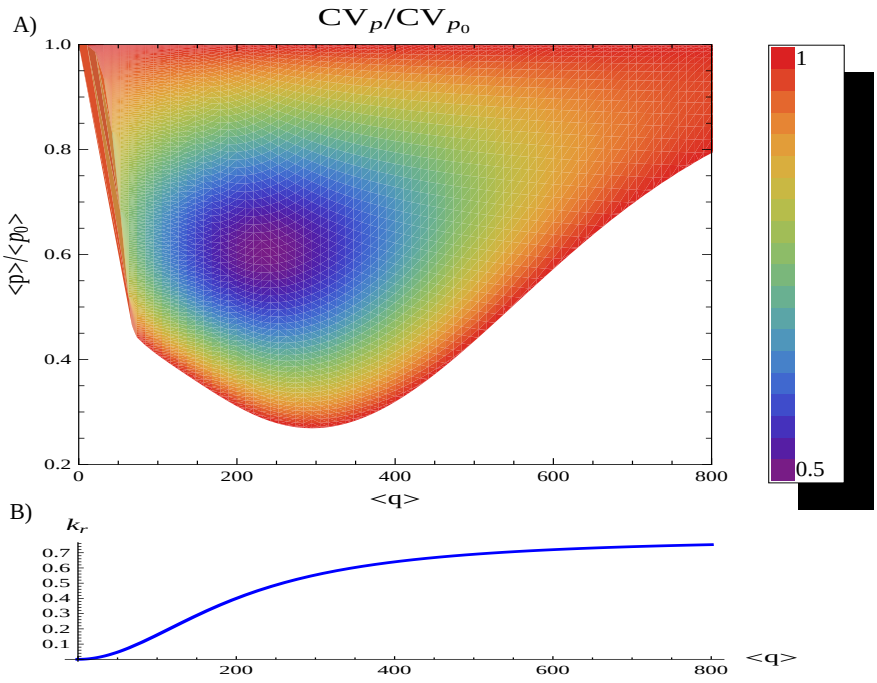

Figure 2.7: Exploring the parameter space. - (A) The target noise CV_p, achieved with the FFL, is evaluated with respect to noise deriving from constitutive expression CV_{p_0} (i.e. in absence of miRNA regulation) for different mean level of the TF $< q >$ and different degrees of reduction of the target protein level $< p > / < p_0 >$ (where $< p_0 >$ is the mean consitutive expression). The TF level is changed through its rate of translation k_q (equivalent results can be obtained changing the rate of transcription k_w), while the target level is tuned by varying the repression strength. All the other parameters have the values reported in caption of Figure 2.3 unless $k_w = 0.01263$ (lower than in Figure 2.3 to explore a more noisy situation). The region where miRNA repression leads to larger fluctuations with respect to constitutive ones is shown in white, while when a noise reduction is gained the value of CV_p/CV_{p_0} is reported with the color code explained in the legend. The best noise control is achieved with a modest suppression of target expression, around the 60% of its constitutive value. (B) The rate of transcription of the target mRNA as a function of the mean number of TFs. The noise reduction shown in the above plot can be obtained outside the saturation regime (where the slope of the activation curve tends to zero).

We consider the characterization of the optimal setting of miRNA-mediated incoherent FFLs for noise buffering, and the resulting U-shaped profile predicted for the noise reduction factor, as one of the major results of our analysis which, we expect, should be amenable of direct experimental validation. The fact that an optimal noise filtering is obtained with a rather modest reduction in the amount of the target protein also agrees with the recent experimental observation that miRNA down-regulation of targets is often modest (73; 74) and apparently dispensable from a functional point of view. In this respect it is tempting to conjecture that, at least for some targets of incoherent FFLs, the down-regulation could only be the side effect of an evolutionary design aiming instead to optimize noise reduction.

2.6 Comparison with purely transcriptional incoherent feedforward loops

The capability of incoherent FFLs to reduce fluctuations was previously studied with simulations in the context of purely transcriptional networks (18). This section presents a comparison of the noise properties of microRNA-mediated FFLs (scheme in Figure 2.1A') and purely transcriptional ones (detailed scheme of reactions in Figure 2.8A), where the miRNA is replaced by a protein that inhibits transcription (rather than translation, as miRNAs do). The transcriptional FFL can be modeled with the same strategy explained previously for the miRNA-mediated version and analogously mean values and standard deviations of the various molecular species can be calculated analytically with the moment generating function method.

The master equation describing a purely transcriptional incoherent FFL (depicted in the scheme of Fig. 2.8A) is:

$$\begin{aligned}
\frac{\partial P_{w,q,s,j,r,p}}{\partial t} = \quad & k_w(P_{w-1,q,s,j,r,p} - P_{w,q,s,j,r,p}) + k_q w(P_{w,q-1,s,j,r,p} - P_{w,q,s,j,r,p}) \\
& + k_s(q)(P_{w,q,s-1,j,r,p} - P_{w,q,s,j,r,p}) + k_j s(P_{w,q,s,j-1,r,p} - P_{w,q,s,j,r,p}) \\
& + k_r(q,j)(P_{w,q,s,j,r-1,p} - P_{w,q,s,j,r,p}) + k_p r(P_{w,q,s,j,r,p-1} - P_{w,q,s,j,r,p}) \\
& + g_w\Big[(w+1)P_{w+1,q,s,j,r,p} - wP_{w,q,s,j,r,p}\Big] + g_q\Big[(q+1)P_{w,q+1,s,j,r,p} - qP_{w,q,s,j,r,p}\Big] \\
& + g_s\Big[(s+1)P_{w,q,s+1,j,r,p} - sP_{w,q,s,j,r,p}\Big] + g_j\Big[(j+1)P_{w,q,s,j+1,r,p} - jP_{w,q,s,j,r,p}\Big] \\
& + g_r\Big[(r+1)P_{w,q,s,j,r+1,p} - rP_{w,q,s,j,r,p}\Big] + g_p\Big[(p+1)P_{w,q,s,j,r,p+1} - pP_{w,q,s,j,r,p}\Big].
\end{aligned} \tag{2.16}$$

The protein j represses target transcription, which is also activated by the master TF q. Consequently, the rate $k_r(q,j)$ can be represented by a function

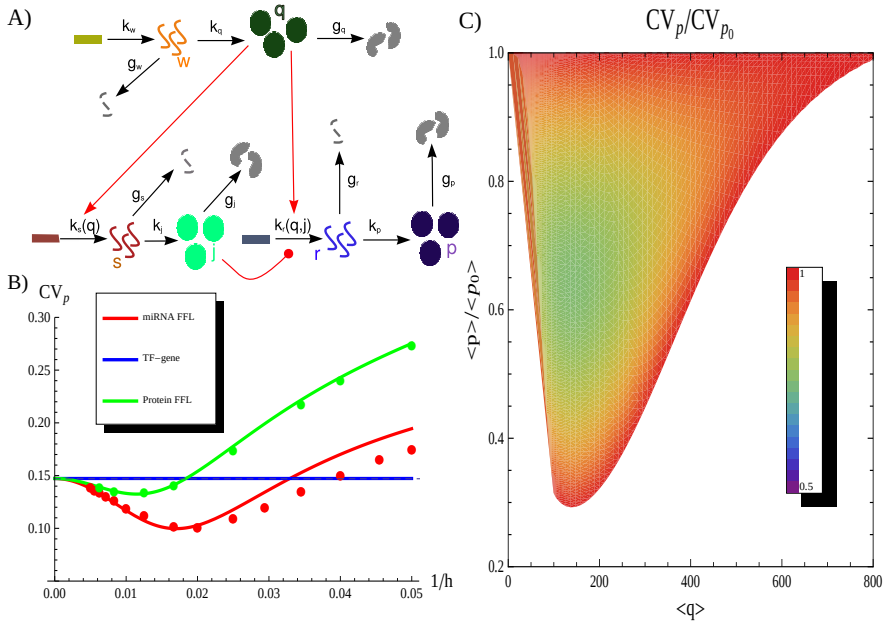

Figure 2.8: Comparison with a purely transcriptional incoherent FFL. - (A) Detailed scheme of a purely transcriptional incoherent FFL. (B) The coefficient of variation of the target protein CV_p as a function of the repression strength $1/h$ for a miRNA-mediated FFL and for its transcriptional counterpart. Thanks to the constraints imposed on parameters we can directly compare their noise-buffering efficiency with respect to a gene only activated by a TF. Both circuitries lead to a CV_p curve with a minimum for an intermediate repression strength, but the miRNA-mediated circuit appears more efficient in filtering out fluctuations. The values of parameters kept constant are the same of Figure 3. Dots are the result of Gillespie simulations with the full nonlinear dynamics while continuous lines are analytical predictions. Also in this case analytical solutions fit nicely with simulation results. (C) The noise reduction CV_p/CV_{p_0}, achieved with a purely transcriptional incoherent FFL, evaluated for different mean level of the TF $< q >$ and different degrees of reduction of the target protein level $< p > / < p_0 >$. The parameter values and the color code are the same of Figure 2.7 so as to allow a direct comparison.

of the concentrations of both regulators, more specifically by a product of Hill functions:

$$k_r(q,j) = k_r \frac{q^c}{h_r^c + q^c} \frac{1}{1 + (\frac{j}{h_j})^c}. \tag{2.17}$$

While the linearization of the Hill function $k_s(q)$ is the one presented in Equations 2.6, the linearization of $k_r(q,j)$ can be written as:

$$\begin{aligned} k_r(q,j) &\sim k_r(q,j)|_{<q>,<j>} + \partial_q k_r(q,j)|_{<q>,<j>}(q - < q >) \\ &+ \partial_j k_r(q,j)|_{<q>,<j>}(j - < j >). \end{aligned} \tag{2.18}$$

Therefore, we can define:

$$\begin{aligned} k_r^0 &= k_r(q,j)|_{<q>,<j>} - \partial_j k_r(q,j)|_{<q>,<j>} < j > -\partial_q k_r(q)|_{<q>,<j>} < q > \\ k_r^1 &= \partial_q k_r(q,j)|_{<q>,<j>} \\ k_r^2 &= \partial_j k_r(q,j)|_{<q>,<j>}. \end{aligned} \tag{2.19}$$

Using the linearization just defined and the moment generating function method previously described, the analytical expressions of $< p >$ and CV_p can be obtained.
In order to make an unbiased comparison of the noise properties of transcriptional and miRNA-mediated FFLs, the corresponding models must be constrained to produce the same amount of target proteins. The constraint can be introduced in several ways. This is due to the fact that there are two additional free parameters in the transcriptional FFL (k_j and g_j) and thus two supplementary degrees of freedom. In fact, a constraint can be inserted for example in the Hill function of target activation (tuning h_j) or in the rate of s transcription $k_s(q)$, choosing accordingly the values of k_j and g_j. This variety of options introduces some arbitrariness in the comparison. Although there is no unambiguous way to put this constraint, a reasonable choice is to keep the shared parameters to the same values (i.e repression/activation efficiencies and production/degradation rates) and choose the two additional ones to make the amount of repressor proteins j in the transcriptional case equal to the amount of miRNAs s in the mixed circuit. With this choice, we end up with the same average amount of repressors in the two circuit versions ($< j >=< s >$), with the same efficiency of repression ($h_j = h_p$), and thus with the same impact on the target expression (making equal the amount of target proteins produced by the two circuits). The target noise CV_p can now be evaluated as a function of the repression strength ($1/h$) for both circuits (Figure 2.8B).
Even though the transcriptional version can potentially reduce target fluctutions, the buffering efficiency appears clearly increased by the use of miRNAs as regulators. Furthermore, a comparison of Figure 2.8C and Figure 2.7B underlines that a miRNA-mediated FFL can buffer fluctuations over a wider range of conditions as well as to a greater extent. This is mainly due to the additional step

of translation required for the production of proteins j, which unavoidably adds noise to the system.
We would like to emphasize that the shown efficiency in noise reduction, achieved with the transcriptional FFL, should be considered as an upper bound. In fact, the constraints imposed on k_j and g_j limit the translational burst size, i.e. the average number of proteins produced from a single mRNA, and this parameter crucially influences the intrinsic fluctuation amplitude of proteins j (2). According to the constraint, in the transcriptional FFL the rate of translation k_j must simply equal the rate of degradation g_j (assuming $g_j = g_q = g_p$). As a result, the average number of proteins j which is produced from a single mRNA is forced to $b = k_j/g_s = g_j/g_s$, where b is the translational burst size. As reported in (23), the fluctuations in the concentration of a single gene product can be expressed as:

$$CV^2 = \frac{1}{< p >}\left(\frac{b}{1+\eta} + 1\right). \tag{2.20}$$

Therefore, the noise level depends on the translational burst size (where η is the ratio of protein to mRNA lifetime). We report the parameter values used in the analysis summarized in Figure 2.8: $k_j = g_j = g_q = g_p = 0.002; g_w = g_s = g_r = 0.006, k_r = 0.8; k_s = 0.5; c = 2; h_r = 200; h_s = 200; k_p = 0.04$. For these values, the translational burst size b, compatible with constraints, is $b = 0.33$, while in eukaryotes it is expected to be larger (certainly larger than one) because of the long average half-life of messenger RNAs, compared to the time required for one translation round (85). This is why a miRNA-mediated FFL can actually outperform its purely transcriptional counterpart as a noise buffer even more than reported in Figure 2.8.
Moreover, some peculiarities (not currently included in our model) of the mixed-motif can further explain why it can be better suited for noise buffering. Firstly, fluctuations in RNA polymerase and ribosome abundance are possible sources of extrinsic noise in gene expression (17). These fluctuations are expected to influence unspecifically the rate of transcription and translation respectively of each gene. In a miRNA-mediated FFL the correlation between target mRNA and miRNA production, which is crucial for noise reduction, is robust to these additional sources of noise as mRNAs and miRNAs are identically affected only by global RNA polymerase fluctuations. On the other hand, in purely transcriptional FFLs the number of repressor proteins j is exposed to the independent fluctuations in ribosome concentration, so the regulator-regulated correlation can be compromised with potentially negative consequences on the circuit's noise reduction efficiency. Secondly, delays in the action of regulators (miRNA or proteins) in the indirect pathway from the master TF to the target can damage the noise buffering function (see section 2.11). However the biogenesis of miRNAs is thougth to be faster than the one of proteins, and thus miRNAs may affect the target expression with a shorter delay with respect to factors regulating nuclear events like a TF (86). This feature should enable miRNAs to produce rapid responses, as required to counteract fluctuations. Finally, the presence of a nucleus makes the eukaryotic cell a two-compartment system with stochastic transport channels, with consequences on gene expression noise (87). In fact, transcriptional regulation requires an additional transport step with respect to

the post-transcriptional one. In a transcriptional FFL, the repressor protein (replacing the miRNA) must return into the nucleus to act on the target. This again potentially reduces the correlation of its fluctuations with the target ones, affecting the noise buffering ability.

2.7 Alternative modes of microRNA regulation

2.7.1 MiRNA-mediated promotion of mRNA degradation

As discussed in section 1.3, by base pairing to mRNAs, miRNAs can mediate translational repression or mRNA degradation (47; 55; 88). So far we considered a repressive action on target translation, making the rate of mRNA translation a nonlinear decreasing function of the number of miRNAs. In this section, the validity of our results will be proven even in the case of a miRNA repression based on promotion of target mRNA degradation. In this case, the miRNA action can be introduced adding to the basal rate of mRNA degradation g_r (in absence of miRNAs) a term that increases with the copy number of miRNAs:

$$g_r(s) = g_r + \frac{g_{max} s^c}{h_{deg}^c + s^c}, \tag{2.21}$$

where g_{max} represents the maximum possible increase of the degradation rate (if $s \to \infty, g_r(s) \to g_r + g_{max}$); h_{deg} is the dissociation constant of miRNA-mRNA interaction; c is the Hill coefficient.
The stochastic model built on this assumption cannot be solved with the same strategy explained in section 2.2. The closure of the equations for $< p >$ and σ_p would require further linearizations. However, we ran simulations for this alternative mechanism of miRNA-mediated promotion of target mRNA degradation to check the robustness of our results. Strikingly enough, these simulations can be fit quite well with the analytical predictions based on the assumption of a miRNA-mediated repression of translation.

2.7.1.1 Modeling

We present only the deterministic equations for the FFL, since the open circuit and the TF-gene analysis can be easily performed following the same steps. The mean field description of the system in Figure 2.9 is:

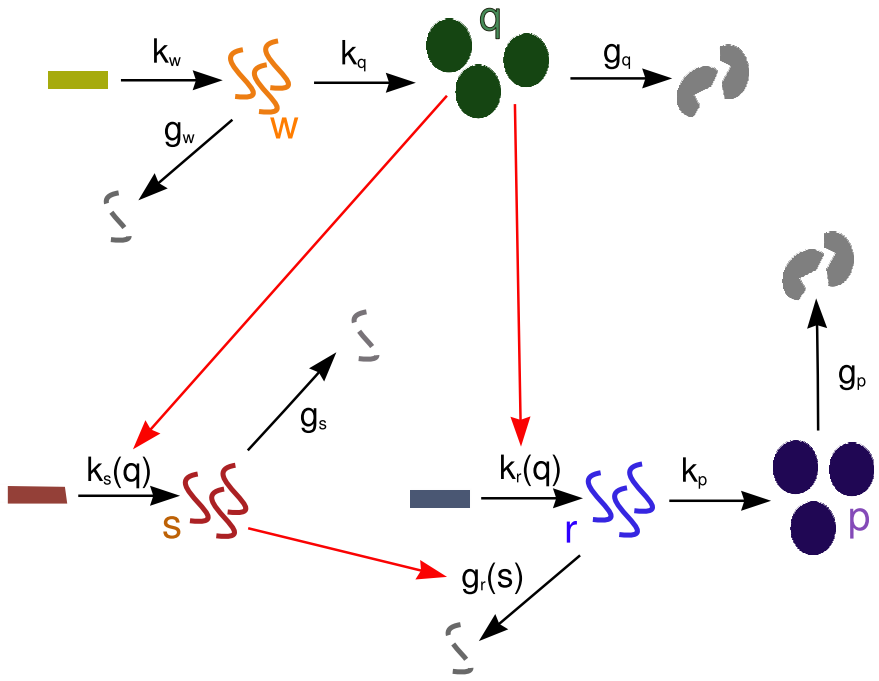

Figure 2.9: Scheme of a miRNA-mediated incoherent FFL, where the miRNA repressive function is a promotion of mRNA degradation. - The notation is the same of Figure 2.1. The red arrow starting from s represents the regulation of the rate of degradation $g_r(s)$, which is a nonlinear increasing function of miRNA concentration.

$$\begin{aligned}
\frac{dw}{dt} &= k_w - g_w w \\
\frac{dq}{dt} &= k_q w - g_q q \\
\frac{ds}{dt} &= k_s(q) - g_s s \\
\frac{dr}{dt} &= k_r(q) - g_r(s) r \\
\frac{dp}{dt} &= k_p r - g_p p, \qquad (2.22)
\end{aligned}$$

where $k_s(q)$ and $k_r(q)$ are the Hill functions of activation shown in Equations 2.1, while the form of $g_r(s)$ is shown in Equation 2.21.
Assuming $c = 2$, the steady state expressions of w_{ss}, q_{ss}, s_{ss} are the same of Equations 2.13, as nothing is changed in their dynamics. On the other hand, the expressions of r_{ss} and p_{ss} become:

$$\begin{aligned}
r_{ss} &= \frac{k_q^2 k_r k_w^2 (g_q^4 g_s^2 g_w^4 h_{deg}^2 h_s^4 + 2 g_q^2 g_s^2 g_w^2 h_{deg}^2 h_s^2 k_q^2 k_w^2}{(g_q^2 g_w^2 h_r^2 + k_q^2 k_w^2)(g_q^4 g_r g_s^2 g_w^4 h_{deg}^2 h_s^4 + 2 g_q^2 g_r g_s^2 g_w^2 h_{deg}^2 h_s^2 k_q^2 k_w^2} \cdot \cdot \\
&\quad \frac{+ k_q^4 (g_s^2 h_{deg}^2 + k_s^2) k_w^4)}{+ k_q^4 (g_r g_s^2 h_{deg}^2 + (g_r + g_{max}) k_s^2) k_w^4)} \\
p_{ss} &= r_{ss} k_p / g_p. \qquad (2.23)
\end{aligned}$$

2.7.1.2 Comparison with miRNA-mediated repression of mRNA translation

In order to compare in an unbiased way the noise properties of the mixed FFL with different mechanisms of miRNA action, we set up the parameters of the two alternative models (Figure 2.2A' and 2.9) so as to achieve the same final level of the target protein p_{ss}. This can be obtained by choosing the same parameters for the two models except those involved in the miRNA regulation. These may then be fixed by equating the values of p_{ss} in Equations 2.13 and 2.23. As Figure 2.10 shows, the result of this comparison is that a mixed FFL with a degradation-based repression presents essentially the same noise properties of the corresponding circuit with a translation-based repression. In particular, Figure 2.10A presents the analogous (for the present repression scheme) of the histograms of Figure 2.3C and 2.4C. The noise buffering effect is clearly visible, thus suggesting that the inchoerent FFL loop performs equally well its noise buffering function with either type of repression mechanism. Superimposing the distributions with mean

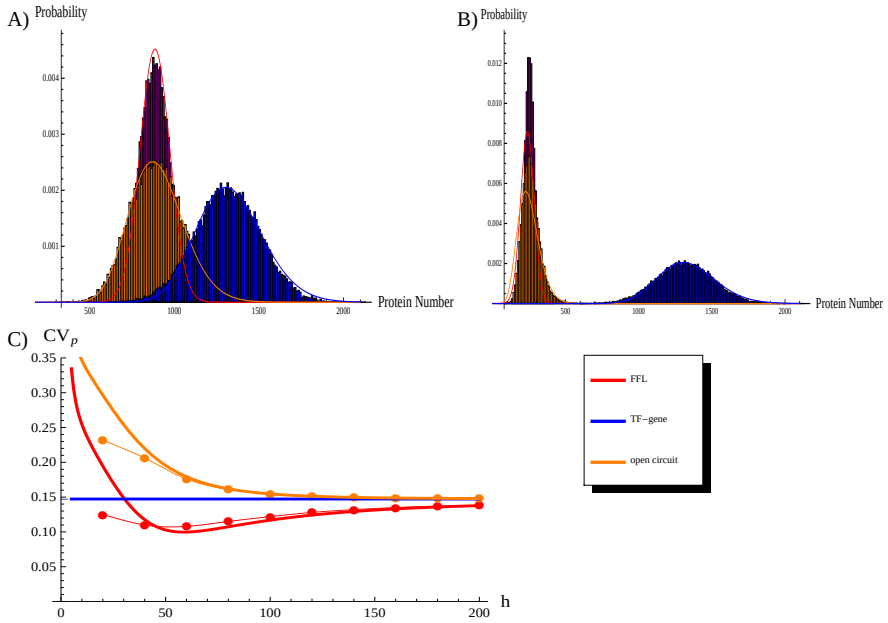

Figure 2.10: The noise attenuation achieved by a FFL with miRNA-mediated promotion of mRNA degradation. - (A) The probability distribution of target protein number for the three circuits in analysis. Histograms are the result of Gillespie simulations with the nonlinear dynamics depicted in Figure 2.9. Continuous lines are empirical distributions (gaussian for the FFL and gamma for the TF-gene and the open circuit) with mean and variance predicted by the analytical stochastic model shown in section 2.2. The parameter values are those explained in caption of Figure 2.3. Even if the analytical model is built on the hypothesis of repression of mRNA-translation, it fits equally well the distributions resulting from simulations based on miRNA-mediated promotion of mRNA degradation. (B) Same histograms of A with a stronger repression ($h = 20$, all other parameters as stated before). In the regime of strong repression, the analytical model tends to overestimate the variance σ_p. (C) The coefficient of variation of the target protein CV_p as a function of the inverse of the repression strength (h) for the three circuits. The figure shows the presence of an optimal repression strength even in the case of a degradation-based miRNA repression.

and variance calculated analytically for the miRNA-mediated repression of translation, we find again a very good agreement, apart from a slight disagreement in the strong repression regime (small h). This can be clearly seen in Figure 2.10, where dots are the results of Gillespie simulations with the hypothesis of a miRNA-mediated promotion of mRNA degradation, while thick lines are analytical predictions of the translation-based model. Apart from the mentioned overestimation in the strong repression region, the model fits quite well Gillespie simulations. In conclusion, all the results presented hold despite the mechanism of miRNA repression and even if the analytical predictions are based on the assumption of a miRNA-mediated repression on mRNA translation, they can be applied more generally.

2.7.2 Stoichiometric mechanism of repression

Regulatory small non-coding RNAs (sRNAs) play a crucial role also in prokaryotes gene regulation. In particular, the class of *trans*-acting sRNAs has many features in common with miRNAs in eukaryotes: most of them bind to the UTR of the target mRNAs through base-pairing (often imperfect) recognition to prevent their translation or to promote their degradation. However, as discussed in (79), unlike their eukaryotic counterpart they usually act stoichiometrically on their targets, since a given sRNA molecule is often degraded along with its target, instead of being used to regulate other targets. Different authors (78; 79; 80) studied the peculiar features of this noncatalytic sRNA-mediated regulation, developing a simple kinetic model for sRNA gene silencing.
In this section, the noise buffering properties of incoherent FFLs will be studied assuming a stoichiometric modality of repression. A detailed comparison with the previously discussed catalytic case will be presented.

2.7.3 Modeling

The scheme of a mixed FFL in which the coupling between sRNAs (s) and mRNAs (r) is stoichiometric is depicted in Fig.2.11. Following (78; 79; 80), we assume that both the sRNA and the mRNA are co-degraded when paired, with a rate that depends on the sRNA-mRNA interaction strength k_{rs}. The mean field kinetics of our system can be described by the equations:

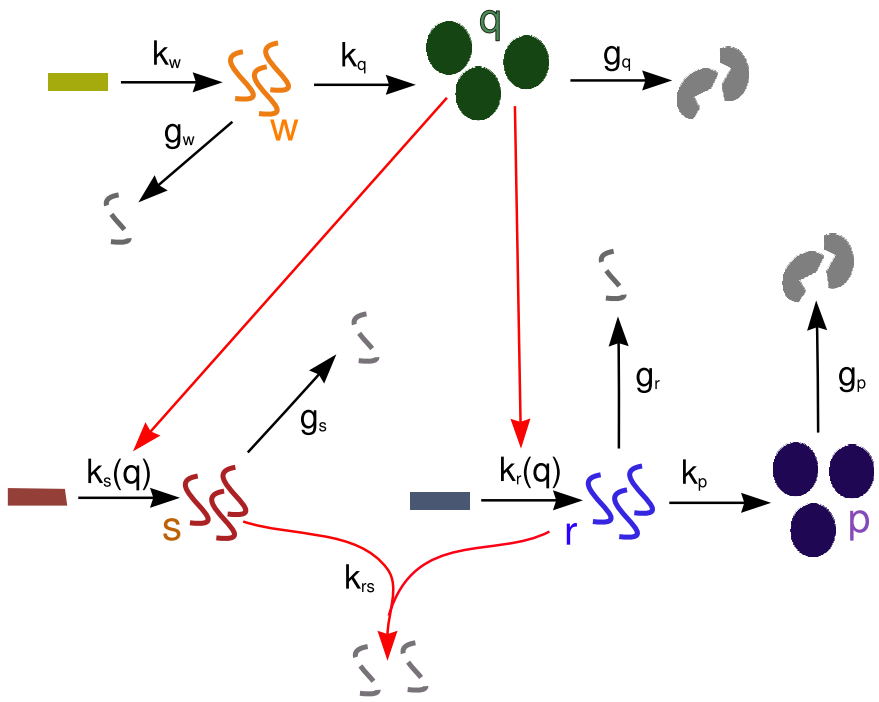

Figure 2.11: Scheme of a miRNA-mediated incoherent FFL, where the miRNA-mRNA pairing exposes both molecules to co-degradation. - The coupled degradation of the miRNA-mRNA pair is described through a second-order kinetic constant k_{rs} .

$$\begin{aligned}
\frac{dw}{dt} &= k_w - g_w w \\
\frac{dq}{dt} &= k_q w - g_q q \\
\frac{ds}{dt} &= k_s(q) - g_s s - k_{rs} r s \\
\frac{dr}{dt} &= k_r(q) - g_r r - k_{rs} r s \\
\frac{dp}{dt} &= k_p r - g_p p.
\end{aligned} \tag{2.24}$$

The stationary solutions ($d_t x_i = 0 \; \forall i \in \{w, q, s, r, p\}$) can be easily calculated.

2.7.4 Comparison with catalytic repression

This section explores the consequences of the nature of sRNA-mRNA interaction (stoichiometric or catalytic) on the noise properties of the mixed FFL. In analogy to section 2.7.1.2, we shall compare the two models choosing the parameters so as to obtain the same steady-state protein number p_{ss} with both types of sRNA action. To this aim, the production and degradation rates of each molecular species are set equal (compare the schemes in Figure 2.2A' and Figure 2.11). The relation between k_{rs} in the stoichiometric model and h in the catalytic model is then calculated by equating the expression for p_{ss} in Equations 2.13 and in steady-state solutions of Equations 2.24.
We ran simulations for the FFL, TF-gene direct regulation and the open circuit for both catalytic and stoichiometric miRNA actions (Figure 2.12). The noise filtering effect is robust with respect to the mechanism of miRNA-mRNA interaction, but a catalytic interaction makes the FFL more efficient in buffering fluctuations (compare the histograms in Figure 2.12A and B). The U-shaped profile of the noise strength CV_p is recovered also in the stoichiometric case. Figure 2.12C and D report the CV_p as a function of the inverse of repression strength for two different sets of parameter values. The maximum of attenuation is achieved for approximately the same value of h of the catalytic case, but a stoichiometric repression lead to a weaker noise reduction. In conclusion, even if the qualitative behaviour is the same, the catalytic modality seems more efficient for a noise control.

As discussed in (78) and (89), in the stoichiometric model described by Equations 2.24 the mean protein number exhibits a threshold linear behaviour as a function of the ratio k_r/k_s, with the threshold in 1 (78). Following (79), protein expression can be classified into three regimes: repressed ($k_r/k_s \ll 1$), crossover ($k_r/k_s \approx 1$) and expressing ($k_r/k_s \gg 1$). A threshold linear behaviour implies ultrasensitivity in the crossover regime. As a consequence, the noise is enhanced

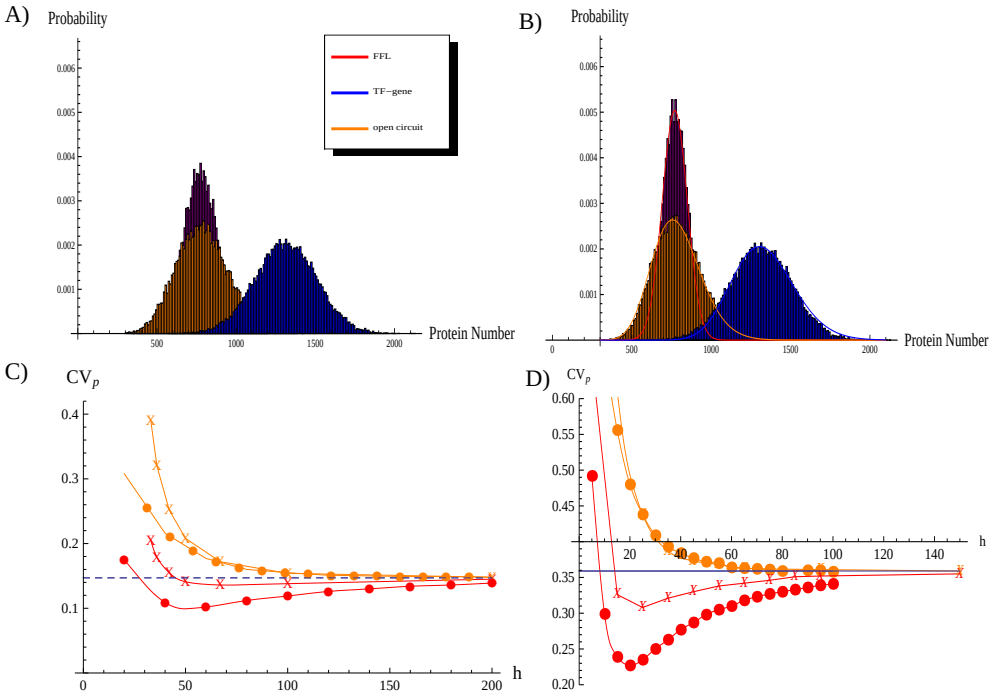

Figure 2.12: Attenuation of noise by a FFL mediated by sRNAs that act stoichiometrically on their mRNA targets. - The the upper part of the Figure shows the probability distributions of the target protein number for the three circuits in the case of stoichiometric action (A) and catalytic action (B). Although in both cases the FFL reduces relative fluctuations with respect to the direct TF regulation and the open circuit, the catalytic modality turns out to be more efficient than the stoichiometric one. For the same parameter set, we report in (C) CV_p as a function of the inverse of the repression strength h. In the stoichiometric case CV_p is actually a function of k_{rs}, which however can be expressed as function of h. To allow a simpler comparison of the various plots, we depicted the stoichiometric results directly as function of h. Dots are the result of simulations based on the hypothesis of a catalytic sRNA action while the x-shaped points derive from simulations with a stoichiometric action. For each regulatory modality the FFL and the open circuit data (which can be recognized because they are always higher than the FFL ones) are reported. (D) Same as (C) but for an alternative set of parameters.

near the threshold due to critical fluctuations ((79) and references therein). However, this threshold-linear response is expected if the mRNA-sRNA interaction is strong, while for a weaker repression the threshold smoothly disappears (see Figure 2.13) and the three regimes become indistinguishable. The analysis presented in Figure 2.12 shows that the attenuation of fluctuations by a mixed FFL can be observed for a weak repression, corresponding to $k_{rs} \sim 10^{-4}$ in Figure 2.13. In this case, the crossover regime is vanishing and the raise in fluctuations at $k_r/k_s = 1$ is negligible.

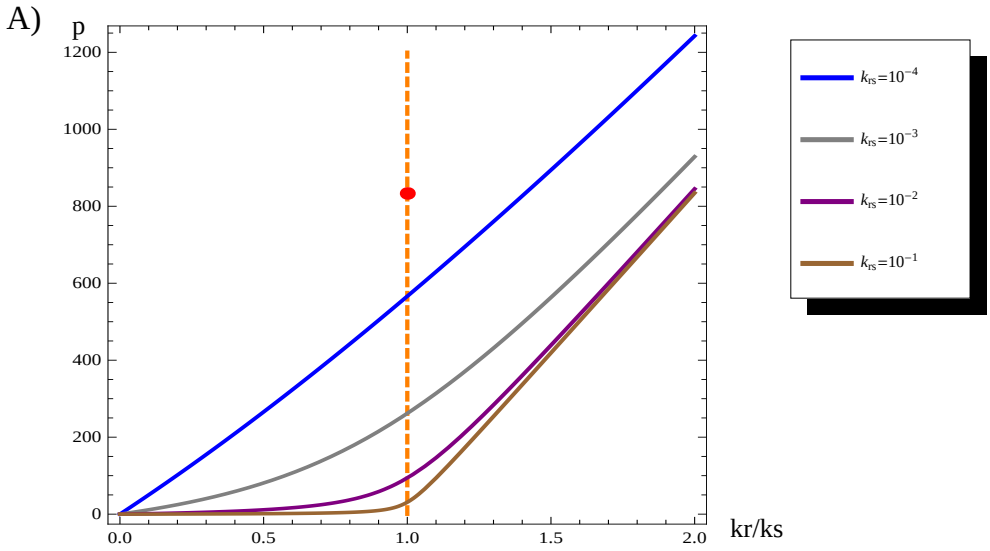

Figure 2.13: Threshold linear behaviour - The response (p) of the FFL is plotted as a function of k_r/k_s for different values of the miRNA-mRNA interaction strength. The red dot represents the protein production in absence of miRNA regulation. The threshold-linear response is evident only for strong repression, while for $k_{rs} = 10^{-4}$, compatible with a fine tuning regulation, the response is almost purely linear.

2.8 Robustness of results

2.8.1 Constraints imposed on the FFL by the requirement of sensitivity to changes in the master TF concentration.

Functional FFLs can be defined as those in which a change in the master TF concentration can cause a change in the concentration of target proteins and miRNAs (90). While the issue of the trade-off between sensitivity to signals and noise control is discussed in detail in section 2.10, in the following we shall define the general conditions that ensure a sufficient dependence of miRNA and target levels on the TF concentration. Noise buffering is functional in the presence of noise propagation, that requires a target dependence on TF concentration. In our context, this dependence implies that the Hill functions of activation by the TF and of repression by the miRNA should not be saturated. In fact, in conditions of complete saturation, signals and fluctuations cannot propagate from the master TF to the target (even in absence of miRNA regulation), therefore a noise control lose any functionality. On the other hand, in the unsaturated regime a change in the number of TFs can alter in a significant way the number of target proteins in the cell together with the number of miRNAs, generating the correlated fluctuations needed for noise buffering. If the TF concentration is too high (with respect to h_r and h_s), the target and miRNA transcription rates become insensitive to variations in TF concentration (unless they are so large that can escape from the region of saturation) limiting the FFL sensitivity to upstream signals. The same considerations hold for the target repression. If there are too many miRNAs (with respect to h), the target expression is drastically shut down and again the system becomes insensitive to changes in the number of TFs. Accordingly, we excluded from our analysis the parameter sets for which: $< q > \ll h_r(h_s)$ or $< q > \gg h_r(h_s)$ and $< s > \ll h$ or $< s > \gg h$. In other words, the circuit functionality imposes that concentrations of regulators must be placed not far from the linear region of the corresponding Hill functions. A high sensitivity corresponds also to an overexposure to noise, in fact noise amplification and sensitivity are correlated quantities (26; 93) (see also section 2.10). Since the aim of our study is to prove the noise buffering role of miRNA-mediated incoherent FFLs, considering the parameter space that strongly exposes to noise makes clearly sense and it seems not a limitation.
With the conditon of unsaturated regulations satisfied, the qualitative results presented apply for virtually all parameter choices. As a partial proof, in the next two sections we shall discuss a few different combinations of parameters. As we shall see, our results turn out to be remarkably robust with respect to changes in the allowed (unsaturated) region of parameters.

2.8.2 Target and miRNA genes differentially expressed

In this section we present the target noise strength for the three circuits as a function of the ratio between the maximum rate of transcription of the miRNA gene (k_s) and the target gene (k_r), keeping fixed the TF concentration ($< q >$) and miRNA repression strength ($1/h$). The aim is to show that the noise buffering role of the mixed FFL shows only a weak dependence on the characteristics of miRNA and target promoters. The upper part of Figure 2.14 shows the target noise strength as a function of k_r/k_s. In the whole range of values, the mixed FFL shows the largest noise reduction effect, and in particular the noise buffering role of the FFL does not require an equal rate of miRNA and mRNA transcription. In fact, the noise attenuation is due to the correlation of fluctuations in the number of mRNAs and miRNAs and not to their absolute values. Different maximum rates of transcription (k_r and k_s) only change the height of peaks in mRNA and miRNA trajectories, without affecting their correlation.

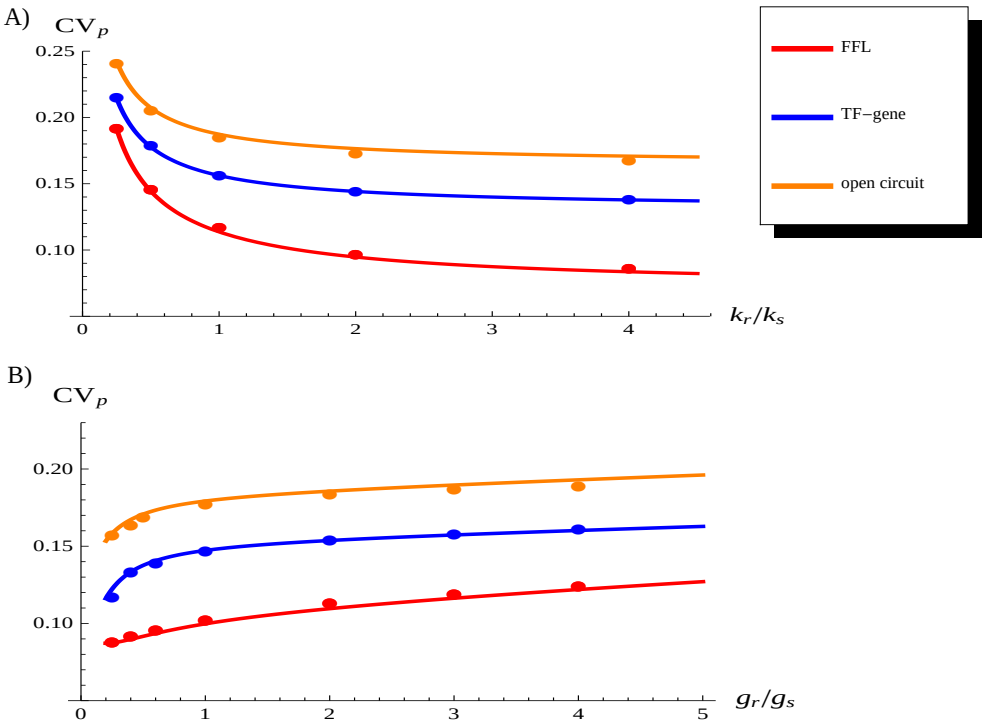

Figure 2.14: Robustness of results. - (A) CV_p as a function of the transcription rate ratio k_r/k_s. (B) CV_p as a function of g_r/g_s.

2.8.3 mRNAs and miRNAs with different stability

Another important robustness test is the dependence of the FFL noise buffering efficiency on the ratio of decay constants g_r/g_s. In principle, one could expect a reduction in the FFL efficiency when $g_r \neq g_s$. In fact, with different values of g_r and g_s the mRNA and miRNA trajectories could start to fluctuate out of phase due to different relaxation times. To answer this question we calculated the CV_p for the three circuits as a function of the ratio g_r/g_s. The results are reported in the lower part of Figure 2.14. As in the previous case, in the whole range of g_r/g_s that we studied the mixed FFL gives the largest noise reduction effect. These two tests together show that noise buffering is a generic feature of mixed FFLs. There is no need to fine tune the half-life and/or the transcription rate of miRNAs and mRNAs to obtain a mixed FFL that efficiently reduces fluctuations.

2.8.4 Optimal TF concentration tuning k_w instead of k_q

We previously discussed the dependence of the noise strength CV_p on the copy number of TFs present at the steady state (Figure 2.6C). The parameter chosen to tune $< q >$ was the rate of translation k_q. For the sake of completeness, the same plot obtained by varying k_w instead of k_q is reported in Figure 2.15. Also with this alternative protocol, the FFL outperforms the other circuits in noise control for an intermediate TF concentration. This is a further proof of the robustness of our results.

2.8.5 Results for another set of parameters

As a final test of robustness, we solved the master equations for the three circuits with a choice of input parameters (reported in the caption of Figure 2.16A) leading to sizeable fluctuations in the number of master TFs ($CV_q \sim 0.4$ for $h = 30$). This should be compared with the values of the case dicussed previoulsy (whose parameter set is reported in the caption of Figure 2.3) for which the noise in the number of TFs was only $CV_q \sim 0.17$. Also in this case the TF fluctuations are efficiently attenuated by the FFL, leading to a target noise strength of $CV_p = 0.25$, to be compared with $CV_p = 0.38$ for the direct TF-gene regulation and $CV_p = 0.46$ for the open circuit (see the histograms in Figure 2.16A). These values agree with the observation that the noise attenuation effect due to the FFL circuit becomes larger and larger as the size of TF fluctuations increases. The U-shaped profile of CV_p for the FFL steps out also for this parameter set, further supporting the idea that this property does not depend on their particular choice but is a generic feature of the model (see Figure 2.16B).

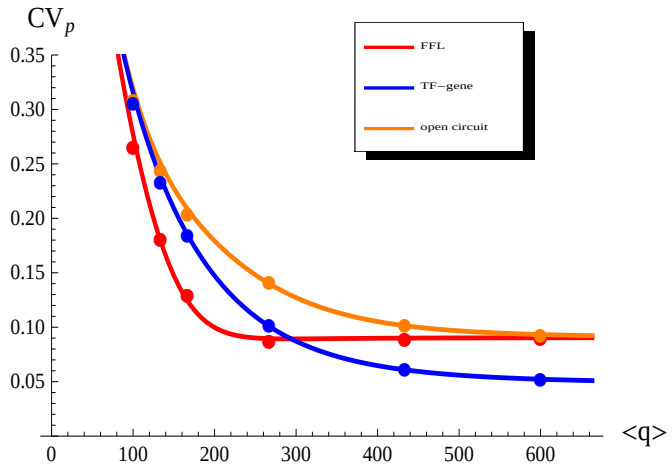

Figure 2.15: Robustness of results 2. - CV_p as a function of the mean number of TFs $< q >$. Dots are the result of Gillespie simulations with the full dynamics while continuous lines are the analytical predictions. The parameter values are the same of Figure 2.3

2.8.6 Testing the effect of Hill function linearization

Besides the robustness against the choice of input parameters another important issue which one would like to address is the effect of the linearization of the Hill functions. This can be easily achieved by comparing analytical versus numerical (Gillespie) results. Since this is the only approximation that we made in our analysis, it is important to understand which is the range of parameters in which we can trust our analytical results not only qualitatively but also quantitatively. It is easy to guess that the linear approximation should give sizeable errors only when the fluctuations in the variables become large enough to cover a wide portion of the Hill functions, thus exploring also their non-linear part. A good example to discuss this issue is given by the set of input parameters discussed in the previous section. In this case, even if the analytical solution still captures qualitatively the main features of the systems, it is less precise in its quantitative predictions. This is clearly visible in Figure 2.16B, where analytical predictions are compared with the results of Gillespie simulations (which keep into account the full non-linear dynamics of the FFL). While for the optimal value of h discussed above ($h = 30$) the agreement is very good, as h decreases the gap between the two curves becomes larger and larger. This is a consequence of the linearization of Hill functions and shows that if fluctuations are too large, as it happens in the strong repression regime, the linear approximation may become too crude. It is interesting to study how the approximation breaks down since it is a typical example of the subtle effects which the two step nature of gene expression may have

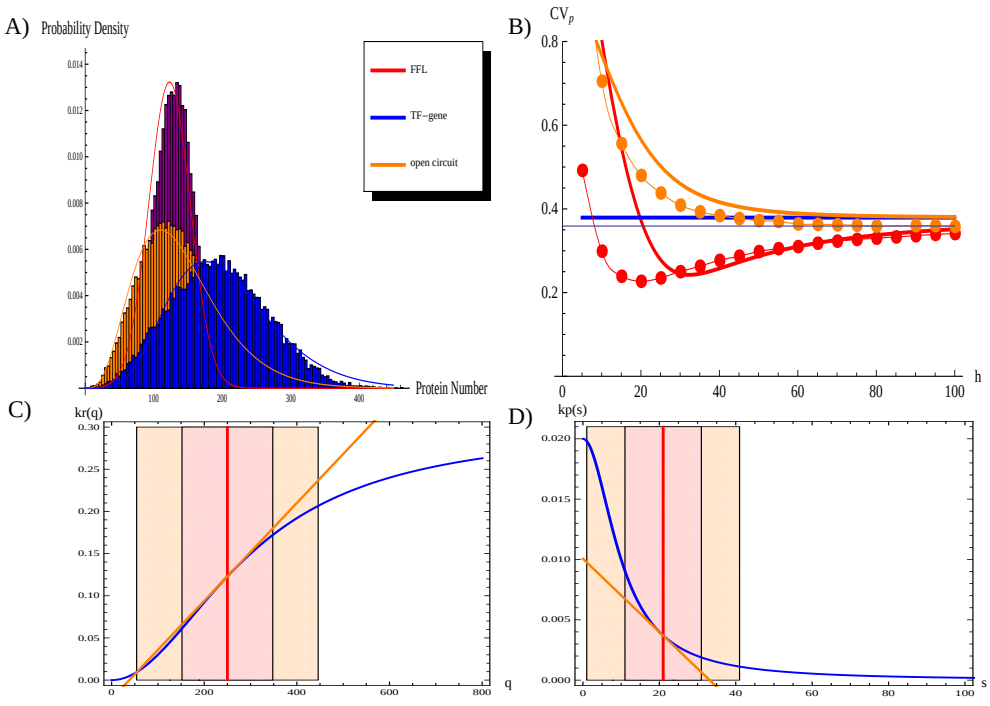

Figure 2.16: Robustness of results 3. - (A) The probability distribution of target protein number for the three circuits in analysis. The parameter values are: $k_w = 0.01s^{-1}$; $k_q = 0.3s^{-1}$; $g_w = g_r = g_s = 0.006s^{-1}$; $g_q = g_p = 0.002s^{-1}$; $k_r = k_s = 0.3s^{-1}$; ; $h_s = h_r = 200$; $c = 2$; $h = 30$. (B) The coefficient of variation of the target protein CV_p as a function of the inverse of repression strength h for the three circuits. (C) The Hill function of transcriptional activation of the target gene (blue line). The red line represents the mean number of TFs $< q >$ at equilibrium, while the shaded region corresponds to intervals $[q - \sigma_q, q + \sigma_q]$ and $[q - 2\sigma_q, q + 2\sigma_q]$. The orange line is the linearized function used for the analytical solution. (D) The Hill function of translational repression of the target gene (blue line) by the miRNA, in the strong repression region ($h = 10$). The red line is the mean number of miRNAs $< s >$ at equilibrium, while the shaded region corresponds to intervals $[s - \sigma_s, s + \sigma_s]$ and $[s - 2\sigma_s, s + 2\sigma_s]$. The orange line represents the linearized function used for the analytical solution.

on noise propagation. With the choice of parameters of the Figure, the q fluctuations cover a wide region of the domain of $k_r(q)$ and $k_s(q)$ (Figure 2.16C), but the line tangent in $< q >$ still captures quite well the Hill function trend, with only a slight overestimation ($< r >=< s >= 20$ from simulations, compared to the predicted value of 21). On the other hand, the large fluctuations in s ($CV_s = 0.48$) make the linearization of $k_p(s)$ a poor approximation (Figure 2.16D). The s distribution spreads on a domain region where the Hill function widely changes its curvature, therefore the tangent line introduces in many trajectories a sizeable underestimation of the rate of target translation. As a result we have $< p >= 43$ from simulations while only $< p >= 28$ from the analytical model. In a similar way also the standard deviation turns out to be uncorrectly estimated by the analytical solution. These disagreements explain the displacement of analytical curves in Figure 2.16B with respect to simulations. This example shows however that, despite its quantitative failure, the analytical model describes fairly well the qualitative behaviour of the system even in the presence of large fluctuations and, as mentioned above, it becomes more and more precise when fluctuations around steady state values cover a domain where the Hill functions are approximately linear (which is the usual assumption in literature).

2.9 Cross-talk between microRNA targets

A recent study pointed out that the action of a miRNA on a specific target gene expression is affected by the total number of miRNA targets and their mRNA abundance (91), a phenomenon called "dilution effect". This is presumably a consequence of target competition for a finite intracellular pool of miRNAs. In particular, the degree of downregulation of an individual target expression is generally reduced by the presence of other transcribed target genes. A similar cross-talk between targets has been previously shown for sRNA regulation in bacteria (78) both theoretically and experimentally. Therefore, the functionality of a genetic circuit that involves miRNA regulations, as the ones studied in this thesis, can be influenced by the expression level of miRNA targets not embedded in the circuit. To address this issue, we evaluate in this section the impact of an additional miRNA target independently transcribed (a situation depicted in Figure 2.17A) on the circuit ability to buffer noise.

2.9.1 Stoichiometric versus catalytic models of miRNA action

The model used so far for miRNA regulation was based on the hypothesis of perfectly catalytic action. The rate of translation of target mRNAs was assumed to be a nonlinear decreasing function of miRNA concentration, neglecting the details of mRNA-miRNA physical coupling with the implicit assumption that the downregulation process does not affect the available miRNA pool. A perfectly catalytic action does not predict any competition effect between multiple targets at equilibrium, since each target can only sense the available number of miRNAs without altering it. On the other hand, a stoichiometric model has been proposed in the context of sRNA regulation in bacteria (78; 79; 80), in which each sRNA can pair with one messenger leading to mutual degradation. In this latter case the expression of a secondary target can capture a significant portion of the sRNAs, with a resulting decrease in the average repression acting on the first target. The nature of miRNA regulation is presumably somewhere in between these two extreme possibilities, although usually generically referred to as catalytic. In this view, in order to address the possible repercussions of target cross-talk on miRNA-mediated FFLs, we consider a deterministic model introduced previously in (78) (here straightforwardly applied to the FFL case), in which the physical coupling of miRNAs and target mRNAs and the catalytic/stoichiometric nature of this coupling is explicitely taken into account. More specifically, the kinetic equations for miRNA-mediated FFLs can be written as:

$$\begin{aligned}
\frac{dw}{dt} &= k_w - g_w w \\
\frac{dq}{dt} &= k_q w - g_q q \\
\frac{ds}{dt} &= k_s(q) - g_s s - (k_+ rs - k_- c) + (1-\alpha)\beta c \\
\frac{dr}{dt} &= k_r(q) - g_r r - (k_+ rs - k_- c) \\
\frac{dc}{dt} &= (k_+ rs - k_- c) - \beta c \\
\frac{dp}{dt} &= k_p r - g_p p,
\end{aligned} \tag{2.25}$$

where c is the concentration of miRNA-mRNA complexes, k_+ is the probability of miRNA-mRNA association, and k_- is the probability of dissociation of the complex c, which can degrade with rate β. The parameter α represents the probability that a mRNA degradation event in the complex is accompained by degradation of the coupled miRNA. As discussed in (78), it is a measure of how much the miRNA action is catalytic. In this section, the variables that describes the state of the system ($\{w, q, r, s, c, p\}$) are continuous variables representing the average number of the various molecular species (we omitted the notation $< .. >$ for averages in Equations 2.25 for brevity). Since we are interested in steady state properties we can simplify the model equilibrating the c complex dynamics:

$$\begin{aligned}
\frac{dw}{dt} &= k_w - g_w w \\
\frac{dq}{dt} &= k_q w - g_q q \\
\frac{ds}{dt} &= k_s(q) - g_s s - \alpha\gamma rs \\
\frac{dr}{dt} &= k_r(q) - g_r r - \gamma rs \\
\frac{dp}{dt} &= k_p r - g_p p,
\end{aligned} \tag{2.26}$$

where $\gamma = \beta k_+/(k_- + \beta)$. The limit of $\alpha = 0$ implies that for each degradation event of c complexes, none of the miRNAs is lost. This corresponds to a perfectly catalytic mode of action. It is indeed a simplification of the model presented in section 2.7.1, in which the rate of mRNA degradation is a linear function of miRNA concentration instead of a nonlinear Hill function. The opposite situation of $\alpha = 1$ reproduces the stoichiometric model described by Equations 2.24 (apart from the sostitution $\gamma \to k_{rs}$).

This description can be generalized to the case of two miRNA targets as in Figure 2.17A, adding an equation describing the dynamics of a second target which is independently transcribed. Restoring the notation for averages, the effective equations describing the dynamics of the mean number of proteins $< p >$, miRNAs $< s >$, mRNAs $< r >$ of the target in the FFL and mRNAs $< r_2 >$ of the secondary miRNA target are:

$$\begin{aligned}
\frac{d< s >}{dt} &= k_s(< q >) - g_s < s > -\alpha\ (\gamma_1 < r >< s > + \gamma_2\ < r_2 >< s >) \\
\frac{d< r >}{dt} &= k_r(< q >) - g_r < r > -\gamma_1 < r >< s > \\
\frac{d< r_2 >}{dt} &= k_{r_2} - g_{r_2} < r_2 > -\gamma_2 < r_2 >< s > \\
\frac{d< p >}{dt} &= k_p < r > -g_p < p >,
\end{aligned} \tag{2.27}$$

where γ_1 and γ_2 are parameters linked to the probability of miRNA-mRNA coupling for the target in the FFL and the secondary target respectively. The probability α that a degradation event of a mRNA is accompanied by the degradation of the miRNA is assumed to be equal for both targets. The corresponding stochastic model, of which Equations 2.27 describe the mean-field limit, cannot be solved analytically starting from the master equation, therefore noise properties will be examined in the following with simulations only.

2.9.2 Dilution effect

In the first place, we evaluate the dependence of the target protein downregulation on the rate of expression of the secondary target, starting from the model described by Equations 2.27. The dilution effect is shown in Figure 2.17B for different values of α: the downregulation exerted on the FFL target results generally dependent on the rate of expression of the secondary target, in line with the observed inverse correlation between target abundance and mean downregulation in higher eukaryotes (91) and in bacteria (78). Similar results can be obtained by varying the coupling constant γ_2 with respect to γ_1 (as reported in (78)), while we are here assuming equal coupling probability $\gamma_1 = \gamma_2 = \gamma$. Therefore, the noise buffering function and the optimality criteria discussed in previous sections could be compromised in the presence of many or highly transcribed independent miRNA targets. This issue will be addressed in details in the following section. As expected, a perfectly catalytic mode does not feel the effect of secondary mRNA targets (red line in Figure 2.17B), while the stoichiometric mechanism is the most sensitive (green line in Figure 2.17B). In the limit of infinite out-of-circuit target expression, the joint target protein level approaches its constitutive value if $\alpha > 0$, while remains constant in the ideal case of perfectly catalytic miRNA repression. This result suggests that a catalytic mode (at least approximately), like the miRNA one, can allow a larger proliferation of the number of targets while limiting the effects of their cross-talk.

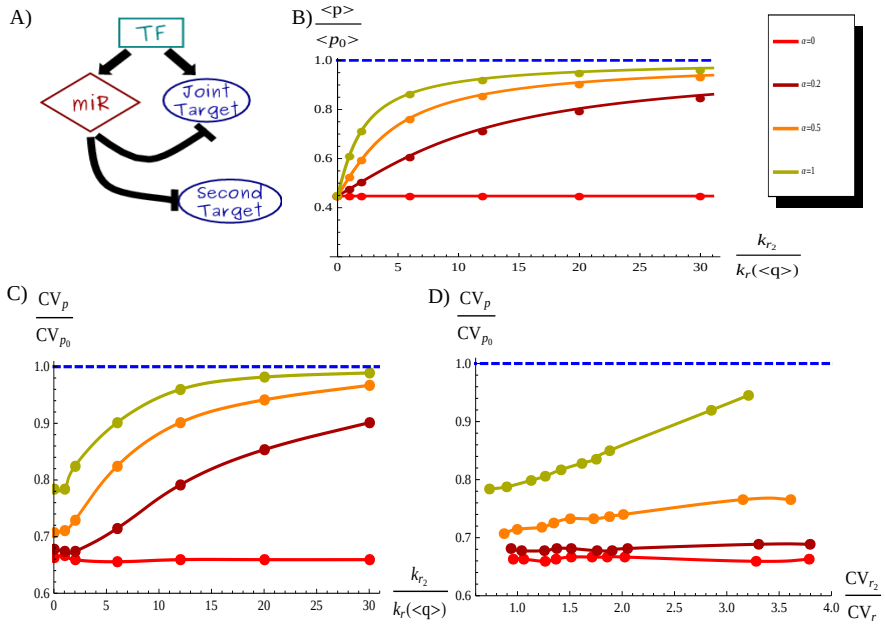

Figure 2.17: Effects of cross-talk between miRNA targets. - (A) Scheme of a miRNA-mediated FFL with an additional independently transcribed target gene (second target). (B) The degree of protein downregulation $< p > / < p_0 >$ is depicted as a function of the ratio of effective transcription rates of the secondary target (k_{r_2}) and of the FFL joint target ($k_r(< q >)$), for different values of α. Continuous lines are analytical solutions of the deterministic model (Equations 2.27), while dots are the result of stochastic simulations. (C) The noise reduction CV_p/CV_{p_0} is evaluated in the same $k_{r_2}/k_r(< q >)$ range of Figure 2.17B. Dots are the result of Gillespie simulations while continuous lines come from trivial interpolation. (D) The noise reduction is evaluated as a function of the out-of-circuit mRNA fluctuations CV_{r_2}, relative to the joint target fluctuations CV_r. Dots are the result of Gillespie simulations, simply interpolated with continuous lines.

2.9.3 Consequences of dilution effect and secondary target fluctuations on noise buffering

Since a high level of expression of secondary targets can determine a decrease of the average downregulation, it can potentially reduce the FFL ability to filter out target fluctuations. In fact, also the noise reduction CV_p/CV_{p_0} (where CV_{p_0} is the constitutive noise in absence of miRNAs) is a function of the additional target expression, as shown in Figure 2.17C. As the expression of the out-of-circuit target increases, its messengers are able to capture more and more miRNAs and the efficiency in noise reduction is gradually compromised. Finally, the FFL target fluctuations CV_p approach the constitutive ones CV_{p_0} when the messengers of the FFL target become a small fraction of the total miRNA targets. The robustness of the circuit functioning with respect to the dilution effect is again dependent on the repression mode (that changes with α). Moreover, as previously discussed, different modes (stoichiometric/catalytic) of miRNA action have a different potential in reducing fluctuations: even in absence of secondary targets, where models with different α have been constrained to produce the same amount of target protein, the noise buffering efficiency decreases with α (Figure 2.17C). This observation highligths that the level of miRNA ability to avoid mutual degradation while targeting a mRNA can play a role in the optimization of fluctuation counteracting, besides conferring stability with respect to target cross-talk.
While the corruption of the noise-buffering ability seems mainly due to the increase in the mean level of secondary messengers, there is another more subtle cause that gives a contribution: the uncorrelated fluctuations of secondary messengers. Since the secondary target is independently transcribed (not under the control of the master TF activating the miRNA gene) its fluctuations are expected to be completely uncorrelated with the miRNA ones, implying a random sequestration of miRNAs. To disentagle this contribution from the merely dilution effect, we studied the case of a secondary target transcribed at the same effective rate of the FFL target, but with different levels of fluctuations (see Figure 2.17D). In the case of equal transcription rates the dilution effect has a negligible impact on the noise buffering activity of the circuit (see Figure 2.17C), nevertheless the level of noise reduction (CV_p/CV_{p_0}) is progressively reduced as the second target concentration becomes more and more noisy, as reported in Figure 2.17D. This effect seems especially relevant for a hypothetically stoichiometric miRNA repression. Therefore, the noise level of additional targets is a variable that must be taken into account in evaluating the cross-talk effect on the noise-buffering efficiency of the circuit. Although the FFLs are overrepresented in the mixed network (60; 61; 62; 63), a single microRNAs can downregulate hundreds of target genes and consequently not every target is expected to be under the control of the same TF regulating the miRNA gene. Therefore, even though most motif function analysis are carried out looking at the motif operating in isolation, we have shown that the presence of additional miRNA targets in the network can alter the functioning of a miRNA-mediated motif. In fact, the efficiency of miRNA-mediated FFLs as noise controllers should be considered context-dependent. While this circuit seems properly designed to filter out fluctuations when the miRNA-target interaction is specific or secondary targets

are poorly transcribed, cell types or conditions that require a high expression of out-of-circuit miRNA targets can significantly corrupt this circuit property. Besides the understanding of the function of endogenous miRNA-mediated FFLs, this analysis of target cross-talk effects can be a useful warning for the growing field of synthetic biology (92): the implementation of genetic circuits incorporating small RNA regulations for specific scopes must take into account the sRNA specificity and the level of expression (and fluctuations) of eventual other targets.

2.9.4 Details on the model setting

The solution for $< p >$ of Equations 2.27 at the steady state depends on α. Therefore, in order to evaluate the impact of the dilution effect for different mechanisms (stoichiometric/catalytic) of miRNA repression, we choose for each α the corresponding γ value that leads to the same mean amount of target proteins $< p >$ at equilibrium. Qualitatively in a catalytic model ($\alpha = 0$) the miRNAs are more efficient since they can affect several target mRNAs without being degraded. Consequently, as α decreases the γ value must be decreased so as to mantain the same target level expression. This is the constraint that makes unique the starting point (for $k_{r_2} = 0$) of the curves corresponding to different α values in Figure 2.17B.
Moreover, in Figure 2.17B and C the transcription of the second target is modeled as an independent birth-death process with birth rate k_{r_2}, while the rate of transcription of the joint target is a function of the TF concentration. Therefore, to compare the transcription rates, the effective mean rate $k_r(< q >)$ is taken as a reference (where $< q >$ is constant as we are not tuning the TF concentration). In fact, $k_r(< q >)$ represents the average rate at which the joint target is transcribed.
Concerning the other parameter values, in Figure 2.17 they are fixed to: $k_q = 0.19s^{-1}; g_q = g_p = 0.002s^{-1}; g_w = g_r = g_s = g_{r_2} = 0.006s^{-1}; k_w = 0.0126s^{-1}; k_r = k_s = 0.8s^{-1}; k_p = 0.04s^{-1}; c = 2; h_r = h_s = 200$, while the value of $\gamma = 0.00011s^{-1}$ is assigned to the catalytic model ($\alpha = 0$) and it corresponds approximately to the optimal buffering value. The γ values for the other α models can be calculated as described above.
Finally, in Figure 2.17D, the fluctuations of the second target mRNAs are modulated considering its rate of transcription as a function of an independent TF and changing the TF noise. The modulating strategy is in perfect analogy with the one explained in section 2.3.3. In brief, the transcrition of the second target is activated by an independent TF q'. Its rates of transcription k'_w and translation k'_q are chosen so as to produce the same mean amount of proteins of the other activator ($< q >=< q' >$). Therefore, the effective mean rates of transcription of both miRNA targets turn out to be equal. Changing the ratio k'_w/k'_q , while keeping costant the product $k'_w k'_q$, allows us to vary the second target fluctuations without altering its mean level.

2.10 Noise reduction and signaling sensitivity

Biological systems present the apparently contraddictory need for both high sensitivity to external signals both homeostatic controls, depending on the specific function in analysis. In fact, while one essential feature of signal transduction systems is the amplification of small changes in input signals (93), the reliable celullar functioning in a fluctuating environment lays on multiple homeostatic controls (the most evident is temperature control in mammals). Similarly, at the level of genetic networks there is an interplay between sensitivity to changes in the input signal and the ability to buffer stochastic fluctuations. An increase in sensitivity to a signal results in an elevated exposure to its fluctuations, as it has been shown for linear cascades of regulations (93; 94). More recently, the sensitivity/noise-buffering analysis has been extended to small genetic circuits, including feedback and feedforward loops (26). The working hypothesis of the authors is that the main function of a genetic circuit is to maximize the amplification of input signals. We argue that while this can be often the case, some circuits can have evolved to mantain reliably a functional steady state, even at the expense of a loss of sensitivity (and even thanks to that loss), to implement in other words a homeostatic control.

Following (26), the steady state sensitivity can be defined as the relative response in output that follows a change in the input. In the context of incoherent FFLs (scheme in Fig. 2.2A'), we can consider as input the mean number of TFs $< q >$ and as output the consequent level of target proteins $< p >$. Following these definitions, the susceptibility takes the form:

$$susceptibility = \frac{< q >}{< p >} \frac{d\ < p >}{d\ < q >} = \frac{d\ ln(< p >)}{d\ ln(< q >)}. \tag{2.28}$$

As a measure of the quantity of noise propagating through the circuit, the noise amplification measure η can be introduced (26):

$$\eta = \frac{CV_p}{CV_q}, \tag{2.29}$$

defined as ratio between output and input noise. As Figure 2.18 A shows, the incoherent miRNA-mediated FFL presents an interplay between noise amplification and susceptibility very similar to that of a gene only activated by a TF, while the same fine-tuning implemented using an independent miRNA would imply a more severe interplay. Therefore, the noise buffering function demonstrated for FFLs is achieved at the expense of steady-state sensitivity. Given a fixed value of susceptibility, the FFL and the TF-gene linear circuit lead to a similar degree of noise amplification, while when the noise is buffered by the FFL there must be a loss of target susceptibility. In fact, the fold change in target expression, that follows a change in the TF mean level $< q >$, is reduced precisely in the region where the noise control is implemented (see Figure 2.18 B). However, we propose that this is precisely the behaviour needed for a homeostatic control. The output is highly sensitive to changes in the input concentration until a finely tuned steady state is reached, then this functional steady state is kept robust to

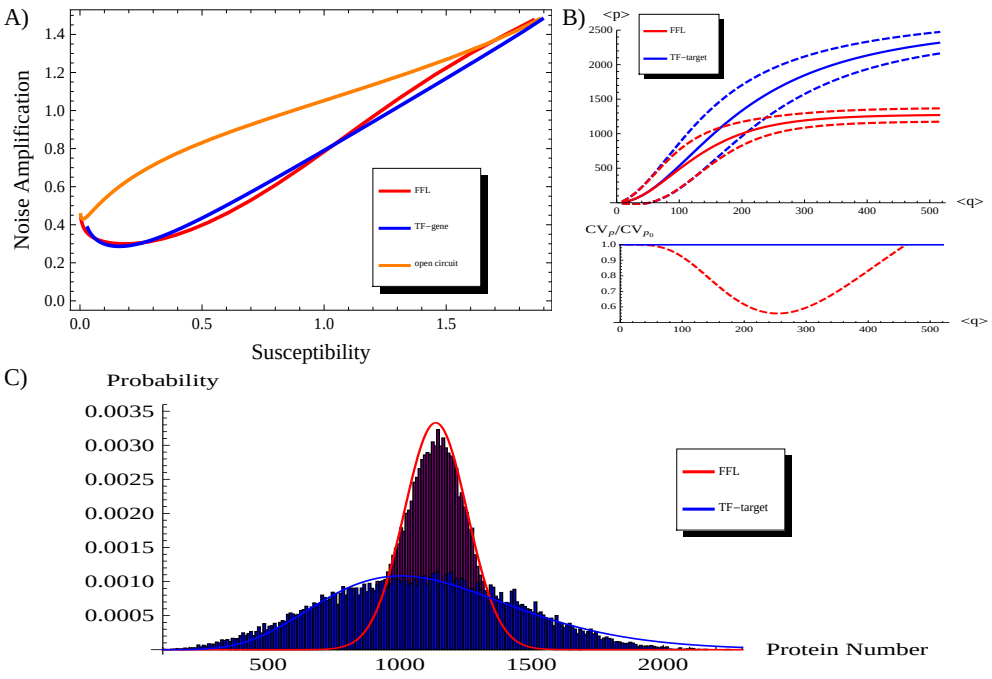

Figure 2.18: Noise-sensitivity analysis. - (A) Noise amplification versus susceptibility for the three circuits: a miRNA-mediated incoherent FFL, a TF-target regulation and an open circuit. The parameter values that are fixed are those reported in caption of Figure 2.3 (unless $k_q = 0.19s^{-1}$ higher than in Figure 2.3 to increase the TF fluctuations). (B) The upper panel shows the fold change in the target level in response to a fold change in the TF level for the miRNA-mediated incoherent FFL and the TF-target linear circuit. Continuous lines represent the behaviour of mean values while dashed lines are depicted at a distance from the mean equal to one standard deviation. In the lower panel the noise reduction CV_p/CV_{p0} is depicted in same range of $< q >$.(C) The probability distribution of protein number for the two circuits (miRNA-mediated FFL and TF-gene). In this case the two regulative circuits are constrained to produce an equal mean amount of target proteins. The same steady state is achieved with a strikingly different control of fluctuations by the two circuitries.

input fluctuations, even if at the expense of a sensitivity loss. The same steady state could be reached more simply without any miRNA regulation, tuning the TF concentration in a TF-gene circuit, so as to conserve a high sensitivity. However, in this case the equilibrium level would be affected by strong fluctuations propagating from the upstream factor, as clearly shown in Figure 2.18 B.
In conclusion, if sensitivity is the function that have to be maximized, as it is probably the case in signaling systems, incoherent FFLs (and miRNA mediated ones) are outperformed by other circuits (like those making use of positive feedbacks loops (26)) that support less noise amplification at a fixed susceptibility. However, in different biological contexts a high sensitivity could be important only until a functional steady state is reached. Then a homeostatic control can be required for keeping the reached level constant in the presence of noisy upstream regulators and miRNA-mediated FFLs seems properly designed to this aim. The proposed functioning is also in agreement with the idea of fine-tuning: when the target expression is switched on by a rise in TF concentration, the maintenance of its level into a narrow functional range can be more important than a reliable transmission of further incoming small signals. A role of miRNA regulation in homeostasis is in line with the observation that miRNAs are often involved in signaling networks to ensure homeostatic controls (see for example (64)).

2.11 Effects of possible delays in miRNA production.

The common lore is that a RNA based post-transcriptional regulation can have a faster action on a target gene expression with respect to TF regulation (86; 92). In fact, a TF must be transported back to the nucleus and find its target promoter to exert its regulative role. However, there is a lack of data to support quantitatively this assumption and the biogenesis of miRNAs actually requires several processing steps. The time needed for the miRNA to be processed, loaded in RISC and in general to become active can introduce a delay between its transcription and its effect on targets. Therefore, it could be interesting to consider possible effects of this time delay on the noise buffering function of mixed FFLs. While in the model presented in section 2.2 the miRNA is supposed to act on its targets instantaneously, this section presents results of simulations performed taking into account the time-delay that can arise from miRNA processing. More specifically, the time-delay has been inserted in the Hill function of regulation of the target translation, mimicking the time required for miRNA activation. With this simulation procedure, for each chosen set of parameter values, it is possible to establish the threshold in delay time below which the circuit is able to reduce target fluctuations.
In Figure 2.19 the noise reduction achieved with a FFL (CV_p/CV_{p_0}) is reported as a function of the time required for miRNA activation. The time-delay is expressed in unit of protein half-life, chosen as a reference since it represents the longest

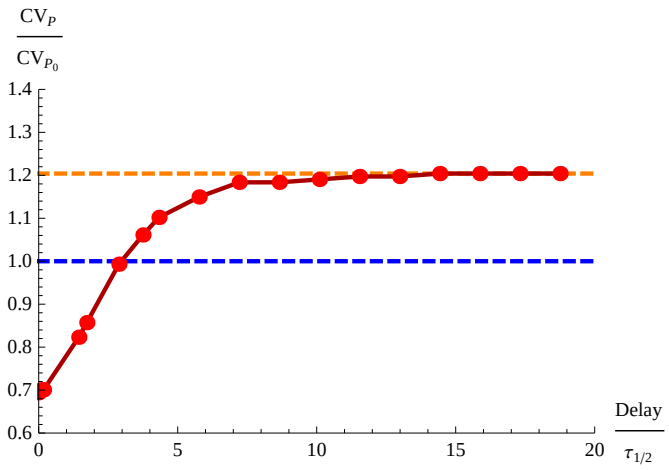

Figure 2.19: The effect on noise-buffering efficiency of a time-delay between miRNA transcription and miRNA repressive action. - The values of parameters are the ones in caption of Figure 2.3. The target noise reduction CV_p/CV_{p_0} is measured as a function of the length of the time-delay expressed in unit of protein half-life $\tau_{1/2}$. CV_{p_0} is the constitutive noise of a TF-gene circuit without the miRNA regulation. Until the delay length approaches approximately 3 protein half-lives, the FFL is still able to filter out fluctuations. After that the noise level tends to the value achieved with an open circuit (dashed orange line) in which miRNAs and target mRNAs have uncorrelated fluctuations in their level. Dots are the result of Gillespie's simulations with the full nonlinear dynamics.

time-scale in the system. The ability of the circuit to filter out fluctuations relies on the correlation between miRNA and target mRNA fluctuations, therefore an eventual time-delay in miRNA action can negatively affect the noise buffering. More specifically, with the parameter values of Figure 2.3, the incoherent FFL is no more able to reduce target fluctuations if the delay is longer than approximately 3 protein half-lives (see Figure 2.19). As the processing time becomes longer and longer, miRNA fluctuations lose any correlation with the target ones and the target noise approches the value corresponding to the open circuit case (dashed orange line in Figure 2.19).
In conclusion, we showed that a significant time-delay between miRNA transcription and target repression can compromise the noise-buffering function. When quantitative measures of the time required for transport and processing of miRNAs and proteins will be available, it will be possible to precisely evaluate the degree of reduction of target fluctuations, inserting the appropriate delays in the Hill functions of regulation of our theoretical model (even if at the expense of its analytical tractability).

2.12 Bioinformatical analysis of miRNAs involved in FFLs in the human mixed network.

Although miRNA-mediated FFLs have been shown to be overrepresented in real mixed networks with respect to randomized networks (60; 61; 62; 63), it is equally important to establish the numerical fraction of miRNAs and miRNA targets that are actually involved in these circuits, to better highlight the effective biological relevance of miRNA-mediated FFLs. To this aim, we take advantage of the genome-wide survey of human miRNA-mediated FFLs previously developed by our group (63), based on a search for overrepresented motif in human and mouse promoters and 3'-UTRs. Of the 464 miRNAs annotated as KNOWN-KNOWN in the Ensembl database (Release 46) (95), using the filters and the software setup of (63) 193 were selected to form the post-transcriptional network (miRNA-target interactions). Integrating this network with the transcriptional one (with TF-target and TF-miRNA interactions), 133 miRNAs have been significantly associated to at least one (usually more than one) of the 5030 mixed FFLs found in the human regulatory network (see (63) for more details). Therefore, miRNAs, at least in the database considered, seem often involved in FFL circuits. Since each miRNA can regulate hundreds of targets, it is also interesting to evaluate what fraction of its targets are part of FFLs. The results of this analysis are reported in the following Table, where the total number of targets and the number of targets in a FFL is presented for each miRNA embedded in a FFL . While some miRNAs preferentially regulate genes through a FFL topology, this is clearly not a general trend, further confirming the importance of considering the possible cross-talk between miRNA targets .
However, it is important to note that the proposed results suffer some limitations. Firstly, incoherent and coherent FFLs cannot be distinguished, since sequence analysis allows the identification of putative interactions but cannot establish

if they are positive or negative. Secondly, the proposed regulations should be considered as potential interactions, because they represent purely bioinformatic predictions and furthermore the miRNA and its targets could be expressed preferentially in different tissues or at different times. In this case, the eventual cross-talk would be limited among co-expressed targets. In spite of the reported limitations, the data presented here underline that miRNA-mediated FFLs can actually represent an often exploited regulative circuitry, further suggesting their importance in real networks of gene regulations.
To partially overcome the discussed limitations of an "ab-initio" analysis, we developed a database of FFLs that have Myc as a master TF, in which all the regulatory links are confirmed experimentally, thus probabily excluding many false positives. Although we will not discuss the details of the database construction in this thesis, the main result is that there is evidence of over-representation of the FFL topology also in this set of exclusively verified interactions.

miRNA gene	Num. of targets	Num. of targets in FFLs	Percentage
hsa-miR-129	44	36	81.8 %
hsa-miR-148b	127	84	66.1 %
hsa-miR-149	55	36	65.5 %
hsa-miR-449b	55	34	61.8 %
hsa-let-7a	83	51	61.4 %
hsa-miR-199a*	138	84	60.9 %
hsa-miR-125b	150	90	60.0 %
hsa-miR-199a	41	24	58.5 %
hsa-miR-101	105	61	58.1 %
hsa-miR-205	38	22	57.9 %
hsa-miR-31	35	20	57.1 %
hsa-miR-203	51	29	56.9 %
hsa-miR-30c	155	87	56.1 %
hsa-miR-425-3p	50	28	56.0 %
hsa-miR-9	106	59	55.7 %
hsa-miR-296	69	38	55.1 %
hsa-miR-194	90	49	54.4 %
hsa-miR-181d	120	64	53.3 %
hsa-miR-219	123	65	52.8 %
hsa-miR-32	148	78	52.7 %
hsa-miR-9*	100	52	52.0 %
hsa-miR-148a	91	47	51.6 %
hsa-miR-24	107	54	50.5 %
hsa-miR-133b	40	20	50.0 %
hsa-miR-499	40	20	50.0 %
hsa-miR-30a-3p	48	23	47.9 %
hsa-miR-218	83	39	47.0 %
hsa-miR-375	113	53	46.9 %
hsa-miR-223	145	67	46.2 %
hsa-miR-100	46	21	45.7 %
hsa-miR-214	62	28	45.2 %
hsa-miR-10a	39	17	43.6 %
hsa-miR-1	46	20	43.5 %
hsa-miR-130a	127	55	43.3 %

miRNA gene	Num. of targets	Num. of targets in FFLs	Percentage
hsa-miR-30a-5p	155	67	43.2 %
hsa-miR-802	76	31	40.8 %
hsa-miR-26a	129	52	40.3 %
hsa-miR-23a	152	60	39.5 %
hsa-miR-99a	46	18	39.1 %
hsa-miR-126*	181	70	38.7 %
hsa-miR-330	50	19	38.0 %
hsa-miR-135b	103	39	37.9 %
hsa-miR-133a	40	15	37.5 %
hsa-miR-155	100	37	37.0 %
hsa-miR-126	109	40	36.7 %
hsa-miR-140	106	38	35.8 %
hsa-miR-506	127	45	35.4 %
hsa-miR-99b	46	16	34.8 %
hsa-miR-202	88	30	34.1 %
hsa-miR-135a	103	35	34.0 %
hsa-let-7f	83	28	33.7 %
hsa-miR-16	57	19	33.3 %
hsa-let-7d	90	29	32.2 %
hsa-let-7e	127	40	31.5 %
hsa-miR-542-3p	39	12	30.8 %
hsa-miR-206	46	14	30.4 %
hsa-miR-34b	55	16	29.1 %
hsa-miR-34c	55	16	29.1 %
hsa-miR-342	49	14	28.6 %
hsa-miR-363	84	24	28.6 %
hsa-miR-365	46	13	28.3 %
hsa-miR-27a	104	29	27.9 %
hsa-miR-29a	115	32	27.8 %
hsa-miR-19a	145	39	26.9 %
hsa-miR-152	127	34	26.8 %
hsa-miR-199b	41	11	26.8 %
hsa-miR-141	146	38	26.0 %
hsa-miR-212	58	15	25.9 %
hsa-miR-302c*	93	24	25.8 %
hsa-miR-106a	126	32	25.4 %

miRNA gene	Num. of targets	Num. of targets in FFLs	Percentage
hsa-miR-17-5p	126	32	25.4 %
hsa-miR-30e-5p	155	39	25.2 %
hsa-miR-495	123	31	25.2 %
hsa-miR-144	146	36	24.7 %
hsa-miR-7	89	22	24.7 %
hsa-miR-20b	126	31	24.6 %
hsa-miR-20a	132	32	24.2 %
hsa-miR-103	97	23	23.7 %
hsa-miR-106b	132	31	23.5 %
hsa-miR-367	111	26	23.4 %
hsa-miR-34a	43	10	23.3 %
hsa-miR-193a	112	26	23.2 %
hsa-miR-200c	143	33	23.1 %
hsa-miR-189	35	8	22.9 %
hsa-miR-93	83	19	22.9 %
hsa-miR-202*	49	11	22.4 %
hsa-miR-451	45	10	22.2 %
hsa-miR-221	50	11	22.0 %
hsa-miR-222	50	11	22.0 %
hsa-miR-138	60	13	21.7 %
hsa-miR-302b	134	29	21.6 %
hsa-miR-302c	134	29	21.6 %
hsa-miR-302d	134	29	21.6 %
hsa-miR-299-5p	108	23	21.3 %
hsa-miR-182	80	17	21.2 %
hsa-miR-142-5p	57	12	21.1 %
hsa-miR-369-3p	101	21	20.8 %
hsa-let-7b	83	17	20.5 %
hsa-miR-494	122	24	19.7 %
hsa-miR-183	92	18	19.6 %
hsa-miR-505	51	10	19.6 %
hsa-miR-377	82	16	19.5 %

miRNA gene	Num. of targets	Num. of targets in FFLs	Percentage
hsa-miR-96	133	26	19.5 %
hsa-miR-195	57	11	19.3 %
hsa-miR-497	57	11	19.3 %
hsa-miR-30e-3p	48	9	18.8 %
hsa-miR-381	165	31	18.8 %
hsa-miR-142-3p	127	23	18.1 %
hsa-miR-139	34	6	17.6 %
hsa-miR-30b	155	27	17.4 %
hsa-miR-30d	155	27	17.4 %
hsa-miR-302b*	76	13	17.1 %
hsa-miR-487b	83	14	16.9 %
hsa-miR-369-5p	90	15	16.7 %
hsa-miR-409-5p	80	13	16.2 %
hsa-miR-410	133	21	15.8 %
hsa-miR-329	93	14	15.1 %
hsa-miR-151	70	10	14.3 %
hsa-miR-412	42	6	14.3 %
hsa-miR-25	74	10	13.5 %
hsa-miR-192	45	6	13.3 %
hsa-miR-496	113	14	12.4 %
hsa-miR-153	100	9	9.0 %
hsa-miR-15a	57	5	8.8 %
hsa-miR-217	102	9	8.8 %
hsa-miR-323	57	5	8.8 %
hsa-miR-484	100	6	6.0 %
hsa-miR-26b	129	7	5.4 %
hsa-miR-146b	40	2	5.0 %
hsa-miR-200a*	90	3	3.3 %
hsa-miR-200a	146	3	2.1 %
hsa-miR-200b	143	2	1.4 %
hsa-miR-429	108	1	0.9 %

2.13 Conclusions and discussion

2.13.1 Experimental and bioinformatic evidences of the relevance of miRNA mediated FFLs in gene regulation.

Few cases of incoherent miRNA-mediated FFLs have been experimentally verified until now: a case involving c-Myc/E2F1 regulation (96) and more recently a miR-7 mediated FFL in *Drosophila* (86). Strikingly, miR-7 has been found to be essential to buffer external fluctuations, providing robustness to the eye developmental program. The fact that miR-7 is interlocked in an incoherent FFL provides a first hint that our model can be biologically relevant.
On the purely computational side, it is interesting to notice that in (63) it was shown that the typical targets of these FFLs are not randomly distributed but are instead remarkably enriched in TFs. These are the typical genes for which a control of stochastic fluctuations should be expected: the noise in a regulator expression propagates to all its targets, affecting the reliability of signal transmission in the downstream network.
Finally, a significant enrichment in oncogenes within the components of the FFLs was also observed (63). The mentioned FFL containing c-Myc/E2F1 is just an example (50). In view of the emerging idea that non-genetic heterogenetity, due to stochastic noise, contributes to tumor progression (97) and affects apoptotic signal response (98), the role of miRNA-mediated FFLs in reducing fluctuations can explain why they are often involved in cancer-related pathways.

2.13.2 Concluding remarks

The type of regulatory action which a miRNA exerts on its targets can be rather well understood looking at the degree of coexpression with the targets (43; 45; 46; 65; 67). In particular, an incoherent mixed-FFL implies a high level of miRNA-target coexpression, so it is suitable to implement a fine-tuning interaction. The target is not switched off by miRNA repression, rather its mean level is adjusted post-transcriptionally to the desired value. However, many cells can have a protein concentration far from the finely controlled mean value, if strong fluctuations are allowed. Hence, a noise buffering mechanism can be crucial at the level of single cells, and a fine-tuning interaction will be effective for a large part of the cell population only if coupled with a noise control. Some authors proposed the conjecture that the incoherent mixed-FFL can actually have a role in noise buffering (61; 65; 72) and biological evidences that miRNAs can effectively be used as expression-buffers have been recently found (72; 86). In this perspective, the miRNA-target interactions classified as neutral (67), as the mean level of the target only changes inside its functional range by the presence/absence of miRNAs, actually could have been selected by evolution to prevent potentially

harmful fluctuations. We demonstrated, through stochastic modeling and simulations, that the incoherent mixed-FFL has the right characteristics to reduce fluctuations, giving a proof to the previously proposed intuitive conjecture and supplying the lacking quantitative description. In particular, we showed that this circuit filters out the noise that is propagating from the master TF, giving robustness to the target gene expression in the presence of noisy upstream factors. Furthermore our theoretical description led to the prediction that there is a value of the miRNA repression strength for which the noise filtering is optimal. A maximum of target-noise attenuation appears likewise by varying the miRNA concentration or the TF concentration and this robust prediction could be tested experimentally. In all cases the implementation of the best noise filter does not imply a strong suppression of the target protein expression, coherently with a fine-tuning function and in agreement with the observation that the miRNA down-regulation of a target is often modest (73; 74).
To the best of our knowledge, we presented the first model explicitly built on the mixed version of the FFL. From a theoretical point of view, we addressed the detailed master equation describing the system (without neglecting the dynamics of mRNA), instead of the approximate Langevin description, and we were able to apply the moment generating function approach despite the presence of nonlinear terms that can give rise to deviant effects. This approach allowed us to take into account extrinsic fluctuations as the noise propagating from upstream genes, without an arbitrary definition of the extrinsic noise distribution. This strategy can be naturally extended to other circuits in the mixed network to test their potential role in the control of stochasticity.
Furthermore, we compared, in terms of noise buffering ability, miRNA-mediated FFLs with their purely transcriptional counterparts, where the miRNA is replaced by a protein that inhibits transcription rather than translation. This comparison shows that a miRNA regulator can be better suited for the noise buffering purpose.
Finally, we tryed to overcome the limitations in the analysis that can arise from considering a genetic circuit as operating in isolation. In this view we evaluated the impact that the recently discovered dilution effect (78; 91) can have on the noise buffering function of miRNA-mediated incoherent FFLs. More specifically we showed than an efficient noise control requires the minimization of the number of miRNA target sites on out-of-circuit genes, especially if highly expressed or strongly fluctuating in the mRNA level.
The hypothesis of a role of miRNAs in noise buffering can shed new light on peculiar characteristics of miRNA regulation. As discussed in (72) and (86), it can explain why miRNAs are often highly conserved, controlling key steps in development, but in many cases they can be deleted with little phenotypic consequences. On the evolutionary side, the origin of vertebrate complexity seems to correspond to the huge expansion of non-coding RNA inventory (including miRNAs) (99). This can suggest a further reasoning: the morphological complexity requires a high degree of signaling precision, with a strict control of stochasticity, and miRNA regulation can satisfy these requirements if embedded in an appropriate circuit, as we showed for the ubiquitous miRNA-mediated FFL.

3

Autoregulation via intronic microRNAs

3.1 Introduction to the problem

As discussed in section 1.3, miRNAs can be classified depending on their genomic location in two main classes: intergenic miRNAs and intragenic miRNAs. The vast majority of intragenic miRNAs can be found in regions annotated as introns, mostly sharing the orientation with the host gene transcript (same strand), as can be seen in Figure 3.1. Whereas intergenic miRNAs present their own promoter region with specific regulatory elements (100; 101), the same-strand intronic miRNAs are belived to be co-transcribed with their host genes (132). Evidence of strong correlation between the expression of intronic miRNAs and their corresponding host genes supports this idea (52; 53). Following the definition in (132), we will simply refer as intronic miRNAs to miRNAs that share the same promoter of their host and that are spliced out of the host transcript and processed into mature miRNAs. In other words, opposite-strand intronic miRNAs or miRNAs in exonic regions will not be considered.
It has been shown that a possible role of an intronic miRNA is to support the function of its host gene by silencing genes that are functionally antagonistic (103), or more generally to coordinate the expression of genes functionally related to the host (104). However, pieces of evidence have been reported that intronic miRNAs can directly regulate the expression of their host genes, establishing a negative feedback regulation. Instances of negative autoregulatory feedbacks by embedded miRNAs were firstly found by expression analysis (61). Subsequently, a more comprehensive study pointed out that the number of occurences of miRNA-mediated selfloops (mSLs) in the human regulatory network can be the effect of an evolutionary positive selection (105), and one of the predicted mSLs has been confirmed experimentally (105). These results suggest that mSLs are an often

exploited and presumably functional regulative circuitry and motivate the search for a complete map of mSLs potential functions that will be the topic of this part of the thesis. In particular, deterministic and stochastic models will be used to analyze the circuit dynamics. It will be shown that the mSL circuit, despite its simple topology, implements different functions: it alters the response-time of the host gene, filters fluctuations propagating from the upstream network and generates adaptation and fold change detection. This one-to-many mapping between circuit topology and functions confers to mSLs a certain degree of flexibility (106). In fact, different functions correspond to different characteristics of the molecules involved (as half-life or repression strength). These parameters can be tuned over evolutionary timescales to optimize the needed function in each specific case (107) without the need of network rewirings.
Besides the understanding of the role of endogenous mSLs, the circuit function-topology map that will be depicted would also be invaluable for synthetic biology, contributing to the draw up of the manual for how to robustly engineer a biological circuit that carry out a target function.

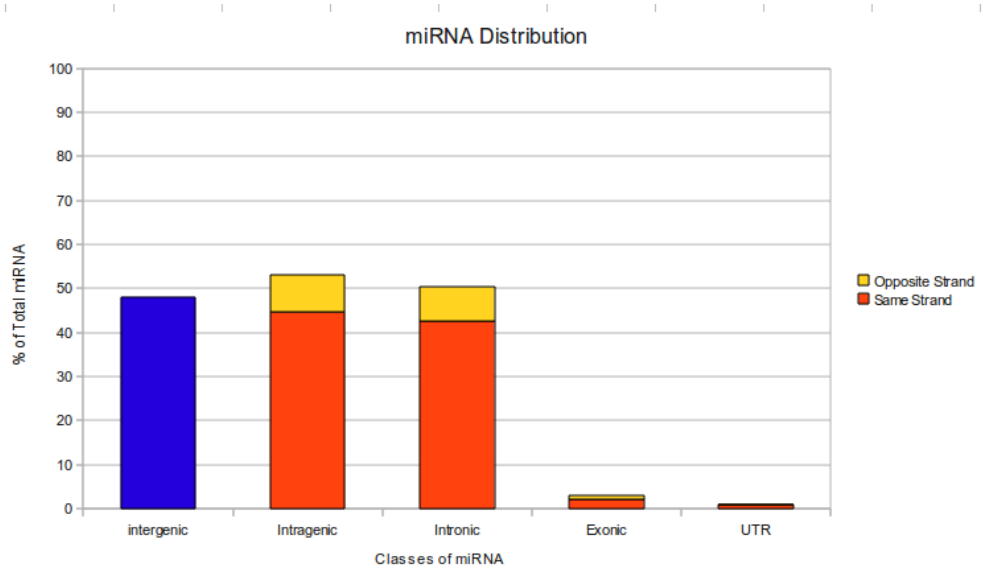

Figure 3.1: Classification of miRNA genes. - Classification based on genomic location of miRNAs. Approximately half of miRNA genes is located in intragenic regions. The vast majority is within introns, sharing the orientation with the host gene. (Data extracted from the Ensembl database - release 57).

3.2 Response time

The response of a transcription network to a stimulus is generally slow and the response time (time to reach half of the final steady state) is governed by the lifetime of the protein products (90; 108), which can be of several hours. Since the production of short-living proteins would have an high metabolic cost, an alternative solution to accelerate the response time is to implement suitable network architectures. In fact, the dynamic response of gene expression depends on the regulatory circuitry in which the gene is embedded. For example, transcriptional negative autoregulation has been proven theoretically and experimentally to speed up the rise-time of a target gene after induction (109), while more complex kinetics can be implemented using feedforward loops (90). This section presents an analysis of the kinetic properties of mSLs in the presence of a sudden change in an activating TF concentration. In particular, it will be shown that this circuit can be a response accelerator after induction and a delay element in the switch-off dynamics. Different strategies to produce the same steady-state protein concentration will be compared: a simple transcriptional unit, a negative transcriptional selfloop (tSL), in which the protein inhibits its own transcription, and autoregulation through intronic microRNAs. In order to explore each circuit ability to switch on and off protein production, the kinetics in response to both an activation of transcription both a complete inactivation will be studied.
Since our analysis is focused on a regulative circuit that is embedded in eukaryotic regulative networks, it is mandatory to take into account explicitely the mRNA dynamics. In fact, while in prokaryotes the mRNA dynamics is usually very fast (half-life of few minutes) compared to the protein one (half-life of many hours) and thus is usually safely neglected (109), in eukaryotes the situation seems more complex. In mammals, the mRNA half-life can range from minutes to about 24 hours (111; 112), with typical values in the range of 5-10 hours (113; 114). Similarly, protein lifetimes cover quite a wide range, from minutes to several days (115). While the common trend (as usually assumed) could be that of proteins more stable than the corresponding transcripts (116), the phenomenology is probably more elaborate, also thanks to the complex regulations of protein and mRNA stability. MicroRNAs are usually stable molecules with an half-life that can span days (117; 118). However, there are cases of short-living miRNAs, as many miRNAs expressed in human brain (119), and also the miRNA turnover seems widely regulated (120). This complex scenario forced us to explore different half-life regimes and it turned out that certain dynamic features are crucially dependent on the relative half-lives of the molecular players.

3.2.1 The theoretical framework

3.2.1.1 Kinetics of a simple transcriptional unit

The mean-field dynamics of a simple transcription unit activated by a TF (see Figure 3.2) can be described by the following kinetic equations:

$$\begin{aligned}
\frac{dw}{dt} &= k_w - g_w w \\
\frac{dq}{dt} &= k_q w - g_q q \\
\frac{dr}{dt} &= k_r(q) - g_r r \\
\frac{dp}{dt} &= k_p r - g_p p,
\end{aligned} \tag{3.1}$$

where (as in the previous chapter) the rate of transcription is assumed to follow a Hill function of the TF concentration:

$$k_r(q) = \frac{k_r q^c}{h_r^c + q^c}. \tag{3.2}$$

We are firstly interested in the dynamics after full induction. The signal of activation of transcription can be represented in our framework as a strong increase in the TF concentration (q). In other words, in an initial condition $x_i(t = 0) = 0$ for each molecular species x_i, the target promoter is exposed to full activation ($q/h_r >> 1$). Therefore, at $t = 0$ the target gene starts to be transcribed at the maximum rate k_r. In this conditions the dynamics of the TF can be neglected as the net effect of its raise is the saturation of the Hill function: $k_r(q) \sim k_r$. How the steady-state ($x_{i_{ss}}$) is approached by the various molecular species can be easily calculated:

$$\begin{aligned}
\frac{r(t)}{r_{ss}} &= (1 - e^{-g_r t}) \\
\frac{p(t)}{p_{ss}} &= \frac{g_p(1 - e^{-g_r t}) - g_r(1 - e^{-g_p t})}{g_p - g_r} \\
\frac{p(t)}{p_{ss}} &= 1 - e^{-gt}(1 - gt) \quad if\ g_r = g_p = g.
\end{aligned} \tag{3.3}$$

The response-time t_r is given by $p(t_r)/p_{ss} = 0.5$. Since t_r does not depend on the steady-state value (only on half-lives), the transcriptional unit can be used as a null model for comparison with more complex circuits without the need of constraints on the final steady state value to avoid numerical biases.
The response time to a switch-off stimulus can be calculated analogously. In this case, the initial condition is the steady state that is reached with a fully activated promoter and at time $t = 0$ the TF concentration is assumed to drop to zero. The response time again does not depend on the absolute value of the starting steady state level but only on the mRNA and protein half-lives:

$$\begin{aligned} \frac{r(t)}{r_{ss}} &= e^{-g_r t} \\ \frac{p(t)}{p_{ss}} &= \frac{g_p e^{-g_r t}) - g_r e^{-g_p t}}{g_p - g_r} \\ \frac{p(t)}{p_{ss}} &= e^{-gt}(1+gt) \quad if\ g_r = g_p = g. \end{aligned} \tag{3.4}$$

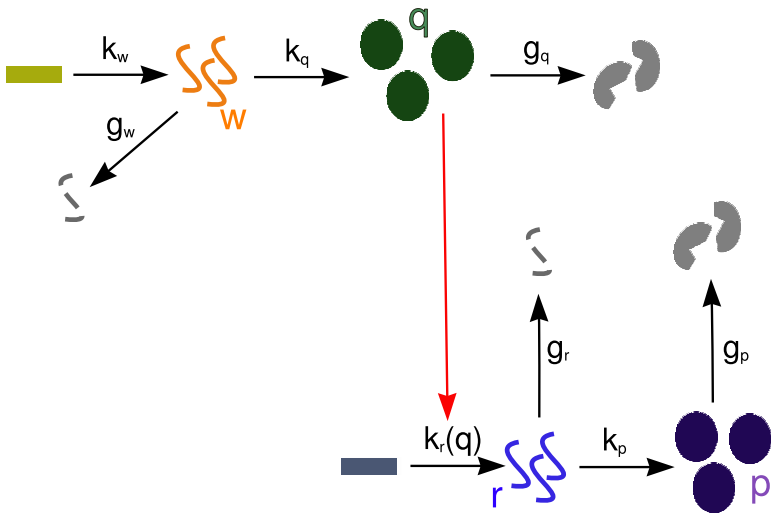

Figure 3.2: Scheme of a TF-gene linear circuit. - Rectangles represent DNA-genes, from which RNAs (w, r). RNAs can be translated into proteins (q is the copy number of TFs while p of target proteins) symbolized by circles. Proteins and mRNAs can be degraded (broken lines and circles). Rates of each process (transcription, translation or degradation) are depicted along the corresponding black arrows. Regulations are represented in red, with the arrow indicating activation by TFs.

3.2.1.2 Kinetics of a miRNA-mediated selfloop

The kinetic equations describing the mSL (Figure 3.3) dynamics are:

$$\begin{aligned}
\frac{dw}{dt} &= k_w - g_w w \\
\frac{dq}{dt} &= k_q w - g_q q \\
\frac{ds}{dt} &= k_r(q) - g_s s \\
\frac{dr}{dt} &= k_r(q) - g_r r \\
\frac{dp}{dt} &= k_p(s) r - g_p p,
\end{aligned} \tag{3.5}$$

where the rate of translation $k_p(s)$ is assumed to be a non linear function of miRNA concentration:

$$k_p(s) = \frac{k_p}{1 + (\frac{s}{h})^c}. \tag{3.6}$$

For the analysis of the response time to a switch-on signal, the TF dynamics (first two equations in 3.5) can be neglected, assuming full induction of the promoter at initial time $t = 0$. Therefore Equations 3.5 simplify to:

$$\begin{aligned}
\frac{ds}{dt} &= k_r - g_s s \\
\frac{dr}{dt} &= k_r - g_r r \\
\frac{dp}{dt} &= k_p(s) r - g_p p.
\end{aligned} \tag{3.7}$$

While it is easy to calculate analytically the steady state, the time evolution of the protein concentration can be extracted numerically. For simplicity, in the following the assumption $c = 1$ will be done, i.e. a Michaelis-Menten-like type of regulation as in reference (109).
The response time required to switch off the gene can be calculated in an analogous manner.

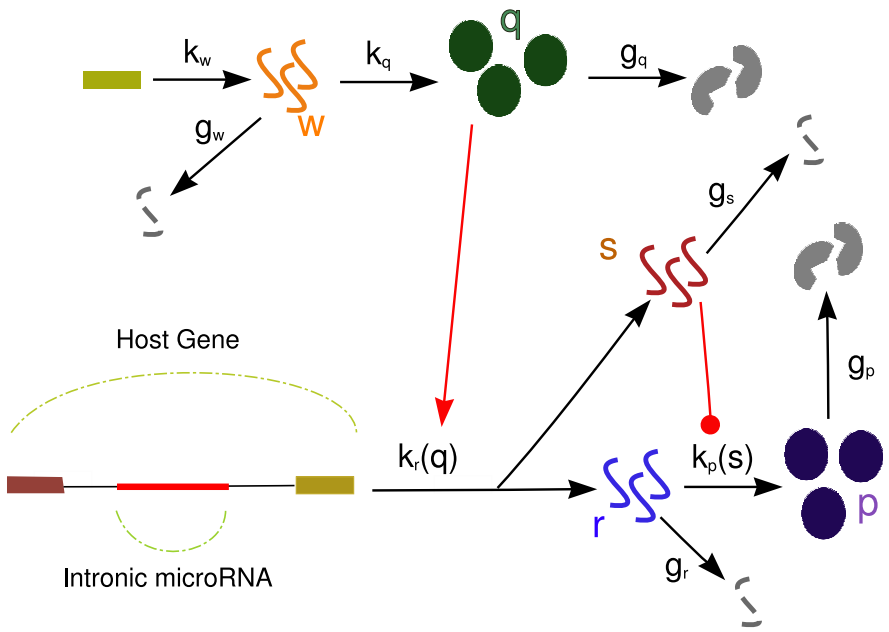

Figure 3.3: Scheme of an autoregulation via intronic microRNAs - The notation is the same of Figure 3.2. In this circuit mRNAs and miRNAs are co-transcribed. The miRNAs can then repress mRNA translation (red arrow with rounded end), making the rate k_p a non linear function of their concentration.

3.2.1.3 Kinetics of a transcriptional selfloop

The kinetic equations relative to the tSL in Figure 3.4 are:

$$\begin{aligned}\frac{dw}{dt} &= k_w - g_w w \\ \frac{dq}{dt} &= k_q w - g_q q \\ \frac{dr}{dt} &= k_r(q,p) - g_r r \\ \frac{dp}{dt} &= k_p r - g_p p. \end{aligned} \tag{3.8}$$

In this case, the rate of target transcription will be assumed to be represented by a product of Hill functions: one corresponding to the TF q activation and the other one taking into account the transcriptional self-repression. Thus, the form of the transcription rate is:

$$k_r(q,p) = k_r \left(\frac{q^c}{h_r^c + q^c} \frac{1}{1 + (\frac{s}{h_p})^c} \right). \tag{3.9}$$

Considering full induction of the promoter and initial conditions $x_i(t=0) = 0 \quad \forall i$, the transcription rate become: $k_r(q,p) \sim k_r \frac{1}{1+(\frac{s}{h_p})^c} = k_r(p)$. Therefore, Equations 3.8 simplify to:

$$\begin{aligned}\frac{dr}{dt} &= k_r(p) - g_r r \\ \frac{dp}{dt} &= k_p r - g_p p. \end{aligned} \tag{3.10}$$

In order to compare the kinetics of this circuit with the miRNA-mediated one, both circuits will be assumed as designed to obtain the same steady-state protein concentration. This condition can be implemented requiring that the solutions p_{ss} of Equations 3.10 and Equations 3.7 coincide. As a consequence, the value of the dissociation constant h_p in the Hill function of the transcriptional self loop becomes a function of the parameters of the miRNA-mediated circuit:

$$h_p = \frac{g_s^2 h^2 k_p}{g_p g_r (g_s h + k_r)}. \tag{3.11}$$

Thanks to this constraint mSLs and tSLs can be safely compared in terms of response time.
It is again straightforward to generalize the calculations to the switch-off dynamics.

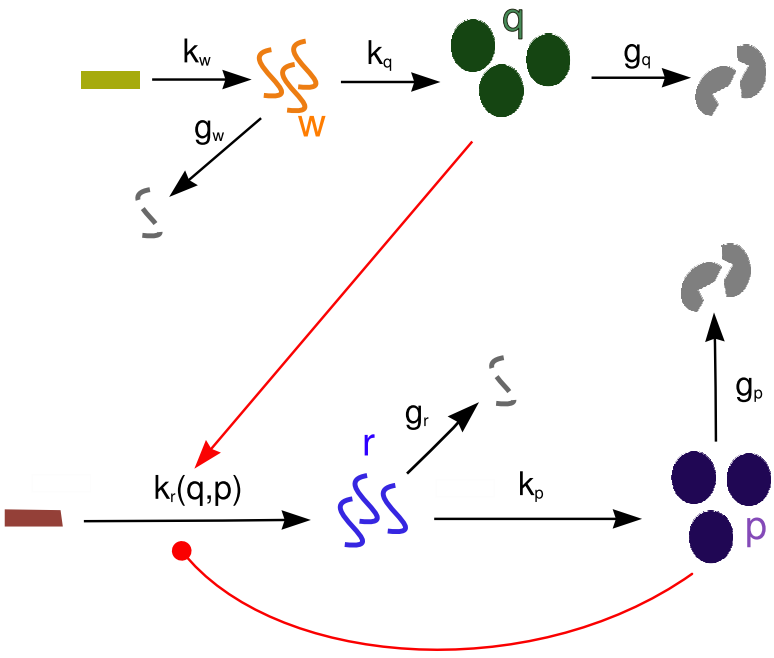

Figure 3.4: Scheme of a transcriptional negative self-regulation. - While the notations are the same of Figure 3.3, in this case the rate of mRNA transcription is a function of the TF concentration and of its own protein product.

3.2.2 Results

3.2.2.1 Switch-on dynamics

The rise-time of the three circuits in analysis is reported in Figure 3.5A, as a function of the repression strength p/p_0 (where p_0 is the target protein level in absence of any repression) and normalized with the response time of the simple transcriptional unit in Figure 3.2. As a first result, both the transcriptional and the post-transcriptional self-loops can speed up the target approach to the steady state. The tSL seems generally more effective, especially when the repression strength is mild (see Figure 3.5C). In fact, in a transcriptional feedback there is a time lag between the activation of transcription and the arrival of an amount of protein sufficient to drive the promoter repression. This lapse of time of unrepressed expression triggers the observed rapid rise of protein concentration. On the other hand, a mSL entails the simultaneous production of mRNAs and their miRNA repressors, therefore the regulative action is immediately present. However, as discussed in Section 2.11, a certain time delay can be also required for miRNA activation. Taking into account larger and larger time delays between miRNA transcription and activation progressively decreases the rise time achieved with the mSL (see Figure 3.5B). For sufficiently large time delays, the mSL can actually outperform the tSL as an accelerator.
This is also the case in the strong repression regime (even in absence of delays), where the acceleration is due to large overshoots in protein concentration (see Figure 3.5D and E). The target protein level can largely exceed the final steady state and then slowly relax toward it, as previoulsy observed for incoherent FFLs (110) and negative feedbacks with delays (109). This overshoot dynamics is more marked with mSLs than tSLs, leading to a slightly higher acceleration for strong repression. This can be simply explained by the fact that in a tSL the regulator and the regulated are represented by the same molecules: as the repression become strong, the final steady state is composed by few protein molecules that in the tSL must be extremely effective in repressing the transcription to keep the low steady-state level, actually limiting the amplitude of overshoots. This is not the case in mSLs where the regulator and the regulated are independent molecular species and consequently the amount of repressors is not compelled to equal the amount of target proteins. Thus, the system can reach a certain final equilibrium with more miRNAs than proteins, allowing a large overshoot before the accumulation of enough miRNAs to bring back the protein level to the low steady state.
It should be noticed that in the extreme limit of almost full repression of the target, the resulting dynamics of the mSL resembles a pulse of target expression. A strong suppression implies that the final steady state, reached after the pulse, is very close to the initial one. This dynamic behaviour is in close connection with the implementation of adaptation that will be fully discussed in a following section.
The response acceleration and the generation of a transient pulse are dynamic properties that mSLs share with incoherent FFLs. This is not surprising since the two circuits have many structural features in common. The main difference

is that in miRNA-mediated incoherent FFLs the mRNA and miRNA concentrations tend to be correlated as they are both activated by the same TF, while in mSLs the correlation is perfect since they share the same promoter.

3.2.2.2 Switch-off dynamics

This section explores the kinetic response of the three regulative strategies in front of a sudden drop of the level q of the activating TF. The time required for the target concentration to reach zero starting from a constitutive level of expression will be evaluated. In a simple transcriptional unit, the dynamics is an exponential convergence to the steady state, both in the presence of an activation input both in front of a switch-off signal (see Equation 3.4). In the presence of a transcription stop signal, the velocity of the protein concentration decrease is established by protein and mRNA degradation rates. In fact, long living mRNAs are more persistent and can be translated for a longer time after the stop of transcription. The same switch-off dynamics is observed for tSLs: as the transcription is already switched off, the repressive proteins cannot exert any regulation and their level simply undergoes the exponential decrease dictated by mRNA and protein degradation. On the other hand, the mSL shows a delay in the switch-off dynamics, as reported in Figure 3.6A. The reason of this larger inertia in shutting down target expression can be qualitatively explained: in a mSL, for each single miRNA that is degraded, many of the still present mRNAs sense an increase in their translation rate and this effect becomes more pronounced as the mRNA half-life increases (Figure 3.6A). Even a counterintuitive behaviour can arise: if the miRNA downregulation is especially effective, the general increase in mRNA translation rate for each miRNA degradation event can lead to a temporary boost in protein concentration (see Figure 3.6B and C).

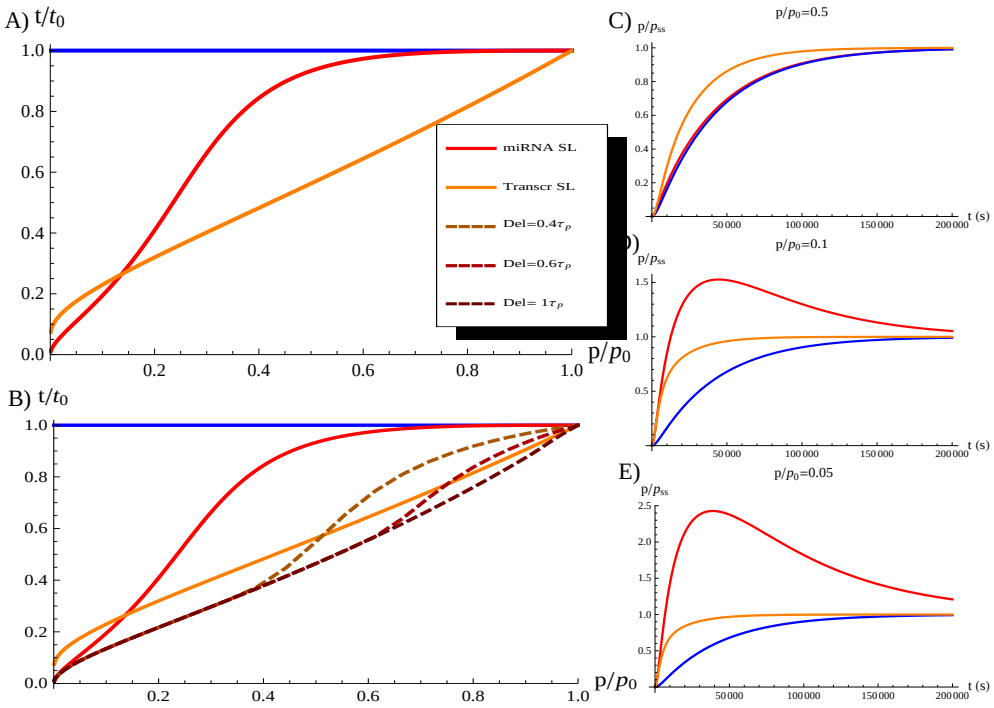

Figure 3.5: Response dynamics to a switch-on signal - (A) The response time normalized with respect to the TF-gene circuit is depicted for transcriptional and miRNA-mediated self loops. The parameters used are: the half-life of proteins and miRNAs is 8 hours; mRNA half-life is 30 minutes; the rate of transcription is $k_r = 0.2s^{-1}$ and of translation $k_p = 0.005s^{-1}$; the dissociation constant for the promoter is $h = 1000$. (B) The effect of inserting a delay between miRNA transcription and activation. The delay time is reported in unit of protein half-life. (C-E) The actual rise in protein concentration (normalized to the steady state value) for the different regulative strategies at different degrees of repression p/p_0

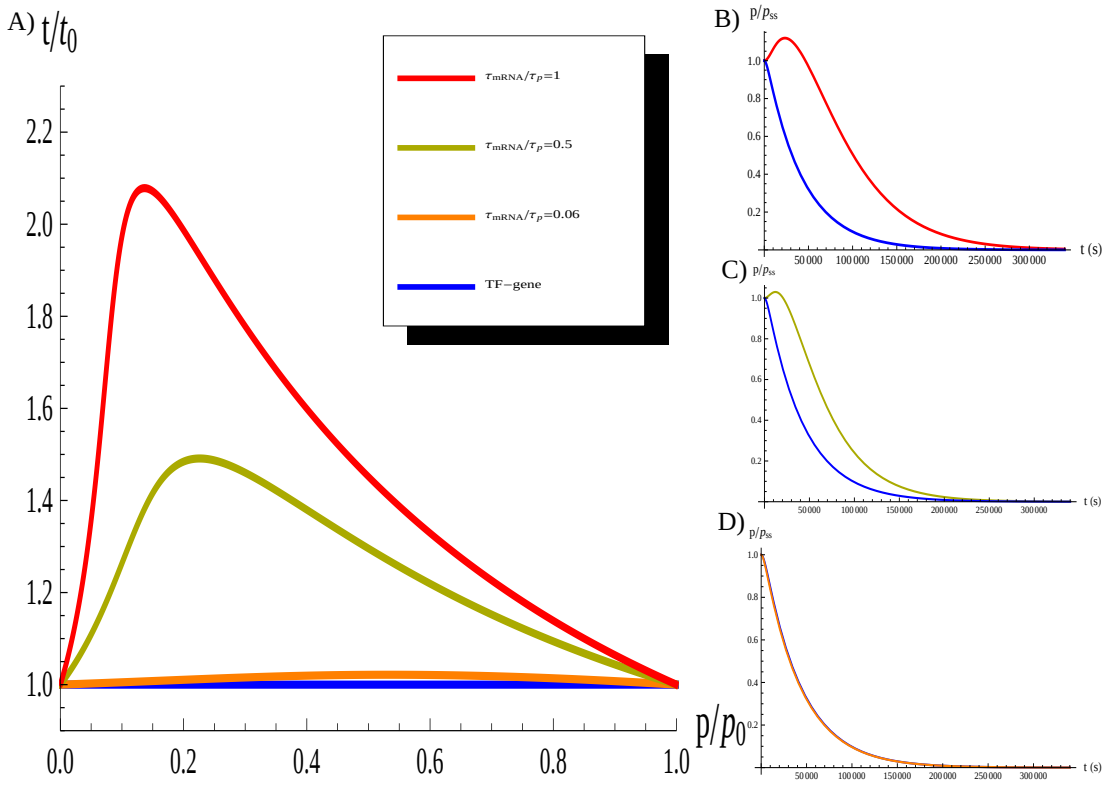

Figure 3.6: Response dynamics to a switch-off signal. - (A) The mSL response time as a function of the repression streght is shown for different values of the mRNA half-life. The response time is again normalized with respect to the dynamics of a simple transcriptional unit and it is depicted as a function of the repression strenght. The parameter values are the same of Figure 3.5. (B-D) In the case of $p/p_0 = 0.5$ the falling of target protein concentration (normalized to the initial steady state value) is shown for different values of the mRNA half-life.

3.2.3 A mechanism for locking the ON-state

While the tSL seems robustly designed for the purpouse of speeding up the rise time (even more efficiently than the mSL without delays), the mSL shows a more complex kinetics since it can both delay the switch-off dynamics both accelerate the switch-on response. Moreover, a comparison of Figures 3.5A and 3.6A reveals that the ranges of repression that make the mSL a rise accelerator and a delay element are largely overlapping. In the overlapping region, the coupling of the two behaviours can robustly keep the system in the ON-state. In the previous sections we implicitely assumed the presence of two steady states for target expression: the ON-state (characterized by full induction of the promoter) and the OFF state in which transcription is shut down. When the mSL target gene is expressed in the ON-state, it responds to a transient fluctuation of the input signal toward zero initially with a slow decrease and then with a quick recovery of the ON-state when the fluctuation is over. In this way, only in the presence of a persistent signal the system would turn-off the host gene expression. A resilient ON-state can be biologically important if it represents the homeostatic level that must be robustly kept high to ensure the physiological phenotype.

3.3 Adaptation and fold-change detection

Adaptation is defined as the ability of a system to respond to an input change and subsequently return to the prestimulus level, even if the stimulus persists. Many signaling systems show adaptation, although implemented in different ways. In all this systems, the benefit of adaptation is the detection of signals irrespective of the background level. Examples of nearly perfect adaptation range from chemotaxis in bacteria (121; 122) to sensor cells in higher organisms (123). Taking the example of chemotaxis, the advantage of an adaptive detection is clear: irrespective of the absolute amount of nutrients stably present in the environment, the bacteria can adjust its sensor system to the background and thus be ready to climbe eventual gradients. This way it can always move toward the best intake condition.
Simple network topologies, as negative feedback loops with a buffering node or incoherent FFLs, can be at the basis of the cellular implementation of adaptation (124).
An additional feature that certain adaptive signalling systems employ is fold-change detection (FCD) (126; 127): the system response depends only on the fold change in the input signal and not on its absolute value (see Figure 3.7). FCD clearly entails exact adaptation (128) since it requires that not only the steady state value does not depend on the input stimuli but also that the shape of the dynamic response is the same in the presence of an equal fold change in the input. The incoherent FFL has been identified as a potential candidate circuit to implement FCD (125). FCD has two potentially useful properties. First, it confers robustness to the response with respect to cell-cell varibility. Second, it can keep a good balance between sensitivity and noise filtering (125) (see section 3.5 for a more detailed discussion).
The analysis presented in this section will show that also the simple mSL topology can be added to the list of successful performers of adaptation and FCD. Furthermore, the range of the biochemical parameters in which the mSL can perform each function will be fully characterized.

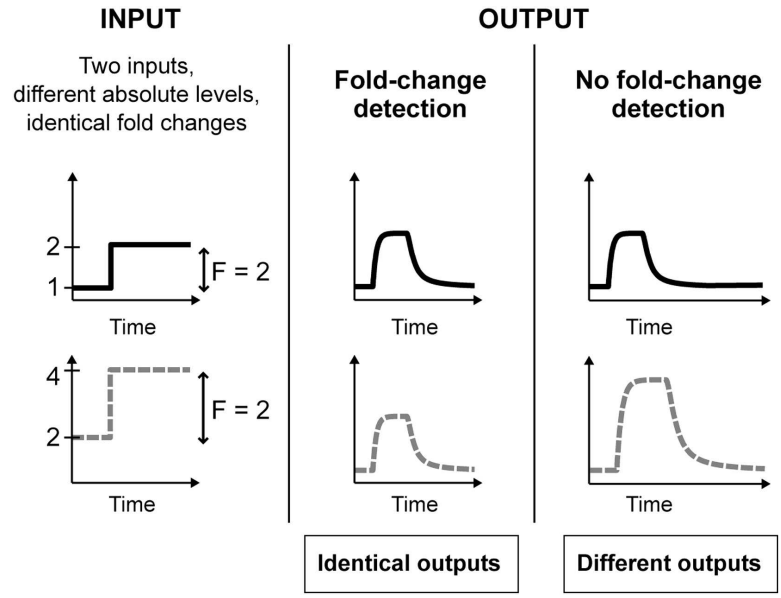

Figure 3.7: Definition of fold-change detection. - In the presence of two stimuli with the same fold change F (left column) the system can show adaptation without FCD (same steady state but different dynamic responses -right column) or adaptation and FCD (identical output -central column). (Adapted from reference (125))

3.3.1 Analytical results

The complete dynamics of the TF (first two Equations in 3.5) will be again neglected and q will be considered as an input parameter.
Assuming a strong miRNA repression, where $s/h >> 1$, the mSL dynamics simplifies to:

$$\begin{aligned} \frac{ds}{dt} &= k_r(q) - g_s s \\ \frac{dr}{dt} &= k_r(q) - g_r r \\ \frac{dp}{dt} &= k_p r/s - g_p p, \end{aligned} \tag{3.12}$$

In the steady state solution of these equations, the protein level does not depend on the quantity of TFs:

$$\begin{aligned} r_{ss} &= \frac{k_r}{g_r} q \\ s_{ss} &= \frac{k_r}{g_s} q \\ p_{ss} &= \frac{k_p g_s}{g_p g_r}. \end{aligned} \tag{3.13}$$

This is the condition required to implement adaptation. If the steady state value does not depend on the input signal q, the system always returns to its specific steady state after the dynamic response to the input, even in the presence of a persistent stimulus.

Introducing the further assumption of approximately linear activation of transcription, i.e. the amount of q is far from saturating the promoter of the target ($q/h << 1$), the rate of transcription can be written as:

$$k_r(q) = \frac{k_r q}{h_r + q} \sim k_r q. \tag{3.14}$$

Consequently, Equations 3.12 simplifies to:

$$\begin{aligned} \frac{ds}{dt} &= k_r q - g_s s \\ \frac{dr}{dt} &= k_r q - g_r r \\ \frac{dp}{dt} &= k_p r/s - g_p p. \end{aligned} \tag{3.15}$$

Finally, we assume that the mRNA half-life is shorter that the other time scales in the system. The situation of short-living mRNAs (with respect to proteins) is usual in prokaryotes, while it can be considered as a specific (but anyway quite diffused) case in eukaryotes. With this condition satisfied, the mRNA dynamics results very fast with respect to protein or miRNA ones. Hence, the mRNA level can be considered at equilibrium. This allows a further reduction of the kinetic equations:

$$\begin{aligned} \frac{ds}{dt} &= k_r q - g_s s \\ \frac{dp}{dt} &= \frac{k_p k_r}{g_r}\frac{q}{s} - g_p p. \end{aligned} \tag{3.16}$$

Let us now define the following dimensionless variables:

$$\begin{aligned} t' &= g_s t, \\ s' &= \frac{g_s s}{k_r q_0}, \\ F &= q/q_0, \\ p' &= \frac{g_r g_p}{g_s k_p}, \\ \phi &= g_s/g_p, \end{aligned} \tag{3.17}$$

where the input stimulus is represented by a change in TF concentration from a basal level q_0 to a new level q. Equations 3.16 can now be rewritten as:

$$\begin{aligned} \frac{ds'}{dt'} &= F - s' \\ \phi\frac{dp'}{dt'} &= \frac{F}{s'} - p'. \end{aligned} \tag{3.18}$$

The dynamics of the target protein depends only on the fold-change F of the input and not on the absolute value, proving the ability of mSLs to implement FCD in suitable conditions. A crucial role in protein dynamics is played by the ratio between half-lives that steps in through the parameter ϕ.

3.3.2 Numerical results

In the previous section we demonstrated that the mSL can perform exact adaptation and FCD in certain mathematical limits. While adaptation only requires strong target repression, to perform FCD the mSL circuit must satisfy:

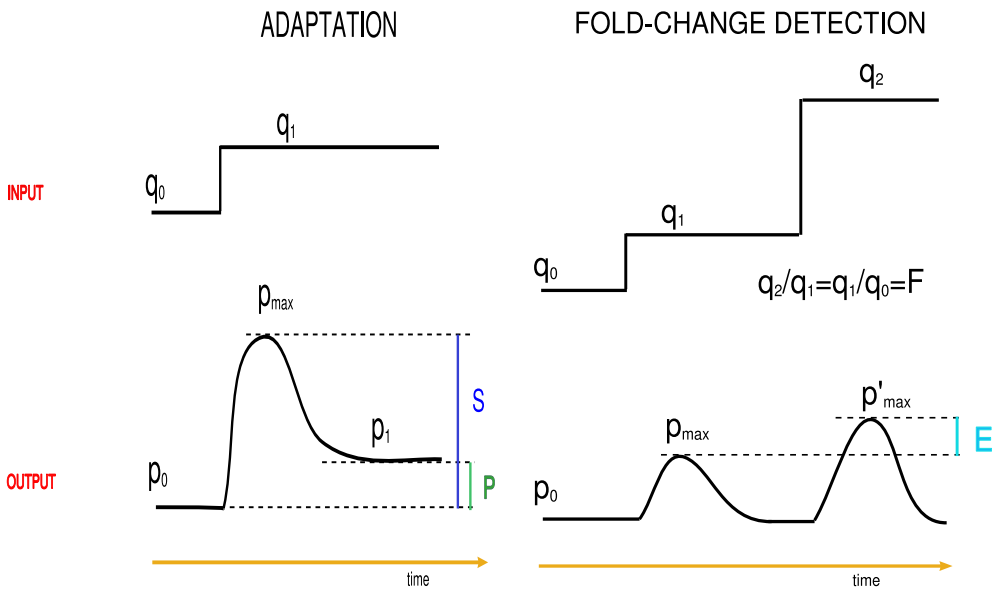

Figure 3.8: Measures of the peformance in adaptation and fold change detection. - To test the degree of adaptation we provide a step input in the TF concentration (from q_0 to q_1) and we measure the distance between the initial and final steady state level (precision P). To check that the system is actually responding to the stimulus, we evaluate the height of the peak (p_{max}) with respect to the prestimuls level (sensitivity S), requiring that S is above the noise level $S > 3CV_{p_0}$. In order to evaluate the ability of adaptive mSLs to implement also FCD, we provide two step stimuli with same fold change F, and we measure the distance between the two maximum target protein levels (E).

- Strong repression $k_p(s) \sim k_p/s$ (condition for adaptation),
- Linear activation of the promoter $k_r(q) \sim k_r\ q$,
- Fast mRNA dynamics (short mRNA half-life): $r(t) \to r_{ss}$.

To go beyond this ideal limiting case, the behaviour of the complete realistic model of mSLs will be tested in the following. In particular, the association between parameter ranges and performance levels of the functions in analysis will be characterized thanks to an exploration of the parameter space (binding affinities, degradation rates, production rates).
First of all, for each set of biochemical parameters a stimulus -as a step change in the TF concentration q- is provided (see Figure 3.8) and two functional quantities indicative of the adaptive behaviour are measured: *Sensitivity* (S) and *Precision* (P).

$$\begin{aligned} S &= \left| \frac{p_{max} - p_0}{p_0} \right| \\ P &= \left| \frac{(p_1 - p_0)/p_0}{(q_1 - q_0)/q_0} \right|^{-1} . \end{aligned} \tag{3.19}$$

P is a measure of the difference in the steady state levels before and after the stimulus, therefore it is actually an estimation of the degree of adaptation. Following (124), we define the minimal threshold of $P > 10$ to select adaptive circuits. On the other hand, a sufficiently high level of S must be required to ensure that adaptation is not a merely consequence of complete insensitivity to changes in TF concentration. We set as a threshold $S > 3CV_p$ where CV_p is the noise level of p at the pre-stimulus steady state, assuming a realistic level of TF relative fluctuations of 10% (see the next section for the stochastic model of the circuit). In this way, we can test if the circuit is able to respond dynamically with a rise in p concentration above the noise level, therefore with a recognizable signal. However, it should be noticed that the amplitude of the p response can be amplified by a slow miRNA dynamics or by the presence of delays between miRNA transcription and activation (125), therefore the level of S can be significantly higher for long-living miRNAs or in case of delays. The two measures P and S were previously introduced in the context of enzimatic networks (124), but we slightly modified the definition of sensitivity. While in long signaling chains sensitivity can require signal amplification (as actually sensitivity is defined in (124)), we think that in basic genetic circuits as mSLs the above discussed signal-to-noise ratio can be more informative of the ability to transmit signals in a noisy background.
Each parameter set that allows an adaptive behaviour is also tested for FCD. To this aim, two step inputs are provided. The inputs have the same fold change but start from different background levels (see Figure 3.8). Following (125), as a

measure of the fold-change property the relative difference (E) in the amplitude of the target protein responses (p_{max} and p'_{max}) to the two inputs is evaluated:

$$E = \left| \frac{p_{max} - p'_{max}}{p'_{max}} \right|. \tag{3.20}$$

Figure 3.9 reports the results of the proposed evaluations (S, P, E) as a function of q/h_r (related to the level of promoter activation) and $1/h$ (a measure of the repression strength). The grey region represents the parameter space that allows adaptation ($P > 10$). Inside this region, the measure of the FCD ability (E) is depicted with a color code. The minimal condition of $E < 0.1$ is required, i.e. the two inputs produce an identical output within 10%. For all the parameter values in the colored area (Figure 3.9) the circuit shows sufficiently high sensitivity ($S > 3CV_p$).
The mSL can perform adaptation and FCD in a wide range of parameters. The requirements of strong repression (in the depicted area $p/p_0 \lesssim 0.2$) and almost linear activation of the promoter ($q/h_r << 1$) seem actually satisfied in a parameter range that span several orders of magnitude. This fact suggests that mSLs can be designed to perform adaptation and FCD in a relatively simple way, without a fine-tuning of the parameters but only keeping their ratio in a quite large interval.

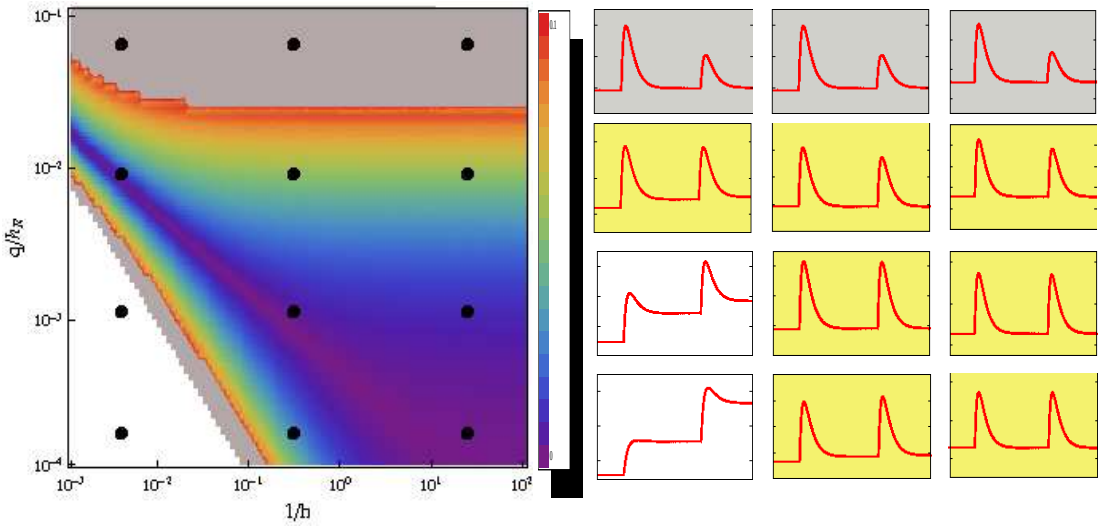

Figure 3.9: Adaptation and fold-change detection with miRNA-mediated autoregulation. - The area defined by repression strength $1/h$ vs promoter activation q/h_r is colored in grey if the mSL circuit shows adaptation ($S > 3CV_{p_0}$ and $P > 10$). For parameter values that ensure adaptation, the measure of FCD efficiency (E) is reported with the color code in the legend. On the right, the dynamic response of the circuit to the two inputs is plotted with the parameter values corresponding to the black dots in the left plot.

3.4 Noise reduction

3.4.1 The theoretical framework

The mSL shares many stuctural features with the miRNA-mediated FFL described in the previous chapter 2. In both circuits a TF activates the expression of a target gene and together with it a miRNA repressor of the target. The difference relies on the fact that the transcription events of target mRNAs and miRNAs are fully coupled due to the common promoter in the mSL case. The analogy can be observed also at the level of the master equation (compare with Equation 2.3):

$$\begin{aligned}\partial_t P_{w,q,s,r,p} = \quad & k_w(P_{w-1,q,s,r,p} - P_{w,q,s,r,p}) + k_q w(P_{w,q-1,s,r,p} - P_{w,q,s,r,p}) \\ & + k_r(q)(P_{w,q,s-1,r-1,p} - P_{w,q,s,r,p}) + k_p(s) r(P_{w,q,s,r,p-1} - P_{w,q,s,r,p}) \\ & + g_w\Big[(w+1)P_{w+1,q,s,r,p} - wP_{w,q,s,r,p}\Big] + g_q\Big[(q+1)P_{w,q+1,s,r,p} - qP_{w,q,s,r,p}\Big] \\ & + g_r\Big[(r+1)P_{w,q,s,r+1,p} - rP_{w,q,s,r,p}\Big] + g_s\Big[(s+1)P_{w,q,s+1,r,p} - sP_{w,q,s,r,p}\Big] \\ & + g_p\Big[(p+1)P_{w,q,s,r,p+1} - pP_{w,q,s,r,p}\Big]. \end{aligned} \tag{3.21}$$

The master equation can be solved analytically using the moment generating function and linearizing the Hill functions of regulation, as explained in section 2.2. The noise properties of the mSL (Figure 3.3), of a simple transcription unit (Figure 3.2) and of a negative transcriptional feedback loop (Figure 3.4) will be compared. All circuits can be modeled and treated using the same mathematical procedure. As previously discussed, the different regulative strategies can be constrained to produce the same mean amount of target proteins, allowing an unbiased comparison of noise properties.

3.4.2 Results

In terms of noise-buffering properties, many of the results presented for miRNA-mediated incoherent FFLs apply also to the mSL case. In particular, mSLs can filter fluctuations propagating from the upstream TF. As a consequence, the steady state level is less noisy if achieved through a mSL circuit with respect to a simple TF-gene activation designed to produce the same target level (see Figure 3.10A). Even if transcriptional self regulation can function as a noise filter, as previously shown (18; 22; 23), the autoregulation via microRNAs seems outperforming. The fact that noise buffers can be more effectively designed using miRNA regulations has been previously discussed in section 2.6.
Both strategies of self regulation show an optimal noise buffering for an intermediate repression strength, but again the degree of attenuation is clearly larger using miRNAs (see Figure 3.10B). A more extensive exploration of the parameter

space points out that a mSL can buffer fluctuations over a wider range of conditions as well as to a greater extent (compare Figure 3.10C and D). The optimal noise buffering does not require necessarily strong repression. In fact, the region of maximum noise reduction corresponds to a suppression of target expression of approximately the 50% of its constitutive value. Therefore, the intronic miRNA can keep its host gene expression in its homeostatic regime while filtering out fluctuations without exerting a strong reduction of its concentration.

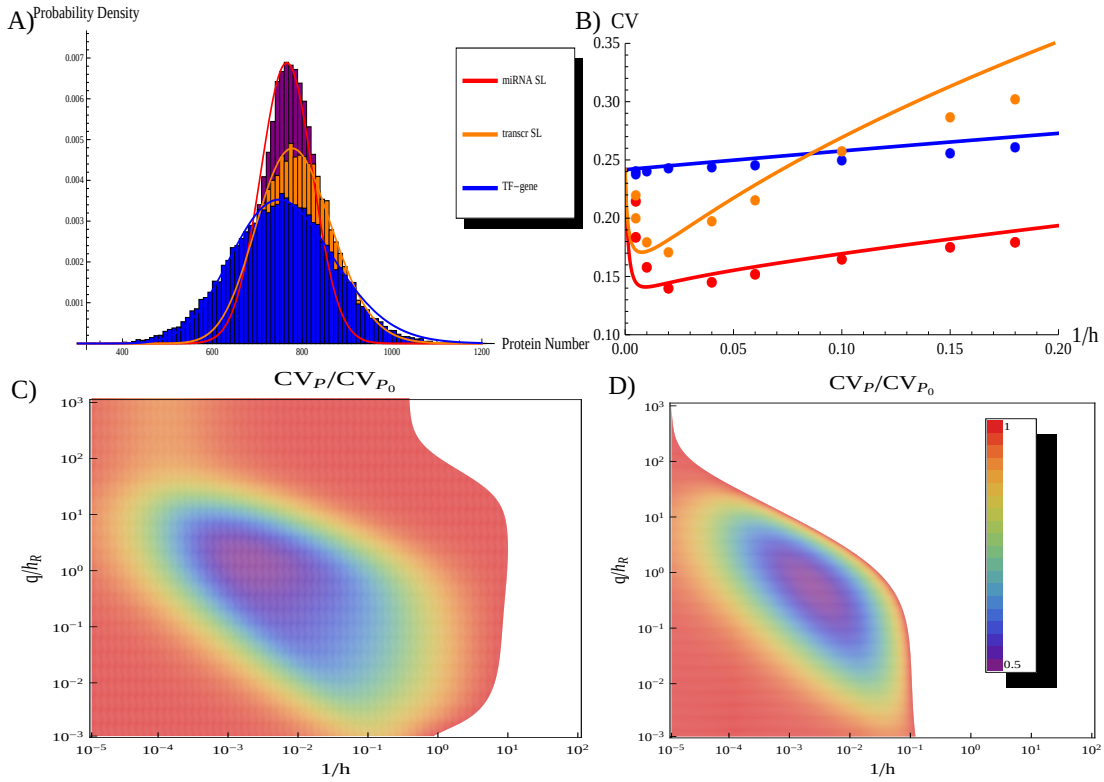

Figure 3.10: Noise buffering with miRNA-mediated selfloops - (A) The probability distribution of protein number for the three circuits. Histograms are the result of Gillespie simulations while continuous lines are empirical distributions with mean and variance predicted by the analytical model.(B) The coefficient of variation of the target protein CV_p as a function of the repression strength $1/h$.(C) The noise reduction CV_p/CV_{p_0} achieved with a mSL, evaluated for different mean levels of the TF $< q >$ and different degrees of repression $1/h$. CV_{p_0} represents here the noise level of a target simply activated by a TF. The region where miRNA repression leads to larger fluctuations with respect to constitutive ones is shown in white, while when a noise reduction is gained the value of CV_p/CV_{p_0} is reported with the color code explained in the legend. All the free parameters are those reported in caption of Figure 3.5.(D) The same plot of (C) in the case of a tSL constrained to produce the same amount of target proteins.

3.5 Conclusions and discussion

Intronic miRNAs often target genes that are functionally related to the host (103; 104), thus coordinating the expression program. However, pieces of evidence that in many cases the host is a target itself suggest a functional role for the so formed mSLs. This final section resumes the different functions -in different (although overlapping) ranges of biochemical parameters - that these circuits can perform.
Firstly, the mSL circuit can keep the expression of the host gene robustly in the ON-state, delaying the switch-off dynamics while shortening the rise time. As a result, the host gene can persist in a widely active expression state until the income of a persistent enough repressive signal. High-frequency downwards fluctuations of the TF concentration are filtered out by the intronic miRNA regulation. An efficient implementation of this function requires a full induction of the promoter (which is actually the definition of the ON-state) and a sufficiently strong repression $0.1 \lesssim p/p_0 \lesssim 0.4$ (see Figures 3.5,3.6), where p_0 is the protein level in absence of miRNA regulation.
Secondly, FCD (and consequently adaptation) can be succesfully encoded in mSLs. The necessary conditions are: strong repression ($p/p_0 \lesssim 0.2$), an almost linear promoter activation ($q/h_r >> 1$) and short living mRNAs. As previously proposed in (125), the ability to respond to the input fold change rather than its absolute value can have potentially useful properties, again conducive to the robustness of the host expression. Since the nuclear level of a transcription factor can vary several-fold between individual cells (127), FCD allows cells with a different basal level to respond identically to a common external signal (see Figure 3.11A). Moreover, FCD can help the discrimination between noise and signals in different backgrounds. Since the noise strength increases with signal strength, when the background is variable an absolute detection would run into trouble as it cannot scale the detection threshold relatively to the background (see Figure 3.11B).
Lastly, we addressed the stochastic properties of mSLs. In complete analogy with the case of incoherent miRNA-mediated FFLs, we found that the optimal attenuation of fluctuations is realized for an intermediate level of repression ($p/p_0 \sim 0.5$) and for an intermediate level of promoter activation ($q/h_r \sim 1$). This is exactly the range where the target is more sensitive to changes (in both directions of increase and decrease) of the TF concentration. In these conditions, a fine-tuning can be necessary to keep the steady-state in a functional intermediate level between full expression and repression. The intronic miRNA can actually fine-tune the host mean level while reducing fluctuations, thus lowering the odds of potentially harmful deviations from the desired protein concentration at the single-cell level.

Summing up (see also Figure 3.12):

- **Robust ON-state** $\Rightarrow$ strong repression; promoter saturation.
- **FCD** $\Rightarrow$ strong repression; linear promoter activation; short living mRNAs.

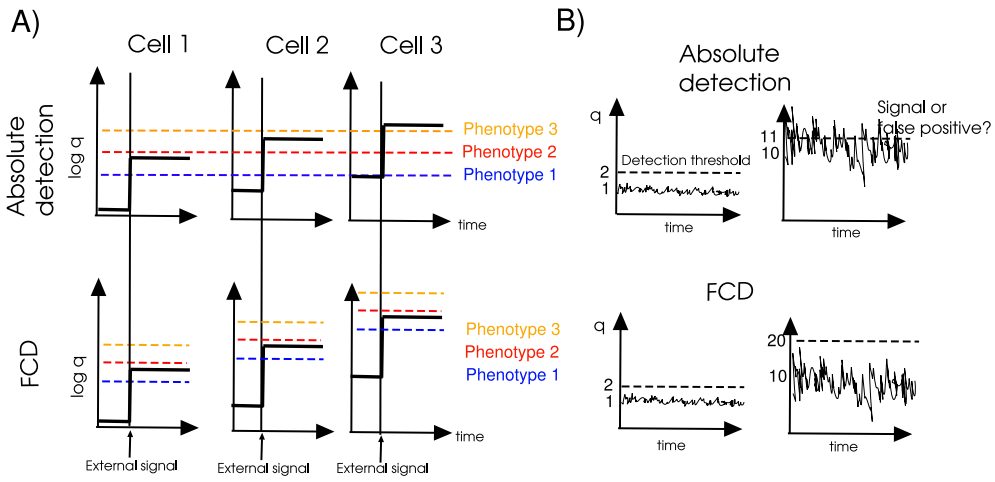

Figure 3.11: FCD and cell-to-cell variability - .(A) Despite variations in the basal level of the TF q, FCD ensures that each cell responds correctly, triggering the right final phenotype. (B) FCD rescales the detection threshold according to the background noise. Incoming signals are evaluated with respect to the background level, allowing a filtering of noise-driven fluctuations. (Adapted from reference (125)).

- **Fine-tuning and noise buffering** ⇒ intermediate repression and activation.

In conclusion all the potential functions associated with mSLs are linked to robustness with respect to fluctuations. However, the definition of robustness can change depending on the steady state that should be mantained resistent to noise and therefore on the specific biological function that is needed. If the target must be widely active at its maximum level of transcription, a relatively strong miRNA repression can lock this ON-state, changing the dynamics of the host gene expression. Alternatively, if the host gene should detect signals while the basal level of its TF presents large cell-to-cell variability, its promoter should be far from saturated and the host mRNA half-life must be sufficiently short (besides again strong repression). Finally, if there is only an intermediate target level which is functional, the miRNA regulation, exerting a relatively mild repression, can fine-tune the host product concentration at the desired level while filtering out fluctuations.
As discussed in (72) and (86), a role of miRNA regulation in robustness to noise can explain why miRNAs are often highly conserved, controlling key steps in development, but in many cases they can be deleted with little phenotypic consequences. While in the usually controlled laboratory conditions the importance of noise resistence could be undervalued, it can be crucial to establish a robust phenotype in the wild-type fluctuating environment. We are currently working on a bioinformatic search for mSLs in different species with the aim of providing a reliable set of candidates amenable to experimental validation of the proposed associated functions.
Besides an attempt of explaining the functional role of endogenous cases of autoregulation via intronic miRNAs, the results presented here are a potential indication for the design of synthetic circuits. The field of synthetic biology is growing incredibly fast with likely forthcoming applications in therapeutics (129) and non-coding RNA regulation represents a helpful resource in engineering genetic networks (130; 131). A detailed map between simple circuit topologies and functions would represent an instruction card for bioengineers trying to encode a specific functional program into a man-made network. In this view, mSLs can be a useful core module for which we provided a first, although yet preliminary, manual of functions. Involving one single gene and its promoter, mSLs represent probably the simplest possible implementation of a complex function like FCD. Moreover, mSLs could be used to confer robustness to specific nodes in larger scale syntethic networks.

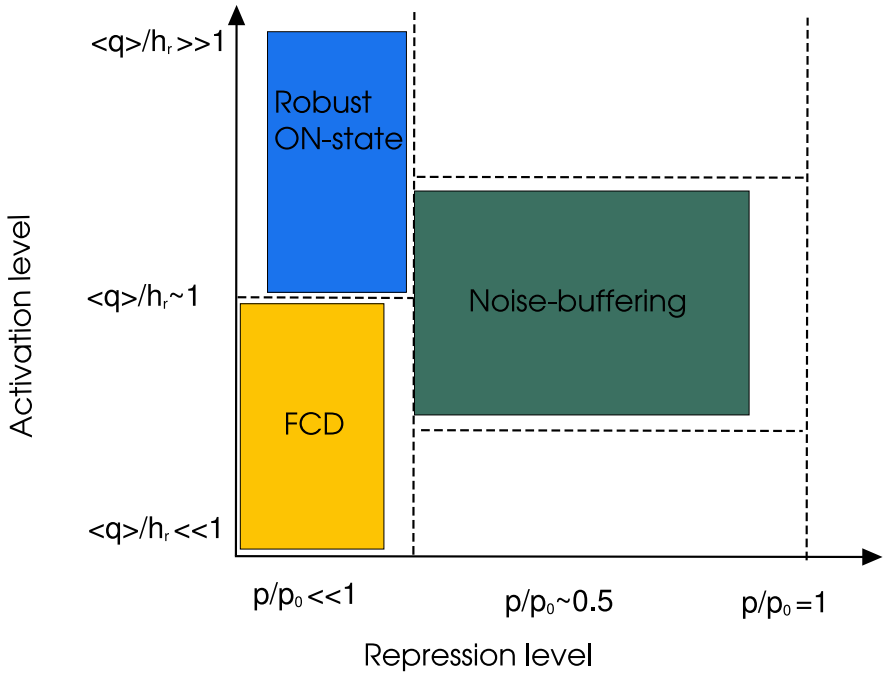

Figure 3.12: Function diagram of miRNA-mediated self loops. - Simplified scheme of the various functions associated with mSLs. The different ranges actually overlap but the optimization of a specific function identifies the general conditions reported in the scheme.

Appendix A

Entropic contributions to the splicing process

A.1 Introduction to the problem

A.1.1 The splicing process and the spliceosome assembly

Eukaryotic genes have a split nature, in which the exons, that encode the information for the final product of a messenger RNA (mRNA), are interrupted by non-coding introns in the DNA and in the precursor mRNA (pre-mRNA) transcript. The intron excision and the concomitant joining of exons, which basically represent the splicing process, are a necessity in order to obtain a mature mRNA that can be exported in the cytoplasm and for example correctly translated into a protein. This process is carried out by the spliceosome, a macromolecular ribonucleoprotein complex, that assembles on pre-mRNA in a stepwise manner (133; 134; 135). The first requirement is the correct recognition of the intron/exon boundaries by small nuclear ribonucleoproteins (snRNPs) and some auxiliary splicing factors by binding to sequences located at the ends of introns. Subsequently the splice-site pairing takes place, bringing the two exons near to each other and looping the intron that have to be cut away.

A.1.2 Exon definition and intron definition

Although the molecular players and the key steps of spliceosome assembly are remarkably conserved through different species (136), there are two alternative pathways of splice-site recognition: *intron definition* and *exon definition* (133; 137; 138; 139; 140).

A. ENTROPIC CONTRIBUTIONS TO THE SPLICING PROCESS

Intron definition (see figure A.1a) begins with the direct interaction of the U1 snRNP with the splice-site in the upstream end of the intron (5' splice-site). The splice-site in the downstream end (3' splice-site) is then recognized by the U2 snRNP and associated auxiliary factors such as U2AF and SF1. When the two complexes are constructed on the intron/exon boundaries they can be juxtaposed, closing an intron loop which is then spliced away in order to correctly glue the exons. The interaction of the splicing factors bound at the splice-sites occurs in this case across the intron. The exon definition (see figure A.1b) requires instead that the initial interaction between the factors bound at the splice-sites occurs across the exon: the U1 and U2 snRNP and associated splicing factors bind to the 3' and 5' ends of an exon and a complex is built across it (usually with the participation of serine/arginine-rich (SR) proteins (138)); then complexes on different exons join together so as to allow intron removal.

Previuos studies have shown that the length of the intron that has to be removed has a key role in the choice of the splice-site recognition modality (139; 142; 143). Short introns are spliced away preferentially through intron definition, while longer introns seem to require an exon definition process. In particular the analysis of (142) suggests the presence of a threshold in intron length - between 200 and 250 nucleotides (nt) long- above which intron-defined splicing ceases almost completely. Lower Eukaryotes present typically short introns, below the threshold, so it is expected that intron removal proceeds through intron definition (137; 140; 143; 144). Higher Eukaryotes instead have an intron length distribution presenting two pronounced peaks, with the threshold in between (see figureA.2), so even if the vast majority of introns are above the threshold (data in table A.1), the first peak contains introns suitable for intron definition. This agrees with several studies (143; 144; 145; 146) which have shown that both ways of splice-site recognition are present in higher Eukaryotes, even if the exon definition pathway seems to be the prevalent one.

As it can be seen in figure A.2, not only the shape of the distribution is quite conserved through different species, but also the position of the peaks.

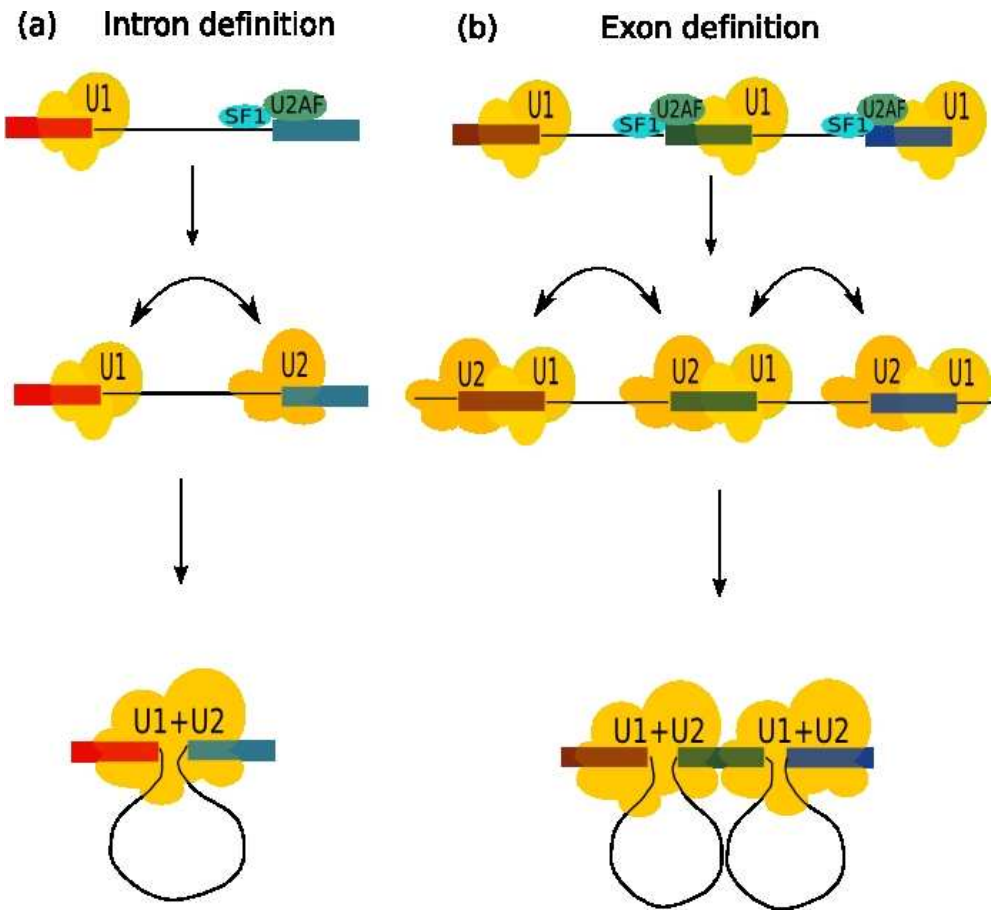

Figure A.1: Intron definition and exon definition: two ways of splice-site recognition. -

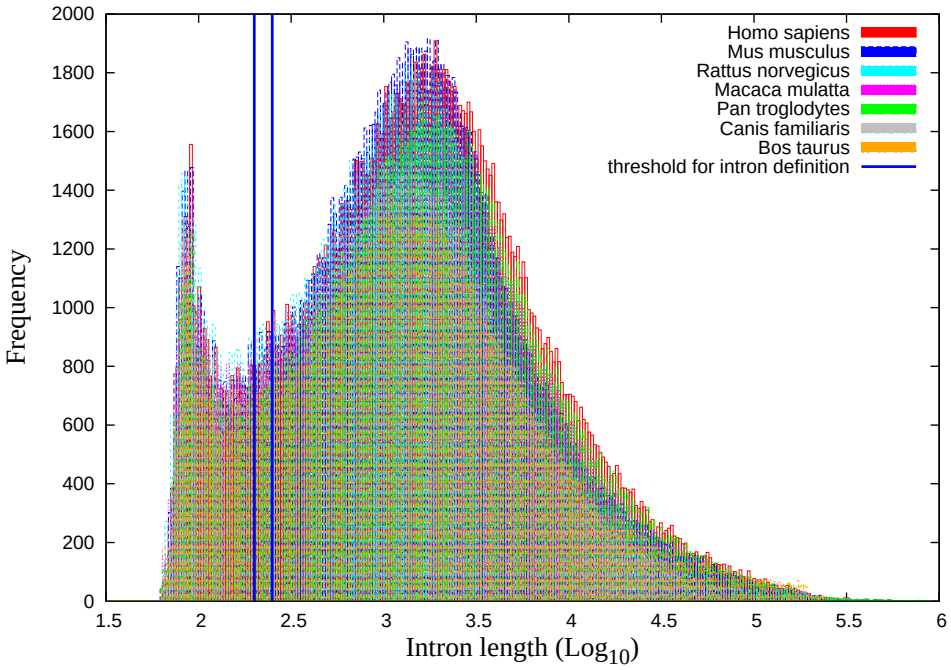

Figure A.2: Intron length distribution for different higher Eukaryotes. - The distribution shows a two peaks structure which is remarkably universal. The intron length threshold mentioned in the main text (blue lines) is located exactly between the two peaks. The right peak contains mostly introns which undergo exon-defined splicing, while the left one can be associated to intron-defined spliced introns. The coordinates of introns used were downloaded from the Ensembl database vers.47 (141) and the distribution was obtained through a logarithmic binning.

A.2 Intron removal and depletion attraction

The first goal of our paper is to propose a simple physical model of early steps of spliceosome assembly on a pre-mRNA, taking into account possible entropic contributions to the splicing process. Subsequently we will show that, despite its simplicity, the model is able to produce quantitative predictions which are in rather good agreement with experimental and bioinformatical observations.
Our starting point is the assumption that the splicing complexes, which are immersed in the crowded nuclear environment ((147) and reference therein), feel the so called "depletion attraction" (148). This interaction is essentially an entropic effect due to the fact that when two large complexes (like the splicing ones) approach each other, they reduce the volume between them excluded to the depleting particles. If the complexes are immersed in an environment crowded of macromolecules of smaller (but comparable) size, then this excluded volume effect induces an attractive interaction between the two complexes.

This simple geometrical reasoning forms the basis for the Asakura-Oosawa (AO) theory (148; 149). In more recent years, a more sophisticated hypernetted-chain-based theory describing depletion forces in fluids has been developed (150; 151) and tested in Monte Carlo simulations (152). However, as discussed in (147), the AO theory is an approximation that remains quite accurate up to $c \sim 0.3$, with c representing the fraction of volume occupied by the crowding molecules. As far as the c value inside a living cell has been estimated between 0.2-0.3 (153; 154) we can safely use in the following the simpler AO description of depletion effects.

Since the two splicing complexes are joined by a freely fluctuating RNA chain the depletion-based interaction becomes effectively long range, with a logarithmic dependence on the chain length. We suggest that this depletion attraction is the driving force which allows the splicing complexes to meet and join one another, in order to start up the splicing process. As we shall see this assumption naturally leads to a smooth cross-over from an intron defined to an exon defined splicing pathway as the chain length increases.

A.2.1 Presentation of the model

Let us model, as a first approximation, the pre-mRNA as a Freely Jointed Chain (FJC) (155; 156), i.e a succession of infinitely penetrable segments, each of length l equal to the Kuhn length of the single strand RNA (ssRNA). The estimated Kuhn length of ssRNA is approximately in the range 2-4 nm, i.e 3-6 nt (157; 158; 159). We chose to neglect the self avoidance in order to use the analytical tractable FJC and moreover the diameter of ssRNA, approximately 2 nm, is not so relevant with respect to long chains: as reported in (157; 160) the FJC modelization is suitable for ssRNA chains with a length greater than 5-6 Kuhn segment, as will always be the case in the following.

The two complexes, composed by U1, U2 and splicing factors, that bind to the exon/intron boundary in the intron definition process, will be modeled as spheres with a diameter D (the dimensions of the major components U1 and U2 are quite similar, both of the order of ~ 10 nm, see (161) and (162) for details). The same geometrical approximation will be done for complexes constructed across exons in exon definition. They will be considered as spheres of diameter D', with $D' \sim 2D$

since they are composed by both the U1 and the U2 subcomplexes plus the exon in between, usually with SR proteins bound to it (140).
The simple FJC model allows the analytical calculation of the radial probability distribution of the end-to-end distance (155):

$$W(r)dr = \left(\frac{\beta}{\sqrt{\pi}}\right)^3 4\pi r^2 \exp(-\beta^2 r^2)dr, \tag{A.1}$$

where $\beta = \left(\frac{3}{2nl^2}\right)^{\frac{1}{2}}$, n is the number of indipendent segments in the FJC and l is the length of a segment (in our case the Kuhn length of mRNA). Following (163), in order to include the depletion attraction contribution, we weighed the radial probability distribution of the end-to-end distance (we assume that the ends of the intron can be considered as the center of the beads) with a Boltzmann factor, which takes into account the depletion attraction potential and which is non-zero in the range $D \leq r \leq D + d$. This potential is easy to evaluate in this "hard sphere" approximation (see for instance (149)) and takes a particularly simple expression in the $d << D$ limit. We can therefore define a new function $W'(r)$ as the weighed FJC radial probability distribution:

$$W'(r)dr = W(r)\exp(\frac{3}{2}c\frac{D}{d}(\frac{D+d-r}{d})^2)dr, \tag{A.2}$$

where c denotes the volumetric concentration of the small molecules and d their typical size. With the typical values of these quantities for the nuclear environment: $c \sim 0.2$ and $d \sim 5nm$ one finds for the problem at hand a potential energy of the order of one hydrogen bond, which is exactly in the range of energies needed to join together the two ends of an intron of length of about 10 Kuhn length (equivalent to 50 nucleotides).

Passing to the variabile $x = r - D$ (distance from the surfaces of the two spheres), we construct our probability distribution $f(x)$ as:

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ W(x)\exp(\frac{3}{2}c\frac{D}{d}(\frac{d-x}{d})^2) & \text{if } 0 \leq x < d \\ W(x) & \text{if } x \geq d. \end{cases} \tag{A.3}$$

which can be simply normalized as:

$$g(x) = \frac{f(x)}{\int_0^\infty f(x)dx}. \tag{A.4}$$

It's now straigthforward to define the looping probability as the probability of finding the surfaces of the two beads at the end of the chain within a sufficiently short distance a (choosen as 5 nm in the following, in line with (163)):

$$P(x < a) = \int_0^a g(x')dx'. \tag{A.5}$$

We reported the equations for the case $D >> d$ for the sake of simplicity, but in the numerical estimates reported in the following sections we used the full effective potential of depletion attraction taken from (148).

The appealing feature of this model is that it introduces in a natural way a logarithmic relation between the intron length and the dimensions of spliceosome subcomplexes attached to its ends, if we constrain the system to keep a fixed looping probability.

This can be seen by looking at figure A.3 where we plotted the looping probability for different intron lengths as a function of the diameter of the spheres attached to its ends. If we increase the intron length of an order of magnitude the beads' diameter must be enlarged by a (roughly constant) multiplicative factor in order to obtain the same looping probability. This observation may be used to explain the switch from intron to exon definition as the intron length increases. When the intron length becomes too large the dimensions of merely U1 and U2 subcomplexes is not sufficient to ensure a reasonable looping probability. This does not mean that such a process is forbidden but simply that it would require much longer times. For large enough introns it becomes more probable that the two complexes instead join across the exon (a process mediated again by the depletion attraction), if it is sufficiently short. The complexes constructed across exons can actually result large enough to maintain a suitable looping probability, even in the case of long introns.

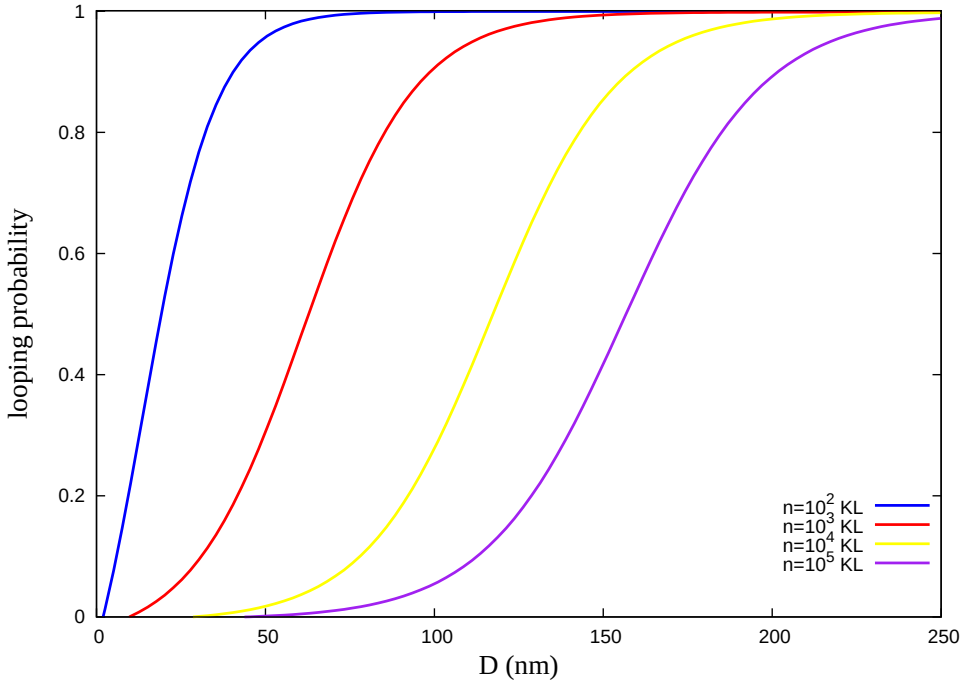

Figure A.3: Looping probability - We show the looping probability (equation A.5 with $a = 5nm$) for different intron lengths as a function of the diameter of the spheres attached to the ends. Following (157; 158) the Kuhn length of the chain was fixed to 5 nt (about 3nm). However it is well known that many regulatory proteins can be bound to the pre-mRNAs and that the latter may fold into rather complex secondary structures. Both these factors have the effect of increasing the stiffness of the pre-mRNA thus increasing its Kuhn length. Unfortunately so far there are no experimental estimates of the Kuhn length in these conditions, so the value derived for ssRNA should be better considered as a lower bound. The diameter of the small crowding molecules is assumed as 5 nm (see (163) and references therein).

A.2.2 Towards a more quantitative model: a compromise between soft and hard hypothesis

Looking at figure A.3 we see that while the model works nicely from a qualitative point of view it predicts intron lengths which are slightly smaller than those actually observed. In fact, in order to make the model more realistic and to be able to obtain also a quantitative agreement with the data, we must take into account two other ingredients. The first one is that pre-mRNAs can be bound to various regulatory proteins which have the effect of increasing their Kuhn length. Unfortunately no direct estimate of the Kuhn length in this conditions exists, thus to obtain the curves reported in figure A.3 we were compelled to use the Kuhn length of pure ssRNAs. Hence the intron length reported in the figure should be better considered as lower bounds.

The second one is that the splicing (sub)complexes are rather far from the hard sphere approximation. If the irregular shape of the molecules allows a snugly fit or if parts of the two subcomplexes can intermingle, the free energy gain will be larger. Again this suggests that our results should be better considered as lower bounds. In this case however we can slightly improve our model and obtain also a reliable upper bound for our looping probability. The maximal relaxation of the hard hypothesis can be achieved considering that the two spheres can fuse with volume conservation (*soft hypothesis*). While we can't actually write the analytical expression of the potential in this "soft beads" case, it's undemanding to calculate the free energy gain obtained by the complete fusion of the two spheres. It's directly related to the portion of volume that becomes available to the crowding molecules:

$$\Delta F_{gain} = cK_BT\left(\frac{2(D+d)^3-(2^{1/3}D+d)^3}{d^3}\right). \tag{A.6}$$

Following again (163) we may at this point assume that the functional dependence on r of the potential is the same as in the hard-hypothesis scenario and that the free energy gain reported in equation A.6 can be a good estimate of the variation of the potential from zero at $r > D + d$ to its maximal absolute value at $r = D$ (i.e. when the beads are in contact). Starting from these resonable assumptions we may write the weighted radial probability distribution as in equation A.2, by simply substituting the maximal free energy gain of the hard beads scenario (which is proportional to $3cD/2d$) with that of equation A.6 :

$$W'_{soft}(r)dr = W(r)\exp(c\left(\frac{2(D+d)^3-(2^{1/3}D+d)^3}{d^3}\right)\left(\frac{D+d-r}{d}\right)^2)dr. \tag{A.7}$$

From this expression it is straightforward to obtain the probability distribution of the end-to-end distance, i.e the corrisponding of equations A.3 and A.4, and obtain curves analogous to those reported in figure A.3.

A.2.3 The intron length distribution of higher Eukaryotes

If the depletion attraction plays a role in exon juxtaposition, the typical length of introns with different splicing fate should be in a range suitable to obtain an high looping probability, given the diameter of the beads attached to their ends. In figure A.4 we report the diameter of the beads needed to have a looping probability of 99%, in the hard sphere hypothesis (blue line) and soft sphere hypothesis (yellow line).

To be more precise, the two colored regions represent the D values, obtained by numerical integrations for different intron lengths, for which $P(x < 5nm) < 0.99$ (see equation A.5), with the radial probability distributions (described by equation A.4), derived starting from equation A.2 (hard-sphere) or from equation A.7 (soft-sphere).

In figure A.4 we also plot two vertical lines corresponding to the intron lengths of the left and right peak of the distribution in figure A.2 as typical values for the introns devoted to intron definition and exon definition respectively. Remarkably enough in both cases the actual dimensions of the splicing complexes (the black dots along the vertical lines in the figure) lie exactly in between the two bounds. Moreover looking at the curves it is easy to see that moving from the first to the second peak, the subcomplexes size must increase roughly of the amount actually observed in the transition from intron definition to exon definition in order to keep the same looping probability. Obviously many other types of specific and elaborate regulation of the splicing dynamic are present in the cell, but the ATP-free depletion attraction could explain the widespread importance of the aspecific intron length variable and the necessity of exon definition when the intron length is increased.

Another interesting extension of the model that we propose occurs if the U1 and U2 subcomplexes can form intermolecular bonding. In this case there would be an additional force driving the intron looping, besides depletion attraction. Unfortunately, even if it is likely that such an interaction is present, there is yet no definitive experimental evidence supporting it and, what is more important for our purposes, the nature and form of its potential is still unclear. In particular even the occurrence of a direct interaction is still under debate: while evidences of such an interaction were proposed in some early papers (164; 165), more recent works suggest instead that intermediate protein(s) are needed to mediate the interaction. For instance the need of the protein Prp5 acting as a bridge between U1 and U2 was recently discussed in (166; 167). In any case, once the interaction potential will be known, it will be rather straigthforward to generalize our model keeping it into account by suitably modifying the Boltzmann factor in equation A.3. Generally speaking, protein-protein interactions are usually short-range (for example an hydrogen bond is formed at distances of the order of 0.1-1 nm) and in a range of energy compatible with the energy gain due to depletion attraction (see section A.2.1). Thus we may safely predict that an additional short-range attraction would only lead to an overall increasing of the looping probability. Qualitatively the effect would be a left translation of the curves in figure A.3 and a lowering of the curves in figure A.4, but this would not change the main results of this paper. As a matter of fact only a contribution of the depletion attraction type, introducing a dependence of the looping probability on the diameter of subcomplexes, could explain the switch from intron definition to exon definition.

A.2.4 Size constraints on introns and exons

Following the idea that the choice of exon or intron definition is related to the looping probability, it is expected that organisms which prevalently use intron definition present a strict constraint on their intron length but not on their exon length, while the opposite is expected for organisms that prevalently use exon definition. As reported in many previously published studies ((137; 140; 144; 145) and reference therein) lower Eukaryotes, that prevalently choose intron definition, present a genomic architecture typified by small introns with flanking exons of variable length. Higher Eukaryotes have the intron length distribution shown in figure A.2, with the vast majority of introns devoted to exon definition (see table A.1), but a strictly conserved distribution of exon length, with a single peack around 100 nt. As shown in the upper right panel of figure A.5, the position of the typical exon length is approximately the same of the length of introns devoted to intron definition. These values, as discussed above, ensure an high probability of juxtapose the two U1 and U2 subcomplexes.
In the case of lower Eukaryotes (three examples in figure A.5) the intron length distribution presents a single narrow peak in a range compatible with high probability of looping. At the same time no constraint on exons are necessary and indeed the distribution of exons' length is quite broad with a long tail towards large lengths.
If the dimensions of merely the U1 and U2 subcomplexes are not enough to ensure an high looping probability across the intron, the exon length is constrained to values that give a sufficient looping probability across the exon, allowing the construction of a larger subcomplex that can then lead to the looping of long introns, as discussed in the previous section.

A.2.5 Cooperative effects

So far we completely neglected the cooperative effects that could arise from the presence of more than two beads on the mRNA string. As discussed in (163), the pairing of more than two beads moves the energetic balance towards the free energy gain. For example, clustering three beads implies three excluded volumes that overlap, but only two loops that have to be closed; four beads give a sixfold free energy gain at the cost of closing only three loops, and so forth. However self avoidance cannot be neglected in this case, as the increasing number of intron chains progressively makes the looping more energetically expensive. As observed in (163) (and reference therein) in three dimensions the entanglement constraints become important when more than eight beads cluster together. Above this threshold the free energy gain/loss ratio starts to decrease, setting the optimal number of beads around eight. In the framework of exon-defined splicing, each bead corresponds to a complex constructed across an exon. Remarkably enough the median value of the number of exons per gene is strongly conserved in higher Eukaryotes (which make an extensive use of exon-defined splicing) and almost coincides with the optimal number of beads in the depletion attraction model (see table A.1 and figure A.6). The same is not true for lower Eukaryotes that

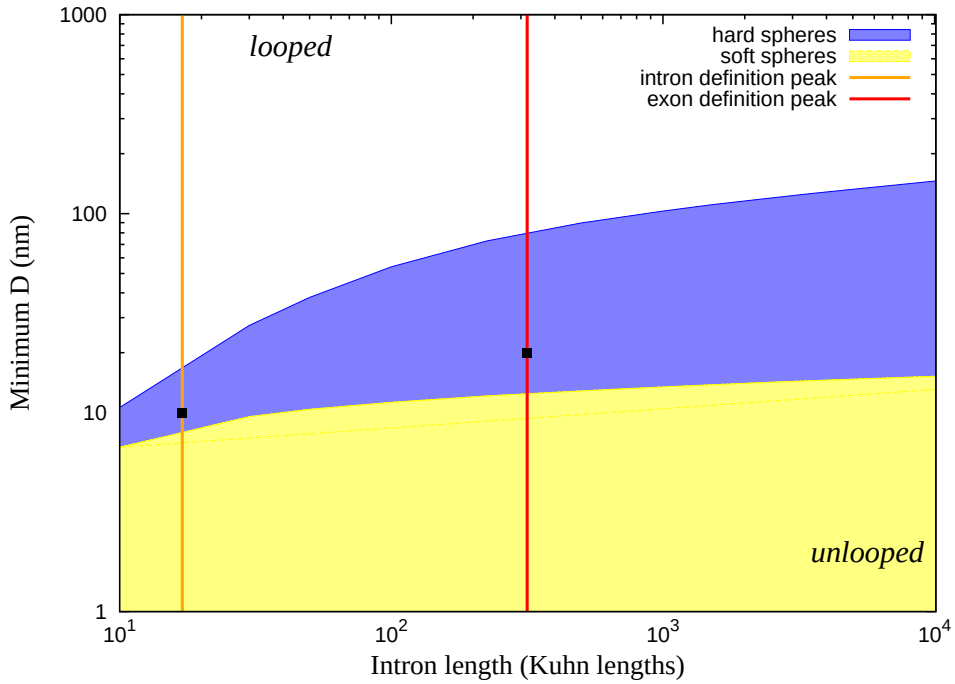

Figure A.4: Comparison with data. - Minimum diameter of beads attached to the ends of an intron needed to obtain a looping probability of 99%. The two curves (blue and yellow) correspond to the looping probability obtained in the hard and soft hypothesis respectively. The two vertical lines (orange and red) correspond to the two peaks in the intron length distribution of H. sapiens (but are quite conserved through different higher Eukaryotes as can be seen in figure A.2). Black squares represent the estimated diameter of spliceosome (sub)complexes for the two corresponding ways of splice-site recognition. While for the intron-definition case estimates for the dimensions of the involved snRNP can be found in literature ((161; 162)), less information is known for the typical size of the complex contructed across exons in exon definition. In the figure we made the (rather conservative) assumption that the diameter of this complex is twice that of the subcomplexes involved in intron-defined splicing.

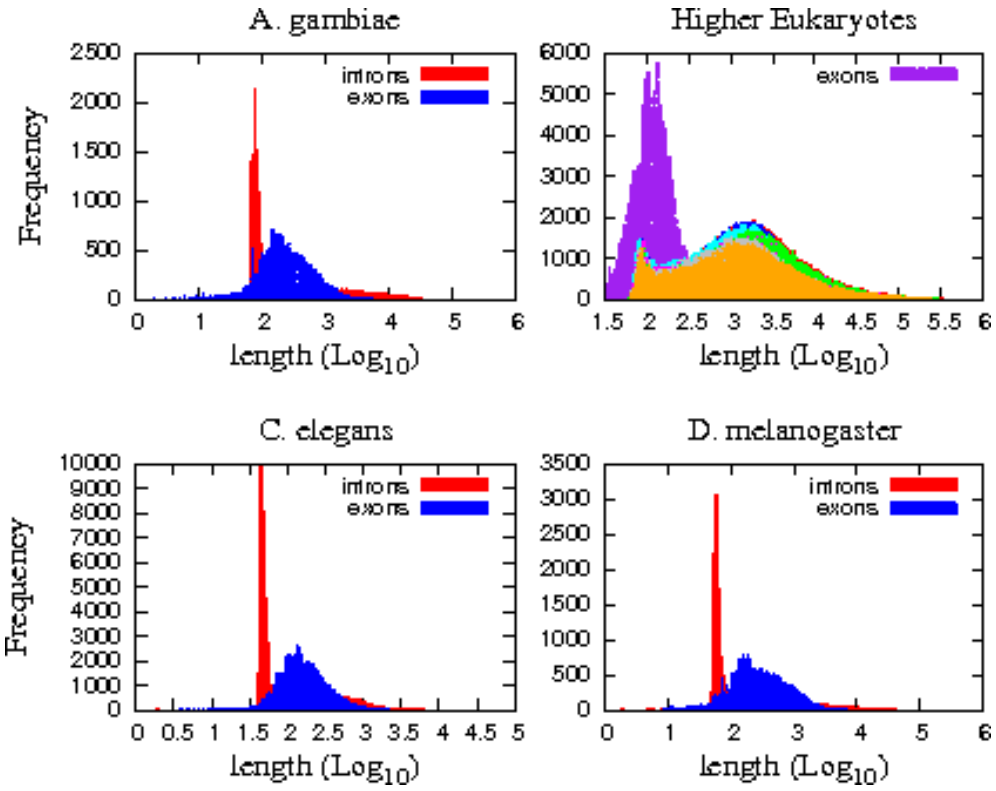

Figure A.5: Intron and exon length distributions in different organisms. - The right upper panel represents the superposition of the intron length distribution of figure A.2 with the exon length distribution of the human genome (but this distribution is again well conserved through different higher Eukaryotes). In the other three panel the superpositions of intron and exon length distributions for three different organisms (*D.melanogaster*,*A.gambiae*,*C.elegans*) that according to (142) prevalently use intron definition

prevalently use intron definition as shown in table A.2 for three model organisms.

Many more refined and energetically costing mechanisms of splicing are surely present in the cell, and many genes present a huge number of exons (up to about 490 in human), but the fact that the typical value is mantained in different organism around, or just below, eight, as predicted by the model, seems to suggests an evolutionary attempt to mantain the number of beads that maximise the depletion attraction effect in exon juxtaposition. Our simple modelization does not ensure the joining of exons in the specific order given by the pre-mRNA transcript, allowing the possibility of scrumbled exons in the mature mRNA. Despite the fact that there are several cases reported in literature of this scrumbling of exons (168; 169; 170), the spliced mRNA usually reproduce the original sequence of exons in the DNA gene, eventually with exon skipping or other splicing variations which however do not affect the exons' order. This is probabily due to the coupling of splicing to transcription by RNA polymerase (171), which naturally introduces a polarity in the transcript and makes the exons available to the splicing machinery in a sequential manner.

Species	Median	Mean of the gaussian fit	Percentage of exon-def introns
Homo sapiens	8	7.7	84
Canis familiaris	8	7.2	78
Pan troglodytes	8	7.8	83
Danio rerio	8	6.7	66
Macaca mulatta	8	6.5	79
Mus musculus	7	6.8	84
Rattus norvegicus	7	6.6	78
Gallus gallus	8	6.7	83
Bos taurus	8	6.8	81

Table A.1: For each species we report: the median (chosen instead of the mean because of the skewness of the distribution) of the overall distribution of the number of exons per gene (first column); the mean of the gaussian fit made over the same distribution, discarding the intronless genes (second column); the percentage of introns which undergo exon-defined splicing according to (142) (third column).

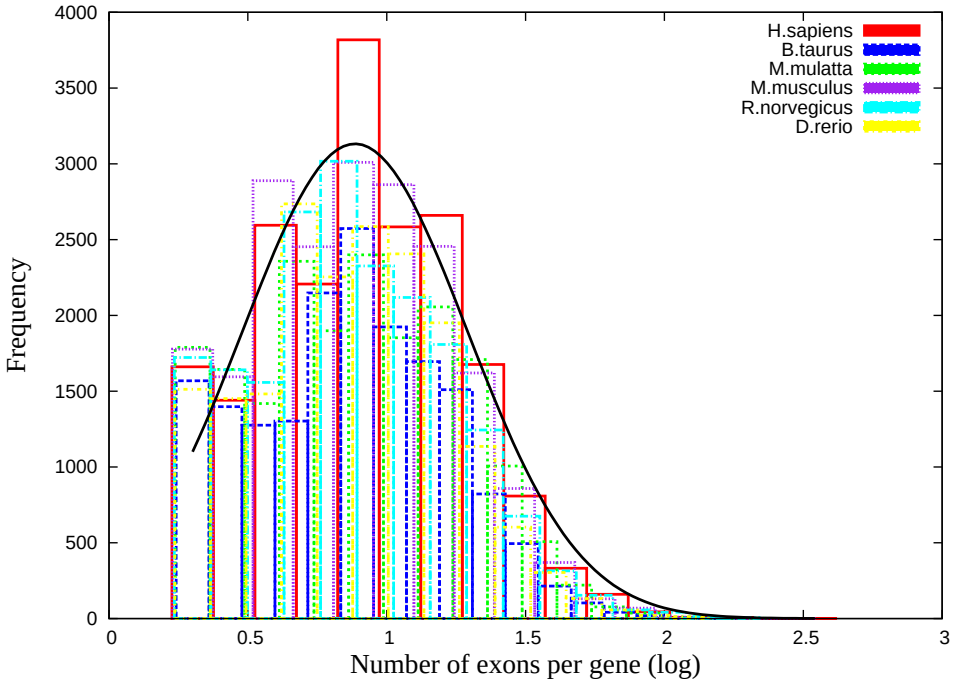

Figure A.6: Cooperative effects - Distribution of the exon number per gene for different higher Eukaryotes (data from (141)). Histograms of frequencies are constructed with a logarithmic binning, discarding the intronless genes. The continuous line is the result of a tentative gaussian fit to the *H. sapiens* distribution.

Species	Median	Percentage of exon-def introns
Anophele Gambiae	3	34
Drosophila melanogaster	3	37
Caenorhabditis elegans	6	40

Table A.2: For each species we report: the median of the distribution of the number of exons per gene (first column); the percentage of introns which undergo exon-defined splicing according to (142) (second column).

A.3 Conclusions and discussion

We presented a model that highlights the possible role of depletion attraction in the splicing process and we showed that this entropic contribution can explain also quantitatively some empirical and bioinformatical observations. Spliceosomal introns can perform various functions (172; 173; 174; 175; 176) and the resulting selective forces to mantain or introduce introns during evolution can explain the genome architecture of higher Eukaryotes, characterized by many introns with a typical large size. The necessity to attain a high regulatory capacity within introns can for example explain the average increase of intron size in the mammal branch of the tree of life (177). At the same time another splice-site recognition modality has been introduced in higher Eukaryotes: the exon definition. In the perspective of our model the exon definition pathway was selected by evolution as the simplest way to mantain a balance between the free energy gain due to depletion attraction and the free energy loss caused by looping longer introns. As shown in section A.2.3 the relation between the dimensions of spliceosome subcomplexes and typical intron lengths is in good agreement with our model predictions. With similar arguments we are able to explain the constraints on exons' length: if the length of introns increases, decreasing their looping probability, the system is compelled to mantain an exons' length suitable for the looping, which is essential to pass to exon definition and obtain diameters of subcomplexes sufficiently large to accomplish the exon juxtaposition.
On the other hand several selective forces can also favour short introns, for example the high fitness of short introns can be due to a reduction of the time and energy cost of transcription and splicing (178), if the conditions favour economy over complexity as in the case of highly expressed genes. Despite the possible selective forces behind - extensively discussed in the case of Drosophila

melanogaster in (179; 180; 181)- usually the introns of lower Eukaryotes have been maintained short by evolution. At the same time, there are no evidences of constraints on exon length, a behaviour again perfectly compatible with our model: the complexes on intron boundaries have a dimension which is sufficient to loop the short introns and proceed with the splicing, so no constraint on exons' length is required. Moreover evolution led to a proliferation of the number of introns in higher Eukaryotes, leading to the genomes with the highest density of introns per gene (182). This contributes significantly to their proteome complexity (138; 183): a gene with many exons can be spliced in many alternative ways to produce different protein products from a single gene. Notwithstanding this, the typical number of exons per gene seems constrained around eight in those species that make an extensive use of exon definition. This coincides precisely with the number that allows an optimal exploitation of the depletion attraction in exon juxtaposition. This result may suggest a trade off between the advantages of a high number of exons - in terms of complexity - and the usage of the uncosting entropic effect of depletion in the splicing process.

References

[1] Alon U An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapman & Hall/Crc Mathematical and Computational Biology.(2006) 1, 18

[2] Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet.**6:** 451-64 (2005) 2, 5, 7, 9, 16, 38

[3] Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell **135** 216 – 226 (2008) 2, 4, 5, 9, 10, 16, 29

[4] Ozbudak EM, Thattai M, Kurster I, Grossman AD, van Oudenaarden A. Regulation of noise in the expression of a single gene. Nat Genet **31:** 69-73 (2002) 2

[5] Elowitz MB, Levine AJ, Siggia ED and Swain PS Stochastic gene expression in a single cell. Science **297:**11831186 (2002) 2, 4, 5

[6] Blake WJ, Kaern M, Cantor CR and Collins JJ. Noise in eukaryotic gene expression. Nature **422:** 663–667 (2003) 2

[7] Raser JM and O'Shea EK Control of Stochasticity in Eukaryotic Gene Expression. Science **304** 1811–1814 (2004) 2, 4

[8] Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature **441:** 840-6 (2006) 2

[9] Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science **329:** 533-8. (2010) 2

[10] Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB. Gene regulation at the single-cell level. Science **307:**1962-5. (2005) 2, 3, 4

[11] Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. Nature **440**:358-62. (2006) 3

[12] Xie XS, Choi PJ, Li GW, Lee NK, Lia G. Single-molecule approach to molecular biology in living bacterial cells. Annu Rev Biophys **37:** 417-44. (2008) 3

[13] Veening JW, Smits WK, Kuipers OP (2008) Bistability, epigenetics, and bet-hedging in bacteria. Annu Rev Microbiol 62:193-210. Bistability, epigenetics, and bet-hedging in bacteria. Annu Rev Microbiol **62:** 193-210. (2008) 4

[14] Davidson CJ and Surette MG Individuality in bacteria. Annu Rev Genet **42:** 253-68 (2008) 4

[15] Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. Nature. **467:** 167-73. (2010) 4

[16] Becskei A, Sraphin B and Luis Serrano L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. EMBO J **20** 2528–2535 (2001) 3, 6

[17] Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in

gene expression. Proc Natl Acad Sci U S A **99:**12795-800 (2002) 4, 9, 38

[18] Shahrezaei V, Ollivier JF, Swain PS. Colored extrinsic fluctuations and stochastic gene expression. Mol Syst Biol **4:**196 (2008) 4, 6, 9, 22, 27, 29, 32, 35, 95

[19] Pedraza JM, van Oudenaarden A. Noise propagation in gene networks. Science **307:** 1965-9 (2005) 4, 21

[20] Volfson D, Marciniak J, Blake WJ, Ostroff N, Tsimring LS and Hasty J Origins of extrinsic variability in eukaryotic gene expression. Nature 439: 861-864.(2006) 4, 6, 21, 27

[21] Tsuru S, Ichinose J, Kashiwagi A, Ying BW, Kaneko K, Yomo T. Noisy cell growth rate leads to fluctuating protein concentration in bacteria. Phys Biol **6:** 036015 (2009) 4

[22] Becskei A, Serrano L. Engineering stability in gene networks by autoregulation. Nature **405:** 590-3 (2000) 6, 22, 95

[23] Thattai M, van Oudenaarden A Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci U S A **98:**8614-9 (2001) 6, 8, 18, 20, 21, 23, 38, 95

[24] Bruggeman FJ, Blthgen N, Westerhoff HV. Noise management by molecular networks. PLoS Comput Biol **5:**e100050 (2009) 6

[25] Simpson ML, Cox CD, Sayler GS. Frequency domain analysis of noise in autoregulated gene circuits. Proc Natl Acad Sci U S A **100:** 4551-6 (2003) 6

[26] Hornung G, Barkai N. Noise propagation and signaling sensitivity in biological networks: a role for positive feedback. PLoS Comput Biol **4:**e8 (2008) 6, 48, 60, 62

[27] Rao CV, Wolf DM, Arkin AP. Control, exploitation and tolerance of intracellular noise. Nature **420:** 231-7 (2002) 6, 9

[28] Isaacs FJ, Hasty J, Cantor CR, Collins JJ. Prediction and measurement of an autoregulatory genetic module. Proc Natl Acad Sci U S A **100:** 7714-9 (2003) 6

[29] Acar M, Becskei A, van Oudenaarden A. Enhancement of cellular memory by reducing stochastic transitions. Nature **435:**228-32 (2005) 6

[30] Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB. Noise minimization in eukaryotic gene expression. PLoS Biol **2:** e137 (2004) 6

[31] Caatay T, Turcotte M, Elowitz MB, Garcia-Ojalvo J, Sel GM. Architecture-dependent noise discriminates functionally analogous differentiation circuits. Cell **139:** 512-22 (2009) 6

[32] van Kampen NG. Stochastic Processes in Physics and Chemistry New York: Elsevier; (1990) 8

[33] Paulsson J Models of Stochastic Gene Expression. Phys. Life Rev **2:** 157-75 (2005) 8, 9, 21

[34] Friedman N, Cai L, Xie XS. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. Phys Rev Lett **97:** 168302 (2006) 8

[35] Shahrezaei V, Swain PS. Analytical distributions for stochastic gene expression. Proc Natl Acad Sci U S A **105:** 17256-61 (2008) 8, 29

[36] Paulsson J. Summing up the noise in gene networks. Nature **427:** 415-8 (2004)

[37] Maheshri N, O'Shea EK. Living with noisy genes: how cells function reliably with inherent variability in gene expression Annu Rev Biophys Biomol Struct **36:** 413-34 (2007) 9, 16

[38] Tnase-Nicola S, Warren PB, ten Wolde PR. Signal detection, modularity, and the correlation between extrinsic and intrinsic noise in biochemical networks. Phys Rev Lett **97:** 068102 (2006) 9, 27

[39] Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. Cell **123:** 1025-36. (2006) 9

[40] Blake WJ, Balzsi G, Kohanski MA, Isaacs FJ, Murphy KF, Kuang Y, Cantor CR, Walt DR, Collins JJ. Phenotypic consequences of promoter-mediated transcriptional noise. Mol Cell **24:** 853-65. (2006) 9

[41] Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. PLoS Biol **4:** e309 (2006) 9

[42] Gillespie DT A general method for numerically simulating the stochastic time evolution of coupled chemmical reactions. J Comput Phys **22:** 403-434. (1976) 9

[43] Flynt AS and Lai EC Biological principles of microRNA-mediated regulation: shared themes amid diversity. Nat Rev Genet **9:** 831-842 (2008) 12, 15, 70

[44] Ambros V The functions of animal microRNAs. Nature **431:** 350-355. (2004) 12

[45] Bartel DP MicroRNAs: genomics, biogenesis, mechanism, and function. Cell **116:** 281-297. (2004) 12, 15, 70

[46] Bushati N and Cohen SN microRNA functions. Annu Rev Cell Dev Biol **23:** 175-205. (2007) 12, 15, 70

[47] Pillai RS, Bhattacharyya SN and Filipowicz W Repression of protein synthesis by miRNAs: how many mechanisms? Trends Cell Biol **17:** 118-126.(2007) 12, 39

[48] Stefani G and Slack FJ Small non-coding RNAs in animal development. Nat Rev Mol Cell Biol **9:** 219-230.(2008) 12

[49] Alvarez-Garcia I and Miska E MicroRNA function in animal development and human disease. Development **132:** 4653-4662.(2005) 12

[50] Esquela-Kerscher A, Slack FJ Oncomirs- microRNAs with a role in cancer. Nat Rev Cancer **6:** 259-269.(2006) 12, 70

[51] Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res **34:** D140-4 (2006) 12

[52] Baskerville S, Bartel DP. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA **11:** 241-7 (2005) 12, 73

[53] Kim YK, Kim VN. Processing of intronic microRNAs EMBO J **26:** 775-83 (2007) 12, 73

[54] Kim VN MicroRNA biogenesis: coordinated cropping and dicing. Nat Rev Mol Cell Biol **6:** 376-85 (2005) 12

[55] Valencia-Sanchez MA, Liu J, Hannon GJ and Parker R Control of translation and mRNA degradation by miRNAs and siRNAs. Genes Dev **20:** 515-524.(2006) 12, 39

[56] Cullen BR. RNAi the natural way. Nat Genet **37:** 1163-5. (2005) 14

[57] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D and Alon U Network motifs: simple building blocks of complex networks. Science **303:** 1538-1542. (2002) 13

[58] Martinez NJ and Walhout AJM The interplay between transcripton factors and microRNAs in genome-scale regulatory networks. BioEssays **31:** 435-445 (2009) 13

[59] Alon U Network motifs: theory and experimental approaches. Nat Rev Genet **8:**450-61 (2007) 13

[60] Shalggi R, Lieber D, Oren M and Pilpel Y Global and local architecture of the mammalian microRNA-transcription factor regulatory network. PLoS Comput Biol **3:** e131.(2007) 13, 15, 58, 64

[61] Tsang J, Zhu J and van Oudenaarden A MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. Mol Cell **26:** 753-767.(2007) 13, 15, 16, 58, 64, 70, 73

[62] Yu X, lin J, Zack DJ, Mendell JT and Qian J Analysis of regulatory network topology reveals functionally distinct classes of microRNAs. Nucleic Acids Res **36:** 6494-6503.(2008) 13, 58, 64

[63] Re A, Cora‘ D, Taverna D and Caselle M Genome-wide survey of microRNA-transcription factor feed-forward regulatory circuits in human. Mol BioSyst **5:** 854-867.(2009) 13, 15, 58, 64, 70

[64] Inui M, Martello G, Piccolo S. MicroRNA control of signal transduction. Nat Rev Mol Cell Biol **11:** 252-63 (2010) 13, 62

[65] Hornstein E and Shomron N Canalization of development by microRNAs. Nat Genet **38:** S20-S2410.1038/ng1803.(2006) 15, 16, 70

[66] Shkumatava A, Stark A, Sive H and Bartel DP Coherent but overlapping expression of microRNAs and their targets during vertebrate development. GenesDev **23:** 466-481.(2009) 15

[67] Bartel DP MicroRNAs: target recognition and regulatory functions. Cell **136:** 215-233. (2009) 15, 70

[68] Stark A, Brennecke J, Bushati N, Russell RB and Cohen SM Animal microRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell **123:** 1133-1146.(2005) 16

[69] Mansfield JH, Harfe BD, Nissen R, Obenauer J, Srineel J, et al. MicroRNA-responsive 'sensor' transgenes uncover hox-like and other developmentally regulated patterns of vertebrate microRNA expression. Nat Genet **36:** 1079-1083. (2004) 16

[70] Hornstein E, Mansfield JH, Yekta S, Hu JKH, Harfe BD, et al. The microRNA mir-196 acts upstream of hoxb8 and shh in limb development. Nature **438:** 671-674. (2005) 16

[71] Karres JS, Hilgers V, Carrera I, Treisman J and Cohen SM The conserved microRNA mir-8 tunes atrophin levels to prevent neurodegeneration in Drosophila. Cell **131:** 136-145 (2007) 16

[72] Wu C, Shen Y and Tang T Evolution under canalization and the dual roles of microRNAs-A hypothesis. Genome Res **19:** 734-743.(2009) 16, 29, 70, 71, 100

[73] Baek D, Villen J, Shin C, Camargo FD, Gygi Sp, Bartel DP The impact of microRNAs on protein output. Nature **455:** 64-71.(2008) 16, 35, 71

[74] Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R and Rajewsky N Widespread changes in protein synthesis induced by microRNAs. Nature **455:** 58-63.(2008) 16, 35, 71

[75] Komorowski M, Miekisz J and Kierzek AM Translational repression contributes grater noise to gene expression than transcriptional repression. Biophys J **96:** 372-84.(2009) 18, 20, 21, 30

[76] Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J and Phillips R. Transcriptional regulation by the numbers: models. Curr Opin Genet Dev **15:** 116-24. (2005) 18

[77] Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P Molecular Biology of the Cell 5th ed. New York: Garland Science.(2008) 18

[78] Levine E, Kuhlman T, Zhang Z, Hwa T Quantitative characteristics of gene regulation by small RNA. PLoS Biology **9:** 229.(2007) 20, 43, 45, 54, 55, 56, 71

[79] Metha P, Goyal S and Wingreen NS A quantitative comparison of sRNA-based and protein-based gene regulation. Mol Syst Biol **4:** 221.(2008) 20, 43, 45, 47, 54

[80] Shimoni Y, Friedlander G, Hetzroni G, Niv G, Altuvia S, Biham O, Margalit H Regulation of gene expression by small non-coding RNAs: a quantitative view. (2007) Mol Syst Biol **3:** 138. 20, 43, 54

[81] Shahrezaei V and Swain PS (2008) Analytical distributions for stochastic gene expression. Proc Natl Acad Sci USA **105:** 17256-17261. 21

[82] Samoilov MS and Arkin AP Deviant effects in molecular reaction pathways. Nat Biotechnol **24:** 1235-1240.(2006) 29

[83] Dublanche Y, Michalodimitralis K, Kummerer N, Foglierini M, Serrano L (2006) Noise in transcription negative feedback loops: simulation and experimental analysis. Mol Syst Biol **2:** 41.(2006) 32

[84] Singh A and Hespanha JP Optimal feedback strength for noise suppression in autoregulatory gene networks. Biophys J **96:** 4013-4023. (2009) 32

[85] Larson DR, Singer RH, Zenklusen D A single molecule view of gene expression. Trends Cell Biol **19:** 630-7 (2009) 38

[86] Li X, Cassidy JJ, Reinke CA, Fischboeck S and Carthew RW A microRNA imparts robustness against environmental fluctuation during development. Cell **137:** 273-282 (2009) 38, 62, 70, 71, 100

[87] Xiong L, Ma Y and Tang L Attenuation of transcriptional bursting in mRNA transport. Phys Biol**7:** 016005 (2010) 38

[88] Filipowicz W, Bhattacharyya SN, Sonenberg N Mechanisms of post-transcriptional regulation by microRNAs: are the answer in sight? Nat Rev Genet **9:** 102-114 (2008) 39

[89] Elf J, Paulsson J, Berg OG, Ehrenberg Mesoscopic kinetics and its applications in protein synthesis. In *Topics in Current Genetics: Systems Biology: Definitions and Perspectives* Alberghina L, Westerhoff HV (eds),pp 95-116. Berlin, Germany: Springer-Verlag. (2005) 45

[90] Mangan S and Alon U Structure and function of the feed-forward loop network motif. Proc Natl Acad Sci USA **100:** 11980-11985 (2003) 48, 75

[91] Arvey A, Larsson E, Sander C, Leslie CS and Marks DS Target mRNA abundance dilutes microRNA and siRNA activity. Mol Syst Biol **6:** 363 (2010) 54, 56, 71

[92] Mukherji S and van Oudenaarden A Synthetic biology: understanding biological design from synthetic circuits. Nat Rev Genet **10:** 859-71 (2009) 59, 62

[93] Shibata T and Fulimoto K Noisy signal amplification in ultrasensitive signal transduction Proc Natl Acad Sci USA **102:** 331-336. (2005) 48, 60

[94] Hooshangi S, Thiberge S, Weiss R Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. Proc Natl Acad Sci U S A **102:** 3581-3586 (2005) 60

[95] Hubbard T, Aken B, Beal K, Ballester B and Caccamo M et al. Ensembl 2007 Nucleic Acids Res **35:** D610-D617 (2007) 64

[96] O'Donnell KA, Wentzel EA, Zeller KI, Dang CV and Mendell JT c-Myc-regulated microRNAs modulate E2F1 expression. Nature **435:** 839-843 (2005) 70

[97] Brock A, Chang H and Huang S Non-genetic heterogeneity – a mutation-independent driving force for the somatic evolution of tumours. Nature Rev Genet **10:** 336-342. (2009) 70

[98] Spencer SL, Gaudet S, Albeck JG, Burke JM and Sorger PK Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature **459:** 428-433. (2009) 70

[99] Heimberg AM, Sempere LF, Moy VN, Donoghue PCJ and Peterson KJ MicroRNAs and the advent of vertebrate morphological complexity. Proc Natl Acad Sci USA **105:** 2946-2950. (2008) 71

[100] Li SC, Tang P, Lin WC. Intronic microRNA: discovery and biological implications DNA Cell Biol. **26:** 195-207 (2007) 73

[101] Saini HK, Griffiths-Jones S, Enright AJ. Genomic analysis of human microRNA transcripts. Proc Natl Acad Sci U S A **104:** 17719-24 (2007) 73

[132] Lin SL, Miller JD, Ying SY Intronic microRNA (miRNA) J Biomed Biotechnol **4:** 26818 (2006) 73

[103] Barik S An intronic microRNA silences genes that are functionally antagonistic to its host gene Nucleic Acids Res **36:** 5232-41 (2008) 73, 98

[104] Lutter D, Marr C, Krumsiek J, Lang EW, Theis FJ Intronic microRNAs support their host genes by mediating synergistic and antagonistic regulatory effects BMC Genomics **11:** 224 (2010) 73, 98

[105] Megraw M, Sethupathy P, Gumireddy K, Jensen ST, Huang Q, Hatzigeorgiou AG Isoform specific gene auto-regulation via miRNAs: a case study on miR-128b and ARPP-21 Theoretical Chemistry accounts **125:** 593-598 (2010) 73

[106] Maca J, Widder S, Sol R. Specialized or flexible feed-forward loop motifs: a question of topology. BMC Syst Biol **3:**84 (2009) 74

[107] Dekel E, Alon U. Optimality and evolutionary tuning of the expression level of a protein. Nature **436:** 588-92 (2005) 74

[108] Rosenfeld N, Alon U. Response delays and the structure of transcription networks J Mol Biol **329:** 645-54 (2003) 75

[109] Rosenfeld N, Elowitz MB, Alon U. Negative autoregulation speeds the response times of transcription networks. J Mol Biol **323:**785-93 (2002) 75, 78, 82

[110] Mangan S, Itzkovitz S, Zaslaver A, Alon U The incoherent feed-forward loop accelerates the response-time of the gal system of Escherichia coli J Mol Biol **356:** 1073-81 (2006) 82

[111] Fan J, Xiaoling Y, Wengong W, Wood W, Becker K, et al. Global analysis of stress-regulated mrna turnover by using cdna arrays. Proc Natl Acad Sci USA **99:** 10611-16 (2002) 75

[112] Chen C, Ezzeddine N, Shyu A Messenger rna half-life measurements in mammalian cells. Methods Enzymol **448:**335-57.(2008) 75

[113] Yang E, van Nimwegen E, Zavolan M, Rajewsky N, Schroeder M, et al. Decay rates of human mrnas: correlation with functional characteristics and sequence attributes. Genome Res **13:** 1863-72 (2003) 75

[114] Sharova L, Sharov A, Nedorezov T, Piao Y, Shaik N, et al. (2009) Database for mrna half-life of 19 977 genes obtained by dna microarray analysis of pluripotent and differentiating mouse embryonic stem cells. DNA Res **16:**45-48.(2009) 75

[115] Yen H, Xu Q, Chou D, Zhao Z, SJ E Global protein stability profiling in mammalian cells. Science **322:** 918-23.(2008) 75

[116] Hargrove J, Schmidt F The role of mrna and protein stability in gene expression. FASEB J **3:** 2360–2370 (1989) 75

[117] van Rooij E, Sutherland L, Qi X, Richardson J, Hill J, et al. (2007) Control of stress-dependent cardiac growth and gene expression by a microrna. Science **316:** 575-9 (2007) 75

[118] Kai Z, Pasquinelli A Microrna assassins: factors that regulate the disappearance of mirnas. Nat Struct Mol Biol **17:** 5-10. (2010) 75

[119] Sethi P, Lukiw WJ. Micro-RNA abundance and stability in human brain: specific alterations in Alzheimer's disease temporal lobe neocortex Neurosci Lett **459:** 100-4 (2009) 75

[120] Chatterjee S, Grosshans H. Active turnover modulates mature microRNA activity in Caenorhabditis elegans Nature **461:** 546-9 (2009) 75

[121] Kollmann M, Lovdok L, Bartholom K, Timmer J, Sourjik V Design principles of a bacterial signalling network Nature **438:** 504-7 (2005) 87

[122] Endres RG, Wingreen NS. Precise adaptation in bacterial chemotaxis through "assistance neighborhoods" Proc Natl Acad Sci U S A **103:** 13040-4 (2006) 87

[123] Matthews HR, Reisert J Calcium, the two-faced messenger of olfactory transduction and adaptation Curr Opin Neurobiol **13:** 469-75 (2003) 87

[124] Ma W, Trusina A, El-Samad H, Lim WA, Tang C Defining network topologies that can achieve biochemical adaptation Cell **138:** 760-73 (2009) 87, 92

[125] Goentoro L, Shoval O, Kirschner MW, Alon U. The incoherent feedforward loop can provide fold-change detection in gene regulation. Mol Cell **36:** 894-9 (2009) 87, 88, 92, 98, 99

[126] Goentoro L, Kirschner MW. Evidence that fold-change, and not absolute level, of beta-catenin dictates Wnt signaling. Mol Cell **36:** 872-84 (2009) 87

[127] Cohen-Saidon C, Cohen AA, Sigal A, Liron Y, Alon U. Dynamics and variability of ERK2 response to EGF in individual living cells. Mol Cell **36:** 885-93 (2009) 87, 98

[128] Shoval O, Goentoro L, Hart Y, Mayo A, Sontag E, Alon U. Fold-change detection and scalar symmetry of sensory input fields. Proc Natl Acad Sci U S A **107:** 15995-6000 (2010) 87

REFERENCES

[129] Khalil AS, Collins JJ Synthetic biology: applications come of age. Synthetic biology: applications come of age. Nat Rev Genet **11:** 367-79 (2010) 100

[130] Rinaudo K, Bleris L, Maddamsetti R, Subramanian S, Weiss R, Benenson Y A universal RNAi-based logic evaluator that operates in mammalian cells. Nat Biotechnol **25:** 795-801 (2007) 100

[131] Deans TL, Cantor CR, Collins JJ A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells. Cell **130:** 363-72 (2007) 100

[132] Lin SL, Chang SJ, Ying SY. Transgene-like animal models using intronic microRNAs. Methods Mol Biol **342:** 321-34 (2006) 73

[133] M.J.Schellenberg et al. Trends Bioch Sci **33:** 243 (2008) 103

[134] M.C. Wahl, C.L. Will and R. Lührmann Cell **136:** 701-718 (2009) 103

[135] L.D. Black Annu Rev Biochem **72:** 291-336 (2003) 103

[136] L. Collins. D. Penny Mol Biol Evol **22:** 1053-1066 (2005) 103

[137] S.M. Berget J Biol Chem **270:** 2411-2414 (1995) 103, 104, 113

[138] G. Ast Nat Rev Genet **5:** 773-782. (2004) 103, 104, 120

[139] Ram O, G. Ast Trends Genet **23:** 5-7.(2007) 103, 104

[140] K.J.Hertel J Biol Chem **283:** 1211-1215.(2008) 103, 104, 108, 113

[141] P.Flicek et al. Nucl Acids Res **36:** D707-D714 (2008) 106, 118

[142] K.L.Fox-Walsh, Y. Dou, B.J. Lam, S.Hung, P.F. Baldi, K.J. Hertel Proc Natl Acad Sci USA **102:** 16176-16181 (2005) 104, 115, 117, 119

[143] L. Collins, D. Penny Mol Biol Evol **23:** 901-910 (2006) 104

[144] A.M. McGuire, M.D. Pearson, D.E. Neafsey, J.E. Galagan Genome Biol **9:** R50 (2008) 104, 113

[145] L.P. Lim, C.B. Burge Proc Natl Acad Sci USA **98:** 11193-11198 (2001) 104, 113

[146] N.J. Sakabe, S.J. Souza BMC Genomics **8:** 59 (2007) 104

[147] D.Marenduzzo, K. Finan, P.R. Cook J.Cell Biol. **175:** 681-686 (2006) 107

[148] S.Asakura and F.Oosawa J Chem Phys **22:** 1255(1954) 107, 109

[149] S. Asakura, F. Oosawa J Polym Sci **33:** 183-192 (1958) 107, 108

[150] P.Attard, G.N.Patey J Chem Phys **92:** 4970. (1990) 107

[151] B.Gotzelmann, R.Evans,S.Dietrich Phys Rev E **57:** 6785-6800.(1998) 107

[152] R.Dickman,P.Attard,V.Simonian J Chem Physics **107:** 205-213.(1997) 107

[153] R.J.Ellis Trends Biochem Sci **26:** 597-604.(2001) 107

[154] A.P.Minton J Biol Chem **256:**10577-10580.(2001) 107

[155] Cantor and Schimmel *Biophysical Chemistry PartIII* (W.H. Freeman and Company, 1980) . 107, 108

[156] P.J. Flory *Statistical mechanics of chain molecules* (Wiley, 1969). 107

[157] K.Rippe Trends Bioch Sci **26:** 733-740 (2001) 107, 110

[158] J.Liphardt, B.Onoa, S.B. Smith,I. Tinoco Jr,C. Bustamante, Science **292:** 733-737 (2001) 107, 110

[159] J. B. Mills, E. Vacano, P. J. Hagerman J Mol Biol **285:** 245-257.(1999) 107

[160] K. Rippe, P.H. von Hippel, J. Langowski Trends Biochem Sci **20:** 500-506.(1995) 107

[161] J.Sperling, M. Azubel,R. Sperling Structure **16:**1605-1615.(2008) 107, 114

[162] B.Kastner, M.Bach, R. Lührmann Proc Natl Acad Sci USA **87:**1710-1714.(1990) 107, 114

[163] D.Marenduzzo, C. Micheletti, P.R. Cook Biophys J **90:** 3712-3721.(2006) 108, 110, 111, 113

[164] I.W.Mattaj, W.J.Habets, W.J.van Venrooij EMBO J **5:** 997-1002.(1986) 112

[165] M.Daugeron, J.Tazi, P. Jeanteur, C. Brunel, G.Cathala Nucl Acids Res **20:** 3625-3630.(1992) 112

[166] Y.Xu, C.M.Newnham, S.Kameoka, T. Huang, M.M.Konarska, C.C. Query EMBO J **23:** 376-385.(2004) 112

[167] G.Dönmez, K.Hartmuth, B.Kastner, C.L.Will, R.Lührmann Molecular Cell **25:** 399-411.(2007) 112

[168] Nigro J.M., Cho K.R., Fearon E.R., Kern S.E., Ruppert J.M., Oliner J.D., Kinzler K.W., Vogelstein B. Cell **64:** 607613 (1991) 116

[169] Cocquerelle C., Daubersies P., Majerus M., Kerckaert J.P., Bailleul B. EMBO J **11:** 10951098 (1992) 116

[170] X. Shao, V. Shepelev,A. Fedorov Bioinformatics **22:** 692-698 (2006) 116

[171] A.R. Kornblihtt, M. de la Mta, J.P. Fededa, M.J. Munoz, G. Nogues RNA **10:** 1489-1498 (2004) 116

[172] S.W. Roy, W. Gilbert Nat Rev Genet **7:** 211-221.(2006) 119

[173] L. Federova, A. Federov Curr Genomics **6:** 589-595. (2005) 119

[174] C. Zhao, T. Hamilton J Biol Chem **282:** 20230-20237.(2007) 119

[175] M. Lynch, A. Kewalramani Mol Biol Evol **20:** 563-571 (2003) 119

[176] M. Lynch Mol Biol Evol **23:** 450-468 (2006) 119

[177] U.Pozzoli, G.Menozzi, G.P. Comi, R. Cagliani,N. Bresolin,M. Sironi Trends Genet **23:** 20-24 (2006) 119

[178] C.I. Castillo-Davis, Mekhedov S.L.,D.L. Hartl, E.V. Koonin, F.A. Kondrashov Nat Genet **31:** 415-418 (2002) 119

[179] G. Marais, P. Nouvellet, P.D. Keightley, B. Charlesworth Genetics **170:**, 481-485(2005) 120

[180] J. Parsch Genetics, **165:** 1843-1851 (2003) 120

[181] D.L. Halligan, P.D. Keightley Genome Res **16** 875884 (2006) 120

[182] Koonin Int J Biochem Cell Biol **2:** 298-306 (2009) 120

[183] B. Modrek, C. Lee Nat Genet **30:** 13-19 (2002) 120

Acknowledgements

First of all, I have to thank prof. Michele Caselle. Three years ago, he introduced me to the amazing world of biology and he gave me the room to steer my projects in the directions I wanted in the following years.
Special thanks to all the current and former members of Caselle's group. In particular, a word of thanks to Carla Bosia that was involved in a lot of the work presented in this thesis, to Ivan Molineris and Davide Cora' that patiently tried to explain to me the rudiments of bioinformatics, and to Luigi Grassi for many stimulating discussions.
Lastly, I would like to thank the friends of taverns and music, my family for their support, and Cristina for being simply Cristina.