UNRESTRICTED

Ph.D. Thesis

A microRNA based genetic clock

Director of studies:

Dr. Diego di Bernardo

External supervisor:

Prof. Mario di Bernardo

Candidate:

Ms. Velia Siciliano

M.Sc. Medical Biotechnology

ARC: Telethon Institute of Genetics and Medicine

YEAR 2011

Date of Submission: 29 August 2011

Date of Award: 13 Febmay 2012.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Contents

LIST OF TABLES v

LIST OF FIGURES v

1 Introduction to Synthetic Biology **6**

1.1 Introduction: from System to Synthetic biology 7

1.2 Synthetic biology 10

2 Synthetic Oscillators **14**

2.1 The logic of circadian clocks 14

2.2 Synthetic oscillators: from bacteria to mammals. 19

2.2.1 The repressilator 19

2.2.2 The first synthetic clock combining positive and negative feedback loop 20

2.2.3 The mammalian genetic oscillator. 21

2.3 Mathematical modeling of natural and synthetic gene networks. 23

2.3.1 Model derivation approaches 25

2.3.2 Modeling approaches 28

3 Construction and modelling of an inducible positive feedback loop stably integrated in a mammalian cell-line. 30

3.0.3 Rational construction and integration of a PFL and of corresponding control network (NOPFL) in CHO cells. 32

3.0.4 Determination of the reporter protein degradation. . . 35

3.0.5 Mathematical modeling of the dynamic behavior of PFL and NOPFL networks. 37

3.0.6 Experimental observation of dynamics properties of PFL and NOPFL networks 39

3.0.7 Derivation of the model parameters 42

3.0.8 Conclusions 48

4 Construction of a post-transcriptional negative feedback loop in mammalian cells 60

4.1 Introduction 62

4.2 Mechanism of RNAi-based gene silencing 63

4.3 Modeling RNA interference in mammalian cells 66

4.4 Design of a microRNA-based negative feedback loop 73

4.4.1 Conclusion 78

5 Synthesis and Analysis of an amplified negative feedback loop oscillator in mammalian cells 86

5.1 Construction of the oscillator in mammalian cells. 87
5.1.1 Repression by transcriptional control 88
5.1.2 Repression by dimerization 91
5.1.3 Repression by proteolysis 92
5.2 A microRNA based oscillator. 92
5.3 Mathematical modeling of synthetic oscillator 99
5.4 Experimental investigation of dynamic properties of the network. 114

6 Synthesis and Analysis of a delayed amplified negative feedback loop oscillator in mammalian cells 124
6.1 Design of a three stage genetic oscillator. 126
6.2 Mathematical modeling of three step oscillator. 135

7 Conclusion 144

A Vector Maps 149

B Publication 161

List of Tables

3.1 Parameters identified after the fitting procedures: parameters values as well as standard deviation are reported for each parameter. 44

5.1 Parameter for the mathematical model; undamped oscillations. 111

6.1 Parameter for the first model; undamped oscillations. 139

List of Figures

2.1 **Schematic representation of the mammalian circadian clock.** 17

2.2 **The Repressilator.** A negative feedback loop composed by three repressors, inhibiting each other. The GFP, used as read-out of the network, is controlled by the promoter responsive to the TetR repressor. 20

2.3 **Genetic clock by combination of positive and negative feedback loop in *E. coli*.** The *Plac-glnAp2* promoter drives the expression of the activator NRI that, upon autophosphorilation (NRI-P), regulates positively itself (positive feedback loop) and induces the negative feedback loop by activating the expression of the repressor Lac I after binding the the *PglnKp* promoter. Lac I represses the *Plac-glnAp2* promoter and also the expression of β-galactosidase, whose oscillations occur with a period of 20h. 22

2.4 **Mammalian oscillator.** The tetracycline-dependent transactivator tTA is driven in sense orientation by the tetracycline responsive promoter *phCMV*-1*, and in the antisense orientation by the streptogramin-responsive (*pPIR*) promoter. tTA expression results in increased GFP and PIT levels. The PIT induces the expression of antisense tTA via binding to the *pPIR* promoter; thus the levels of tTA, PIR and GFP decrease. The whole combination of positive and time-delaying negative feedback loop mediate robust oscillations. 24

3.1 **Design of the expression system.** (A) PFL: the promoter *CMV-TET* consists of seven direct repeats of a 42-bp sequence containing the tet operator sequences (tetO), located just upstream of the minimal *CMV* promoter (PminCMV). The Tetracycline-controlled transactivator tTA derives from the addition of the VP16 activation domain to the transcriptional repressor TetR to optimize the expression in mammalian cells. The d2EYFP is the destabilised yellow-green variant of enhanced green fluorescent protein. (B) NOPFL: the *CMV* promoter drives the expression of the tTA, which in turns drives the transcription of the d2EYFP from the *CMV-TET* promoter. (Inset) RealTime PCR performed on DNA extracted from PFL and NOPLF cells shows that the DNA levels of tTA and d2EYFP are comparable among the two clonal cell populations. 34

3.2 **Degradation kinetics of d2EYFP.** *CMV-TET*-d2EYFP stably integrated CHO AA8 TET-OFF cells were treated at t=0 with different concentrations of Cyclohexamide (CHX): panel A, 10μg/mL; panel B, 50μg/mL; panel C, 100μg/mL; panel D, 500μg/mL. Fluorescence intensity was followed up to 750 minutes. Sampling time is equal to 15 min. The thin line represents the mean over biological triplicates; the shaded area represents the standard error. Experimental data were used to fit the exponential decay of d2EYFP protein levels, and thus to derive its half-life (τ). 36

3.3 **Experimental and simulated switch off time-course across the PFL and NPFL cell population.** Experimental data (thin lines) and model simulations (thick lines) were reported for the PFL (left) and NOPFL (right) cells. Shaded areas represent standard deviations from replicate experiments. 40

3.4 **Replicates of the experimental time-courses across the PFL and NPFL cell population.** Replicates of the experimental time-courses for the PFL (left) and NOPFL (right) cells. Each line in each panel represent the average fluorescence intensity across the cell population in one switch-off experiment. 41

3.5 **Switch off time τ_{off} for varying Doxycycline concentrations from experimental data and model predictions.** The model predictions for the switch off times τ_{off} are shown for PFL (dashed thick line) and NOPFL (solid line). Experimental quantification of the τ_{off} for PFL and NOPFL models have been reported for comparison with + and × respectively. Observe that the experimental τ_{off} for the PFL at 1ng/mL and 10ng/mL could not be estimated since the PFL is not switching off in the experimental observation time (43h). 47

3.6 **Phase portrait of the PFL model.** The tTA-d2EYFP mRNA concentration (y axis) has been plotted against tTA protein concentration (x axis). Varying Doxycycline concentrations (1 ng/mL through 1 μg/mL) were used to investigate the dependence of the two stable equilibria ("ON" and "OFF" in the graph) on the amplitude of the input. The shape and dimensions of the two basins of attraction (the set of initial conditions ending up in one of the two stable steady states) can be studied with the same technique: in this figure the grey shaded area represents the basin of attraction of the "OFF" equilibrium for Doxycycline= 0 nM. 49

4.1 **Schematic representation of RNA interference in a mammalian cell.** 65

4.2 **Ratio of *EGFP* mRNA levels between cells transfected with the siRNAs specific for EGFP, and negative control cells, transfected with a non-specific siRNAs, measured 48 hours after transfection.** Error-bars represent the standard-error from three biological replicates for each point. The x-axis reports the different quantities of siRNA oligomers tested. mRNA levels were measured using real-time PCR. The error-bars have the length of one standard error. 67

4.3 **Ratio of EGFP protein levels between cells transfected with the siRNAs specific for EGFP, and negative control cells, transfected with a non-specific siRNAs, measured 60 hours after transfection.** Error-bars represent the standard-error from three biological replicates for each point. The x-axis reports the different quantities of siRNA oligomers tested. Protein levels were measured using FACS analysis quantifying EGFP protein fluorescence. The error-bars have the length of one standard error. 67

4.4 **Ratio of *tTA* mRNA levels between cells transfected with the siRNAs specific for tTA, and negative control cells, transfected with a non-specific siRNAs, measured 48 hours after transfection.** Error-bars represent the standard-error from three biological replicates for each point. The x-axis reports the different quantities of siRNA oligomers tested. mRNA levels were measured using real-time PCR. The error-bars have the length of one standard error. 68

4.5 **Numerical fitting of the model on the in vitro experimental results on mRNA *EGFP*.** 71

4.6 **Numerical fitting of the model on the in vitro experimental results on protein EGFP.** 71

4.7 **Numerical fitting of the model on the in vitro experimental results on mRNA *tTA*.** 72

4.8 **Schematic representation of the miR223-based negative feedback loop.** The expression of the *EF1α* intron including miR223 and of the *mCherry* is driven by the *CMV-TET* promoter. WPRE, woodchuck hepatitis virus post-transcriptional element; SD, splice donor site; SA, splice acceptor site. 75

4.9 **HEK cells were co-transfected with the lentiviral vector carrying the miR223 controlled by a *CMV* promoter and a plasmid containing an *EGFP* with four tags for the binding of the miR223 (EGFP tag).** As negative control we used the same plasmid expressing the EGFP but lacking the tags (EGFP no tag). The efficiency of the knocking down due to the microRNA effect is shown in the real-time PCR. 76

4.10 **miR223 plasmid transfection in PFL-HEK cells.** PFL-HEK cells were transfected with the *CMV*-miR223 plasmid. The mRNA level of the d2EYFP were compared to a mock control (same lentiviral plasmid not carrying the *miR223-mCherry* gene expression cassette), as shown in the real-time PCR. . . . 77

5.1 **Genetic clock by combination of activator and repressor modules in *E.coli*.** 90

5.2 **Amplified negative feedback loop topology**. Gene X activates itself and geneY, while gene Y represses transcription from X. The topology is conserved for the respression by dimerization, and repression by proteolysis. 93

5.3 **Mammalian oscillator.** (A) Scheme of the genetic oscillator. The tTA transcriptional activator binds the *CMV-TET* promoter auto-regulating itself, together with the expression of the d2EYFP fluorescent protein (positive feedback loop). The activation of the *CMV-TET* promoter triggers the miR223-mCherry production and the consequent degradation of the *tTA-d2EYFP* mRNA through the binding to four tags placed at the 3' end of the construct. (B) Autoregulated *CMV-TET* promoter-driven tTA expression triggers d2EYFP production (1). As the tTA and d2EYFP reach the peak (2) the miR223 is transcribed together with the mCherry (3) resulting in the downregulation of tTA and d2EYFP (4). 95

5.4 **Selection of CHO cells stably expressing the positive and negative feedback loops.** (1)CHO cells were infected with a virus carrying the positive feedback loop (*CMV-TET-tTA-d2EYFP*). (2) CHO cells positive to viral infection were selected either with blasticidine antibiotic or with FACS sorting. (3)Serial dilution of CHO PFL cells were made to generate monoclonal population (CHO-PFL cells). (4-5) CHO-PFL clonal population number 2 was infected with the lentivirus harboring the NFL; double infected cells were sorted via FACS according to the contemporary expression of green and red fluorescence . For simplicity they were called Red on Green cells (RoG cells). (7-8) Following serial dilutions, clonal populations of RoG cells were generated. 96

5.5 **Real-time PCR of genomic DNA (gDNA) and mRNA (cDNA) extracted from the monoclonal population of double infected RoG CHO cells.** 98

5.6 **Real-time PCR of genomic DNA (gDNA) and mRNA (cDNA) extracted from the monoclonal population of double infected GoR HEK 293 cells.** 99

5.7 **Bifurcation diagram for basal activity of *CMV-TET* promoter (α_1) *versus* degradation rate of miR223 (δ).** . . 103

5.8 **Bifurcation diagram for basal activity of *CMV-TET* promoter (α_1) *versus* maximal rate of silencing of miR223 (λ).**Big stars in the diagram correspond a longer period of the oscillations. 104

5.9 **Bifurcation diagram for degradation rate of miR223 (δ) *versus* maximal rate of silencing of miR223 (λ).** Big stars in the diagram correspond a longer period of the oscillations. 105

5.10 **Bifurcation diagram for hill constant for d2EYFP equation (K_3) *versus* hill costant for miR223 equation (K_1).** Big stars in the diagram correspond a longer period of the oscillations. 106

5.11 **Bifurcation diagram for hill costant for d2EYFP equation (K_3) *versus* folding rate for miR223 (K_D).** Big stars in the diagram correspond a longer period of the oscillations. . 107

5.12 **Bifurcation diagram for folding rate for miR223 (K_D) *versus* maximal rate of silencing of miR223 (λ).** Big stars in the diagram correspond a longer period of the oscillations. 108

5.13 **Zoom of the folded and unfolded form of mCherry on the simulation of all equations for the second model with parameters of Table 5.1. Due to the stability of the protein, mCherry levels are up to 40 times more than d2EYFP protein levels.** 109

5.14 **Simulation for the second model with parameters of Table 5.1.** 110

5.15 ***In silico* simulation of a mathematical model lacking of the equations that discriminate the inactive and active form of the miR223.** The network does no more behave as an oscillator. 113

5.16 **Negative control of the oscillator in CHO cells.** CHO NO PFL cells were infected with the virus carrying the negative feedback loop. The resulting cell line was used as control in our experiments since cells lack of the tag for the miR223; therefore, variances in the level of fluorescence, are due only to the intrinsic cell variability. 115

5.17 **CHO RoG cells show oscillatory behavior.** d2EYFP fluorescence was detected for up to 4000 minutes using a time lapse microscopy. Oscillations occur with a period of about 1000 minutes. (A) CHO-Osc. (B) CHO-nOsc. 117

5.18 **HEK 293 GoR cells show oscillatory behavior.** d2EYFP fluorescence was detected for up to 4000 minutes using a time lapse microscopy. Oscillations occur with a period of about 1000 minutes. (A) HEK GoR cells. (B) HEK PFL cells. 118

6.1 **Three stage genetic clock.** 127

6.2 **RealTime PCR of genomic DNA (gDNA) and cDNA of CHO cells infected with the lentivirus carrying the dGAL4-VP16 under the control of *CMV-TET* promoter.** Third column: transfection of infected CHO cells with pTET-OFF vector expressing the tTA to test the activation of the system. 129

6.3 **Co-trasfections of CHO-GAL4 cells with a pTET-OFF vector expressing the tTA transactivator, and a vector where the miR223 is driven by the *UAS* promoter.** Upon the production of GAL4 induced by the tTA, also the mRNA levels of *mCherry* increase, compared to the negative control where we transfected a vector of comparable size, but lacking of the activator of GAL4 expression 130

6.4 **RealTime PCR of clonal population of CHO cells infected with the lentivirus carrying the dGAL4-VP16 under the control of *CMV-TET* promoter.** Side by side are reported *GAL4-VP16* expression values in absence and presence of the activator tTA. 131

6.5 **RealTime PCR of cDNA of clonal population of CHO cells infected with the PFL.** 133

6.6 **RealTime PCR of genomic DNA of clonal population of CHO cells infected with the PFL.** 134

6.7 **Simulation for the three stage oscillator.** 138

6.8 **Quantitative analysis of the period of the oscillations changing the degradation rate for *dGal4* mRNA (d_G).** 140

6.9 **Quantitative analysis of the period of the oscillations changing the hill costant for *Gal4-UAS* promoter (K_{UAS}).** 141

A.1 **pENTR-CMV-TET vector.** Flanking are the attL4 and attR1 recombination sites. 150

A.2 **pENTR-CMV vector.** Flanking are the attL4 and attR1 recombination sites. 151

A.3 **pENTR-UAS vector.** Flanking are the attL4 and attR1 recombination sites. 152

A.4 **pENTR-PFL vector.** Flanking the tTA-IRES-d2EYFP-WPRE-223TAG are the attL1 and attL2 recombination sites. 153

A.5 **pENTR-NFL vector.** Flanking the miR223-mCherry are the attL1 and attL2 recombination sites. 154

A.6 **pENTR-d2EYFP vector.** Flanking the d2EYFP are the attL1 and attL2 recombination sites. 155

A.7 **pENTR-GAL4-VP16 vector.** Flanking the gene cassette are the attL1 and attL2 recombination sites. 156

A.8 **pDEST-PFL vector.** From the recombination reaction between three vectors we obtained the positive feedback loop . . 157

A.9 **pDEST-NFL vector.** Obtained from recombination reaction between the plasmids carrying promoter, gene expression cassette e viral backbone. 158

A.10 **pDEST-GAL4 vector.** Obtained from recombination reaction between the plasmids carrying promoter, gene expression cassette e viral backbone. 159

A.11 **pDEST-UAS-miR223-mCherry vector.** Obtained from recombination reaction between the plasmids carrying promoter, gene expression cassette e viral backbone. 160

Acknowledgement

A mia Madre. Tutto l'amore che c'e'.

Mi sono sempre domandata quali fossero i miei limiti, fin dove potessi spingermi. Oggi posso dire di non saperlo ancora, la voglia di mettermi alla prova e' ancora troppo forte. Ma ci sono molte persone che hanno fatto parte di un percorso di crescita che vorrei oggi ringraziare. Saro' breve, perche' pare che l'ultima volta abbia parlato davvero troppo! In questi anni trascorsi al Tigem i confronti e gli scontri con Diego sono stati estremamente formativi, sia da un punto di vista personale che professionale. La mia forza e' la mia caparbieta' mi hai detto una sera. Tu l'hai riconosciuta tutta, e tra discussioni, risate e sfoghi nel tuo ufficio, nei corridoi e in lab l'hai fatta emergere. Grazie. Grazie a Lucia e Giulia, senza le quali l'inizio di quest'esperienza, e l'esperienza tutta, non sarebbe stata la stessa. A Giusy, con cui da forti discordanze, birre, e chiacchierate, e' nata una sincera amicizia. Grazie. Ringrazio Vincenzo, Francesco, Maury, Filippo,Mario, Santosh, Luisa, Gopu, Margherita, Alberto, Mukesh, Annamaria, Giampiero e Barbara

per le tavolate, le bevute, le innumerevoli pacche sulle spalle all'ennesimo esperimento non riuscito...siete stati fantastici! Alda la tua forza mi ha sempre spiazzata, sei un mito! A tutta la nuova generazione del gruppo di Bernardo: ai balli con Chiara sulla colonna sonora di Dirty dancing, alla saggezza e sottile ironia di Roberto, alle grasse risate di Simona...che belli siete! A Stefania sempre brillante nel distrarmi con le sue strampalate domande, a Gianfranco pronto a caricarmi sulle spalle, a Gennaro per la pazienza con cui sopportava le mie dormite durante i film. Grazie ad Imma, per la delicatezza con cui mi sostiene, e per la determinazione. Quando avrai piena consapevolezza di te sarai inarrestabile! A Nicoletta, dura di capa e fragile di cuore...grazie per il sostegno, gli sfoghi, i pianti, la rabbia e i sorrisi. Sei un'amica preziosa. A Rossella, senza cui io sicuramente non sarei quella che sono oggi. Grazie per la sincerita' della tua amicizia, grazie perche' mi hai preso per i capelli quando non riuscivo a vedere altro che il fondo. GRAZIE! A Michele per la buona volonta' nel barcamenarsi in questo tesoro di femmina (che sarei io)...la tua presenza e' stata fondamentale in tutte quelle sere in cui, io influenzata cercavo di capire cosa avrei dovuto difendere all'esame (!!!!), mentre tu, cadendo addormentato mi dicevi: "non ti preoccupare sarai bravissima"..il mio cavalier russante! Ed infine grazie alla mia famiglia, i miei zii ai miei nonni, alla mia cuginetta Nunzia. A mia sorella, che e' il mio piu' grande amore, a mio padre che mi ha sempre sostenuta nei miei progetti (anche lui si e' sempre vantato della mia caparbieta'!), a mia madre che e' la mia luce. Il cuore trabocca di troppo amore nel pensarvi, grazie! E infine, un ultimo grazie: grazie

a me! Tutto sommato avro' fatto qualcosa anche io.. no??!!

Velia

Contributions

Maria Nicoletta Moretti, Chiara Fracassi *Clonal population generation, biological validation of reporter half-life.*

Filippo Menolascina, Lucia Marucci, Immacolata Garzilli *Mathematical modeling, data simulation.*

Chapter 1

Introduction to Synthetic Biology

1.1 Introduction: from System to Synthetic biology

'...the pluralism of causes and effects in biological networks is better addressed by observing, through quantitative measures, multiple components simultaneously and by rigorous data integration with mathematical models'.

[71]

Over the last decades, the standard way of studying a biological function, or the underlying molecular mechanisms responsible for a disease, by focusing on a single gene has changed; the so called 'reductionist approach' according to which the biological systems are divided into their smallest possible parts, analyzed separately, is now taken over by modern biology. Most biological functions depends on the interactions among hundreds of different molecular species orchestrating the complex processes needed to sustain life.

The advent of new technologies transformed the modern concept of a biological process into a "system" rather than the sum of its single components; thus, one of the main challenges in the post-genomic research era is to find methods to interpret and to extract the huge amount of information

produced by the high-troughput techniques which integrate the single gene function to the entire process it belongs to [23].

Two new disciplines, Systems and Synthetic biology work in a complementary way to advance our understanding of how life works. Systems biology aims at developing a formal understanding of biological processes through the development of quantitative mathematical models. Systems biology takes advantage of quantitative sciences such as physics, engineering and computer science, and acts in two possible directions; in the 'bottom up' approach, well known biological processes are quantitatively described via mathematical and logical formalisms; thus the process is represented by a predictive model, which is then validated by *in silico* simulations. For example, this approach allows to depict a known biological pathway as a network of interactions between genes, proteins, and metabolites. Describing this 'prior knowledge' using a mathematical model, allows its behavior to be explored and analysed via computer simulations and mathematical analysis, and then tested *in vivo*. On the contrary, the 'top down' approach consists in uncovering the network of gene regulatory interactions (gene networks) of an almost unknown biological process. This method is called 'reverse engineering', and is based on high-troughtput gene expression profiling in different conditions [28]. The gene network model identified via reverse-engineering is usually not a quantitative model, as in the case of the bottom-up approach, but a qualitative, graphical representation of the likely regulatory and physical

interactions occuring among genes and proteins.

The payoff for systems biology is not only a quantitative representation, and hence a better undersanding, of how biological pathways work, but also the empowerment to design new and improved biological functions via 'synthetic biology' [23]; this refers to the possibility of engineering and combining well characterized biological parts in order to create new synthetic systems that overcome the complexity, cross-talking and non-modularity of natural systems.

1.2 Synthetic biology

'The reductionist approach has successfully identified most of the components and many of the interactions but, unfortunately, offers no convincing concepts or methods to understand how system properties emerge...'

[71]

When the Human Genome Project revealed that our genome contains the same number of genes of *Drosophila melanogaster*, this astounding finding begged to the question: how it was possible that one organism looks like a fly and the other one likes a human? One hypothesis was that non-protein-coding sequences were responsible for such a complexity; but a decade of research has put forward the rather different idea that the determinant of organismal complexity is not the length of 'parts list', but how these 'parts' fit together [64].

Synthetic Biology can be defined as the engineering of biology. The core aim of this discipline is to develop and apply engineering approaches,to build new synthetic pathways (or circuits as they are referred to in this discipline) using well characterized biological 'parts'. The aims are two-fold: (1) to uncover the design principles of natural biological systems through the design of very simple synthetic circuit; (2) to engineer new functions in the cell for

biotechnological or medical applications.

Up to now, two major goals have been actively investigated: the building of new biological networks in the cell that perform a specific task (e.g. periodic expression of a gene [38] or genetic switching [40]), and the modification of networks that occur in nature in order to achieve some desired functionalities (e.g. production of a specific compound useful for medical applications [70]).

The first challenge for synthetic biologists is to assemble a synthetic circuit with a predictable behavior. In order to achieve this task, there is the need to precisely characterise the steps involved in gene expression, such as the processes of transcription and translation, in order to construct the apprioriate 'parts', such as a promoter with a given transcription rate, or a protein with a given degradation rate.

Earlier studies were focused on the construction of systems in simple prokaryotes, in which the qualitative notions of transcriptional activation, repression and post-transcriptional regulation could be quantified; these constructs were based on the combinatorial promoter libraries driving the expression of reporter genes [44].

Later on an increasing interests in the post-transcriptional and post-translational regulation triggered the design of synthetic circuits harboring elements from the RNA interference pathway, aptamers, riboswitches, or mutated ribosome binding sites [13, 88, 82, 30].

All these basic elements were manipulated so that they could work synergistically toward the desired goal, forming small modules including switches,

pulse generators, time-delayed circuits, oscillators and so on [40, 51, 7, 87, 77, 81].

Synthetic Biology is an interdisciplinary area requiring a deep synergy between biology, biotechnology and nanotechnology on one side and mathematical modelling, information technology and control theory on the other. Such combination of disciplines is needed to construct robust and predictable synthetic networks.

Aim of the project

This thesis focuses on the design, construction and stable integration in mammalian cells of a natural microRNA-based genetic oscillator. This will help both in better understanding the rules underlying the periodic expression of genes observed in major biological processes, such as the circadian clock and cell-cycle, as well as, in generating new tools to probe and investigate the function of a gene in a cell, by allowing not only its over-expression or knock-down, but also its cyclic expression.

Chapter 2

Synthetic Oscillators

2.1 The logic of circadian clocks

'Life is a cyclical chemical process... Birth to death, a cycle, and there are cycles within the cycles- circannual rhythms, menstrual cycles, and daily 24h or circadian cycles.'

[35]

Oscillations can occur when a system is perturbed from its dynamical equilibrium and restoration forces try to return the system to the equilibrium state [8]. However, oscillations can also be the 'natural' equilibrium state of the system, which mathematician call the 'limit-cycle', which consists of

perpertual, robust, periodic oscillations, to which the systems returns to, even if perturbations try to drive the system out of it.

Oscillations are everywhere; they can be found in several physiological processes such as controlling key repair [52], metabolic and signaling pathways [27], somite segmentation during vertebrate embryogenesis [59], the cardiac pacemakers [83, 84], the hormone-controlled female reproduction cycle.

The periodic expression of genes underlies the functionality of one of the most striking and general processes found also in every mammalian cell: the circadian clock.

Circadian clocks are essential for coordinating the physiology of the whole organism. Circadian rhythms are driven by biological clocks, with two key characteristics: first, they are temperature-compensated, meaning that they are not affected by alterations of temperature (they do not run slower at lower temperature nor speed up at higher temperature); second, they can synchronize to temporally relevant stimuli such as light, temperature or feeding schedules; this is why their definition of internal time becomes predictive of external (solar) time.

The molecular basis of this process is a conserved transcriptional-translational autoregulatory loop with a product which slows down the rate of the process itself with a certain delay (negative feedback loop), and a source of activation, which keeps the oscillator from winding down (positive feedback loop), generating molecular oscillations of 'clock genes' at the cellular level.

In mammals, the circadian system is organized in a hierarchical manner in which a master pacemaker in the suprachiasmatic nucleus (SCN) regulates downstream oscillators in peripheral tissues. The SCN coordinates independent peripheral oscillators so that a coherent rhythm is orchestrated at the organismal level. The clock mechanisms is, again, based on a combination of positive and negative transcriptional/translational feedback loops that drive rhythmic, 24h expression patterns of core clock components. The core clock component consists of genes whose protein products are responsible for the generation and regulation of circadian rhythms within individual cells throughout the organism.

As shown in Figure 2.1, in the primary feedback loop of the circadian clock, the positive elements include members of the basic helix-loop-helix (bHLH)-PAS (Period-Arnt-Single-minded) transcription factor family, CLOCK and BMAL1. CLOCK and BMAL1 heterodimerize and initiate transcription of target genes containing E-box cis-regulatory enhancer sequences, including *Period* (in mice, *Per1, Per2 and Per3*) and *Cryptochrome* (*Cry1* and *Cry2*). Then the heterodimerization of PER:CRY and their translocation back to the nucleus induces the repression of their own transcription by acting on the CLOCK:BMAL1 complex, thus achieving the delayed negative feedback loop.

Another regulatory loop is induced by CLOCK:BMAL1 heterodimers activating transcription of retinoic acid-related orphan nuclear receptors, *Rev-erbα* and *Rorα*. REV-ERBα and RORα compete to bind retinoic acid-related

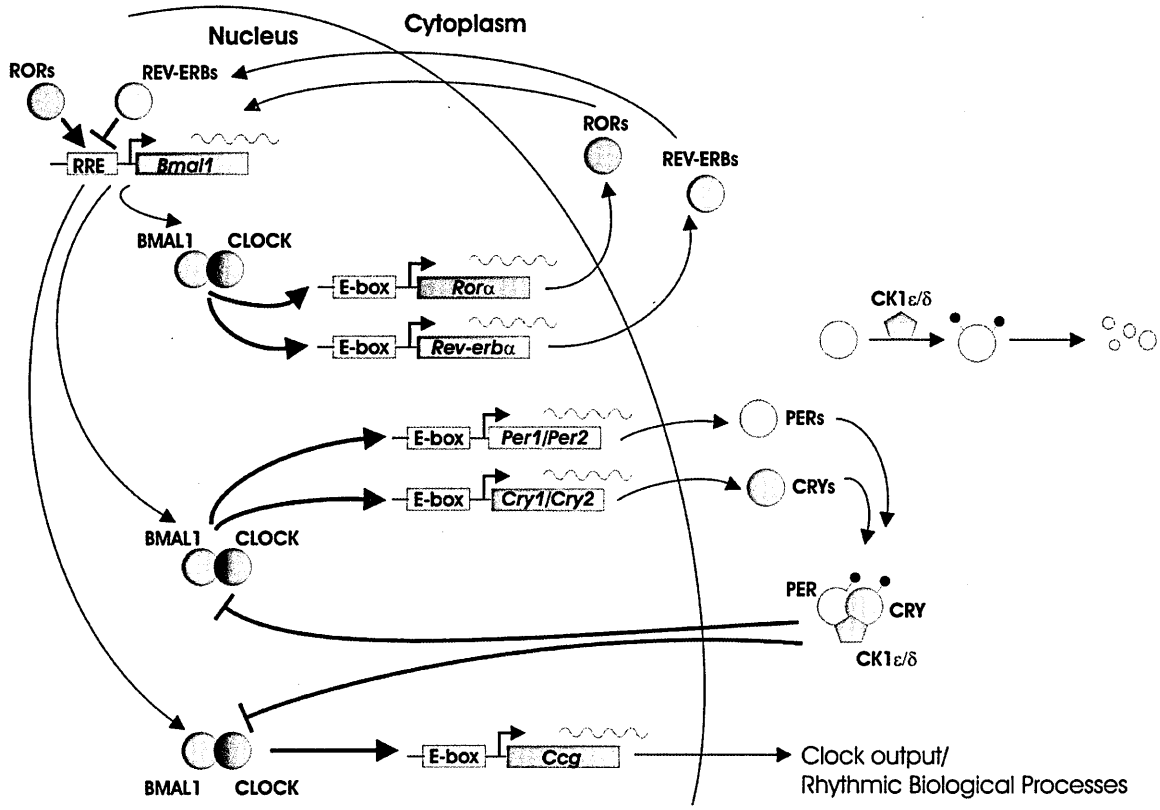

Figure 2.1: **Schematic representation of the mammalian circadian clock.**

orphan receptor response elements (ROREs) present in *Bmal1* promoter. It has been shown that members of ROR (α,β andγ) and REV-ERB (α and β) are able to regulate *Bmal1* through ROREs. RORs activate transcription of*Bmal1* , whereas REV-ERBs repress the transcription process. Hence, the circadian oscillation of *Bmal1* is both positively and negatively regulated by RORs and REV-ERBs [50].

The autoregulatory feedback loops described take 24 h to complete a cycle and constitute a circadian molecular clock. This generation of the 24 h

molecular clock is governed by post-translational modifications such as phosphorylation and ubiquitination, that significantly contribute to the precision of the mammalian clock by affecting the stability and nuclear translocation of the core clock proteins. A critical role in clock protein turnover is played by Casein kinase 1 *epsilon* and Casein kinase 1 *delta* (CK1ϵ and CK1δ) [36]. Recently, a small ubiquitin-related modifier protein modification of BMAL1 has also been proposed as another level of post-translational regulation [19].

It has been recently demonstrated that the circadian clock is active also in erytrocytes, which have no nucleus, thus showing that the core clock components can work also using just post-translational mechanisms[66].

Inspired by natural devices controlling the circadian clock, synthetic biologists have arranged standardized biological parts in circuits exhibiting oscillatory behavior. As the natural biological design, these circuits are based on a combination of positive and negative feedback loops with integrated time-delay dynamics, but unlike natural clocks, the synthetic clocks do not exhibit robust temperature-compensated stable oscillations.

In the next section, I will provide an overview of the most successful results in engineering synthetic biological clocks.

2.2 Synthetic oscillators: from bacteria to mammals.

2.2.1 The repressilator

The pioneering design of the first synthetic oscillator was published in 2002 by Elowitz and Leibler [38] and shown in Fig. 2.2. The network is based on an engineered transcriptional repressor system (termed repressilator), in which three repressors are assembled in a negative feedback loop that induces the rhythmic expression of a green fluorescent protein (GFP) in *Escherichia coli*. In the network, the LacI protein from *E.coli* inhibits the production of the second repressor gene *tetR* from the tetracyclin-resistance transposon Tn*10*, which in turn represses the production of a third repressor CI from lambda phage. This circuit produced fluctuating levels of each repressor protein that could be visualized by plugging in a GFP expression unit driven by the promoter target of the repressor TetR. Despite the large variability in oscillatory behavour among cells as result of biological noise, the average period of oscillations was about 150 minutes, that is threefold longer than the *E.coli* doubling time, demonstrating that cell division and repressilator cycle were uncoupled. Nevertheless oscillations were not robust and damped rapidly, probably because of the missing positive feedback loop, which has been shown to increase the robustness to noise of genetic circuits [10, 12]. The key features the authors highlighted as necessary for oscillations to ap-

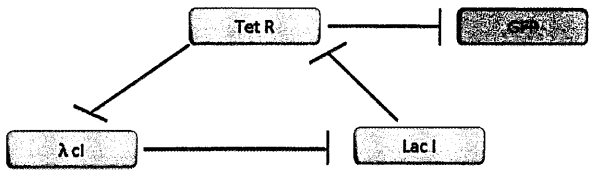

Figure 2.2: **The Repressilator.** A negative feedback loop composed by three repressors, inhibiting each other. The GFP, used as readout of the network, is controlled by the promoter responsive to the TetR repressor.

pear were the use of strong promoters with minimal leakiness coupled to efficient ribosome-binding sites, co-operative repression and low, as well as, comparable protein and mRNA decay rates.

2.2.2 The first synthetic clock combining positive and negative feedback loop

From the repressilator and theoretical speculations, it appeared clear that a robust oscillator required the functional combination of two circuits providing both a positive and a negative feedback loop. In Atkinson et al [7], a genetic oscillator is described, which combines the *E.coli* nitrogen-controlling

glnALG, and *lac* operons. Specifically, it consists of a dual-input promoter *Plac-glnAp2*, containing the operator sites for LacI and NRI P, which is repressed by LacI and induced by the (auto-)phosphorylated nitrogen-response regulator NRI P (Fig. 2.3). NRI P not only regulates its own transcription via *Plac-glnAp2* (positive feedback loop) but, by binding the *PglnKp* promoter (also containing an NRI P operator), it induces the expression of *LacI*, resulting in the LacI-mediated repression of *Plac-glnAp2* (negative feedback loop). Oscillations were monitored by β-galactosidase level produced by the LacI-controlled lactose operon. The circuit was integrated in *E.coli* strains deficient for the individual clock components and grown in turbidostat cultures. By synchronizing the cells with IPTG, damped oscillation were observed with a period (20h) much longer than the cell cycle (2h).

2.2.3 The mammalian genetic oscillator.

In 2009, Tigges et al [81] described the first mammalian oscillator. The circuit, shown in Fig. 2.4, consists of a sense-antisense expression "pendulum" with the tetracycline-dependent transactivator tTA auto-regulating itself, thereby forming a positive feedback loop. The tTA drives also the transcription of the streptogramin-dependent transactivator, which induces the tTA in antisense orientation, thus reducing the tTA levels (negative feedback). The fluctuation of tTA levels were monitored via a few-minutes half-life-GFP, whose expression is driven by the tTA. The different components of the genetic oscillator were carried by three different plasmids. The oscillator showed automatic,

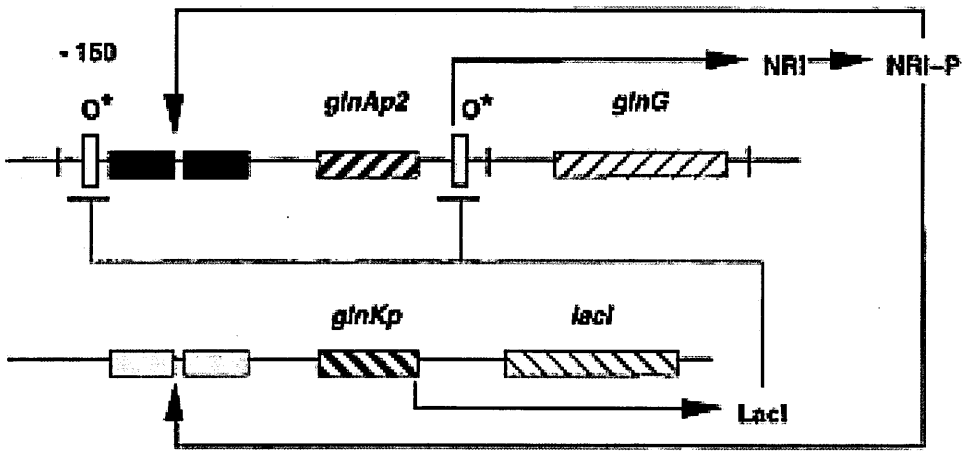

Figure 2.3: **Genetic clock by combination of positive and negative feedback loop in *E. coli*.** The *Plac-glnAp2* promoter drives the expression of the activator NRI that, upon auto-phosphorilation (NRI-P), regulates positively itself (positive feedback loop) and induces the negative feedback loop by activating the expression of the repressor Lac I after binding the the *PglnKp* promoter. Lac I represses the *Plac-glnAp2* promoter and also the expression of β-galactosidase, whose oscillations occur with a period of 20h.

self-sustained oscillations with a frequency of about 147 minutes, when cells were transfected with the same amount of each expression vector (1:1:1 100 ng each). The main limit of such an approach is that since the uptake of plasmids by mammalian cells is random, only few individual cells that received the right mix of each component exhibited oscillations. Moreover plasmids transfected in mammalian cells are not maintained as episomes, hence the number of oscillations are limited, and oscillations are not sustained.

Beside the biological development of designs for oscillatory networks, a fundamental role is played by mathematics. All the circuits described above consist not only of an experimental part, but also of a modeling part, where mathematical models of these circuits are used to investigate the best combination of biological parts in order to obtain the desired behaviour.

In the next section, I will provided an overview of the general mathematical formalisms used to describe biological systems.

2.3 Mathematical modeling of natural and synthetic gene networks.

In the fields of Systems and Synthetic Biolgy, theory and experiments need to be viewed in close interplay. *In silico* predictions of the behaviour of a biological system can be used to complement in vivo experimental observations and accelerate the hypothesis generation-validation cycle of research [56]. Modelling a cellular process can highlight which experiments are likely

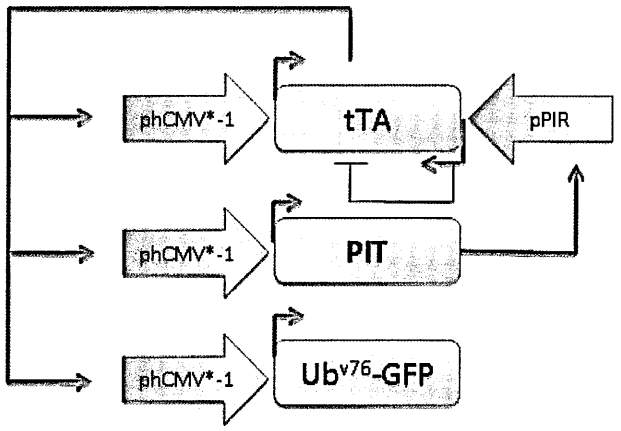

Figure 2.4: **Mammalian oscillator.** The tetracycline-dependent transactivator tTA is driven in sense orientation by the tetracycline responsive promoter *phCMV*-1*, and in the antisense orientation by the streptogramin-responsive (*pPIR*) promoter. tTA expression results in increased GFP and PIT levels. The PIT induces the expression of antisense tTA via binding to the *pPIR* promoter; thus the levels of tTA, PIR and GFP decrease. The whole combination of positive and time-delaying negative feedback loop mediate robust oscillations.

to be the most informative in testing model hypotheses, and allow testing for the effect of drugs [34] or mutant phenotypes [72] on cellular processes, thus paving the way for individualized medicine.

A mathematical model is the formalization of the biological knowledge about a certain system, where each component of the system is described by an equation, which represents its behaviour as a function of its regulators. *A priori* knowledge, which derives from experiments and/or literature, is essential and needs to be formalized for the chosen framework. Unfortunately, even for the best-studied systems, the mass of accumulated data still falls short of describing, even qualitatively, the variety of elementary processes that each molecular species engages in (post-translational modifications, degradation, complex formation, and so on); even less known are details of spatial information and the timing of events. Consequently, assumptions are necessary (for example, that all gene copies of a multi-copy plasmid are transcriptionally active, or that a certain molecule freely diffuses inside a cell or is always monomeric). On the other hand, sometimes it can be beneficial to exclude some known data to accommodate available computational power and to facilitate the analysis (even at the expense of accuracy).

2.3.1 Model derivation approaches

Model derivation from experimental data can be carried out following three major approaches: white-box, black-box and gray-box.

In white-box modelling, the model and parameter values are entirely derived from first principles; in black-box modeling the model is completely derived from input-output data. The gray-box approach [65], combines the two above approaches. Specifically, first principles are used to partially derive the model structure, while parameters or terms in the model are determined by measurement data. In the following, we will focus on the gray-box approach. Gray-box modeling entails three main steps to be executed iteratively: (i) derivation of the model equations; (ii) identification of the model parameters from experimental data and/or literature; (iii) validation (or invalidation [3]) of the model. Step (i) requires introducing simplifying hypothesis and choosing a proper formal framework. Step (ii) is required to estimate unknown model parameters from the available experimental data. A crucial issue that arises when estimating model parameters, is the structural identifiability. Identifiability addresses feasibility of estimating unknown parameters from data collected in well-defined stimulus-response experiments [24]. We can distinguish between structural non-identifiability and practical non-identifiability. The first is related to the model structure independently from experimental data; the second also takes into account the amount and the quality of measured data used for parameters calibration. Of note, a parameter that is structurally identifiable may still be practically non-identifiable, due to the unavoidable presence of noise in biological experimental data [69]. Unfortunately, while being well assessed in the case of linear dynamical systems, the identifiability analysis of highly non-linear systems remains an open problem

[15].

The parameter estimation problem can be formulated from the mathematical viewpoint as a constrained optimization problem where the goal is to minimize the objective function, defined as the error between model predictions and real data. In biological applications, the objective function usually displays a large number of local optima as measurements are strongly affected by noise. For this kind of problems, classical optimization methods, based on gradient descent from an arbitrary initial guess of the solution, can be unfeasible and show convergence difficulties. The above considerations suggest to look at stochastic optimization algorithms, like evolutionary strategies, which rely on random explorations of the whole space of solutions, are not sensitive to initial conditions and avoid trapping in local optimal points. In [63] the performance of both local and global-search optimization methods is compared in the identification of the 36 unknown parameters of a non-linear biochemical network. The authors show that only evolutionary strategies are able to successfully solve the parameters estimation problem, while gradient based methods tend to converge to local minima. Among the stochastic techniques, Genetic Algorithms (GA) [62] provide a very flexible approach to non-linear optimization. Their application showed good results in the parametrisation of synthetic networks [87, 81].

Finally, step (iii) is required to check the validity and usefulness of the model, that is to evaluate its ability in predicting the behaviour of the actual physical system. Theoretically, the modeller should be confident that the

formalism is able to describe *all* input-output behaviours of the system [76]. This condition can be never guaranteed, since it would require an infinite number of experiments. However, it is possible to test a necessary condition: the model is able to describe *all observed* input-output behaviours of the system [76]. To this aim, one possible approach is to use a cross-validation like procedure [4] by splitting the experimental data in two sets: one of them is used for the parameter identification, while the other one is used to validate the predictive power of the model. If the predictive performance of the model is not satisfactory, it is invalidated [3]. Thus, it is necessary to refine the model (for example, by increasing the level of detail) and/or to perform new experiments, going back to step (i) of the modelling procedure.

2.3.2 Modeling approaches

In general a major distinction can be done between qualitative and quantitative models. In qualitative modelling, for simulations to be applied and useful in drawing non-obvious conclusions, we need to retrieve from biological data at least the information required for the formulation of logical statements describing, for instance, causal relationships between events involving model components. On the other side, quantitative modeling have a natural appeal in that they offer greater detail in mimicking reality. Moreover, rich qualitative insights on the system are possible using theoretical tools such as bifurcation and stability analysis, which, for example, indicate the precise boundaries of parameter ranges to which steady states or sustained

oscillations correspond, or reveal the stability of the solutions before actually solving the dynamical equations representing the system. Quantitative modeling is further divided in deterministic or stochastic modeling.

Deterministic formalisms are commonly used to describe the average behaviour of a population of cells [31]. They have been shown to be viable for the analysis of synthetic networks in a number of works (e. g. [38, 40, 51, 81, 78]). The reaction mechanism is described by applying the law of mass action: the rate of any given elementary reaction is proportional to the product of the concentrations of the species reacting in the elementary process (reactants) [1].

When Differential Equations (DEs) are used, the cellular concentration of proteins, mRNAs and other molecules are represented by continuous time variables with the constraint that a concentration can not be negative. The DEs modelling approach is based on the following biological assumptions: the quantified concentrations are continuous functions of time. These assumptions hold for processes evolving on long time scales in which the number of molecules of the species in the reaction volume is sufficiently large. As described in Chapter 5 we used the Differential Equation based approach to model the synthetic oscillator, since this system satisfies the above assumptions.

Chapter 3

Construction and modelling of an inducible positive feedback loop stably integrated in a mammalian cell-line.

This chapter concerns the characterization of the positive feedback loop motif, which constitutes one of the parts of the oscillatory network described in Chapter 5. In particular it focuses on the investigation of the dynamical properties determined by the topology of this network motif.

The positive feedback loop is a regulatory motif often found in transcriptional and signalling pathways. This motif exhibits a dynamic behaviour which is very different from that obtained when autoregulation is removed. This difference is intrinsic to the specific wiring diagram chosen by the cell to control its behaviour (feedback versus non-feedback configurations), and can be instrumental in understanding the complex network of regulation occurring in a cell. Part of the results described in this chapter have been published in [74].

3.0.3 Rational construction and integration of a PFL and of corresponding control network (NOPFL) in CHO cells.

Our approach is based on the use of well known and characterised regulators of gene expression, in order to achieve a complete control of the network behaviour. The positive feedback loop (PFL) is shown in Fig. 3.1A. In designing the PFL, we took advantage of the inducible Tet regulatory system; the expression of Tetracycline-controlled transactivator tTA is self-controlled by a *CMV-TET* promoter, responsive to the tTA itself unless Tetracycline, or its analogous Doxycycline, is added to the medium in which cells are grown. To follow the dynamics of the PFL, a destabilised yellow-green variant of the enhanced green fluorescent protein (d2EYFP) (Clontech), with a reported half-life of approximately two hours, was placed under the control of the same promoter. To this end, we constructed a unique cassette with an Intra Ribosomal Entry Sequence (IRES) in between of the transactivator tTA and the d2EYFP, which enables a single mRNA to encode for two different proteins (Fig. 3.1A).

We embedded the positive feedback loop (PFL) in a lentiviral vector [2, 45] in order to infect and select clonal population of Chinese Hamster Ovary cells (CHO) cells stably expressing the circuit. We decided not to use plasmid transfections, in order to prevent precise quantitative measurements due to the unpredictable amount of plasmids that enters each cell, and to

the transient nature of transfection.

Infected cells were first sorted by Fluorescence Activated Cell Sorter (FACS) according to their fluorescent levels (d2YFP) and then a clonal population of CHO cells carrying the PFL was generated by single cell expansion (PFL cells).

In order to capture the dynamic properties intrinsic to the PFL, we generated a control network lacking the positive feedback loop (NOPFL), but using the same biological "parts" as in the PFL network. As shown in Fig. 3.1 B, we constructed a cassette containing the same *CMV-TET* promoter upstream of the d2EYFP. The tTA protein, this time, was placed under the control of a constitutive promoter, thus breaking the feedback loop. Using the same strategy described above, we generated a clonal population of CHO cells carrying the NOPFL network (NOPFL cells). We experimentally verified that both PFL and NOPFL cells have the same number of tTA/d2EYFP DNA integrations (Fig.3.1 inset). Further details about lentiviral vectors we designed are provided in Appendix A.

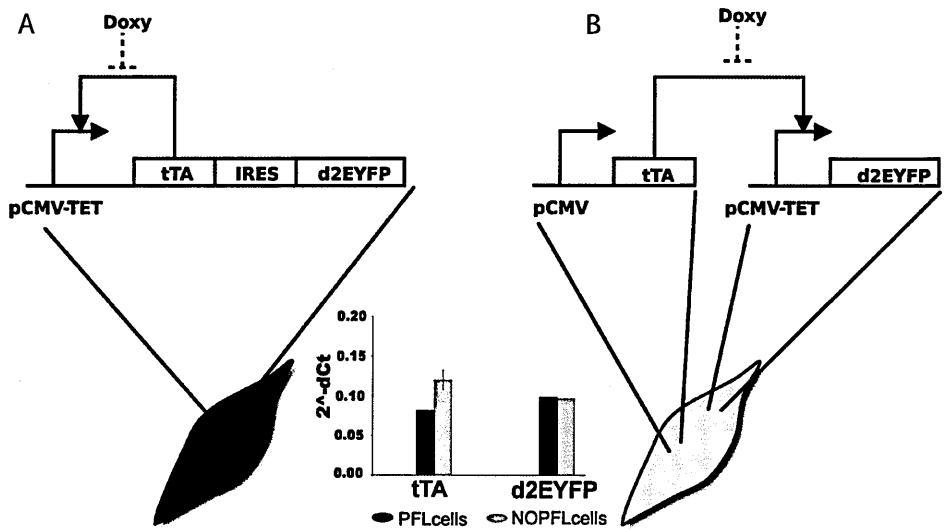

Figure 3.1: **Design of the expression system.** (A) PFL: the promoter *CMV-TET* consists of seven direct repeats of a 42-bp sequence containing the tet operator sequences (tetO), located just upstream of the minimal *CMV* promoter (PminCMV). The Tetracycline-controlled transactivator tTA derives from the addition of the VP16 activation domain to the transcriptional repressor TetR to optimize the expression in mammalian cells. The d2EYFP is the destabilised yellow-green variant of enhanced green fluorescent protein. (B) NOPFL: the *CMV* promoter drives the expression of the tTA, which in turns drives the transcription of the d2EYFP from the *CMV-TET* promoter. (Inset) RealTime PCR performed on DNA extracted from PFL and NOPLF cells shows that the DNA levels of tTA and d2EYFP are comparable among the two clonal cell populations.

3.0.4 Determination of the reporter protein degradation.

Since the degradation rate of the reporter protein is a key parameter in determining both the observed fluorescence dynamics and the precision of the mathematical model, we experimentally evaluated the half life of the d2EYFP. Stably integrated NOPFL cells were treated with Cyclohexamide to a final concentration of 10μg/mL, 50μg/mL, 100μg/mL or 500μg/mL, to inhibit protein synthesis[75]. The fluorescence intensity of NOFPL cells was followed for 12 hrs and images were acquired at 15 min intervals via an optical inverted microscope (Nikon Eclipse T). The resulting d2EYFP dynamics are shown on Fig. 3.2 and appear very similar, independently of the Cyclohexamide concentrations. The experimental data were fitted to an exponential curve $ke^{-d_3 t}$, and the degradation coefficient d_3 was used to obtain the half-life ($\tau_{\frac{1}{2}}$) of the d2EYFP protein: $\tau_{\frac{1}{2}}$=log(2)/ d_3 (Fig. 3.2 and Table 1). We estimated $\tau_{\frac{1}{2}}$ to be in the range 3.6h-4.4h. (Experimental procedure: determination of d2EYFP half-life). The estimated value is about two-fold the reported d2EYFP half-life of 2h [85]; we believe that this discrepancy is likely due to the fact that cells were grown at a temperature 32 °C, rather than the usual 37 °C.

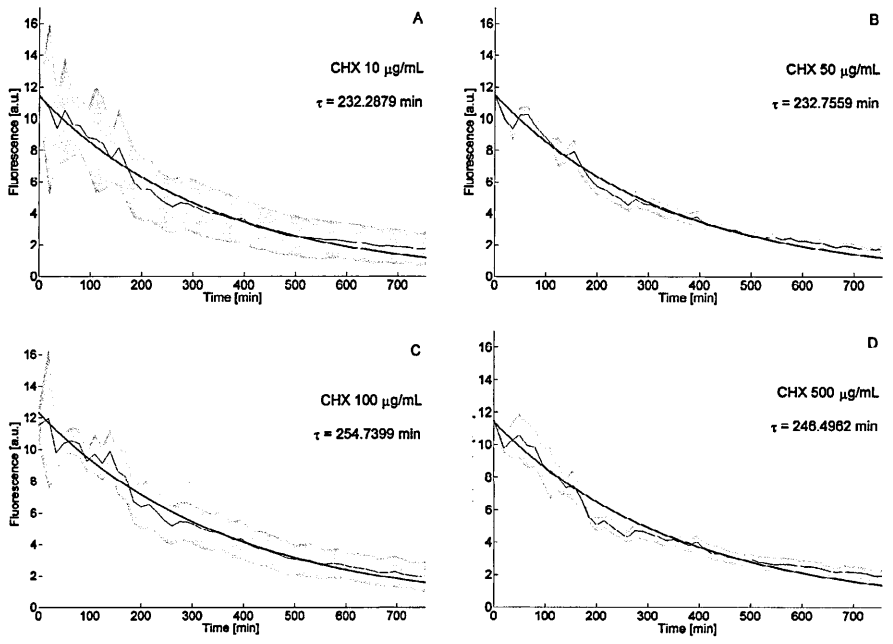

Figure 3.2: **Degradation kinetics of d2EYFP.** *CMV-TET*-d2EYFP stably integrated CHO AA8 TET-OFF cells were treated at t=0 with different concentrations of Cyclohexamide (CHX): panel A, 10μg/mL; panel B, 50μg/mL; panel C, 100μg/mL; panel D, 500μg/mL. Fluorescence intensity was followed up to 750 minutes. Sampling time is equal to 15 min. The thin line represents the mean over biological triplicates; the shaded area represents the standard error. Experimental data were used to fit the exponential decay of d2EYFP protein levels, and thus to derive its half-life (τ).

3.0.5 Mathematical modeling of the dynamic behavior of PFL and NOPFL networks.

We derived a model of the PFL and NOPFL networks using Ordinary Differential Equations (ODEs), that, as described in Chapter 2, are suitable since we are measuring the average behaviour of a clonal population of cells ([81, 40, 51, 78]) .

For each species, i.e. each mRNA and correspondent protein concentration, we wrote one equation, which expresses the change in concentration of the species in a given time interval, as the result of a production term and a degradation term. We assumed:

- Hill functions to model the rate of gene transcription, including basal activity to describe the leakiness of the *CMV-TET* promoter;
- linear degradation for all genes and proteins;
- linear dynamics for the translation;
- Hill functions to model the effect of the inducer (Doxycycline);
- distinct dynamics for the unfolded (inactive) and folded (active) forms of the reporter protein (d2EYFP).

The last assumption was introduced in order to take into account d2EYFP maturation time needed for correct protein folding[81]. Thus, we introduced two differential equations as in [81]: one for the translation of mRNA to the unfolded d2EYFP protein, and one for the folded protein d2EYFP.

Letting x_1 be the *tTA* IRES *d2EYFP* mRNA concentration, x_2 the tTA protein concentration, x_3 the unfolded d2EYFP protein concentration and x_4 the folded d2EYFP protein concentration, the PFL network can be described as follows:

$$\frac{dx_1}{dt} = v_1 \left(\alpha_1 + (1-\alpha_1) \frac{\left(\frac{\theta^{h_2}}{\theta^{h_2}+D^{h_2}} x_2\right)^{h_1}}{K_1^{h_1} + \left(\frac{\theta^{h_2}}{\theta^{h_2}+D^{h_2}} x_2\right)^{h_1}} \right) - d_1 x_1, \quad (3.1)$$

$$\frac{dx_2}{dt} = v_2 x_1 - d_2 x_2, \quad (3.2)$$

$$\frac{dx_3}{dt} = v_2 x_1 - (d_3 + K_f) x_3, \quad (3.3)$$

$$\frac{dx_4}{dt} = K_f x_3 - d_3 x_4. \quad (3.4)$$

Note that, due to the presence of the IRES sequence, the concentrations of tTA protein and d2EYPF protein depend on the same variable (x_1), that is the concentration of the single mRNA transcript encoding for both proteins.

For the NOPFL network, we let x_1 represent only the *d2EYFP* mRNA concentration, and we assumed a constant level ($\bar{x}_2$) of the tTA protein, due to the constitutive promoter driving *tTA* expression in the NOPFL cells. The equations thus become:

$$\frac{dx_1}{dt} = v_1 \left(\alpha_1 + (1-\alpha_1) \frac{\left(\frac{\theta^{h_2}}{\theta^{h_2}+D^{h_2}} \bar{x}_2 \right)^{h_1}}{K_1^{h_1} + \left(\frac{\theta^{h_2}}{\theta^{h_2}+D^{h_2}} \bar{x}_2 \right)^{h_1}} \right) - d_1 x_1, \quad (3.5)$$

$$\frac{dx_3}{dt} = v_2 x_1 - (d_3 + K_f) x_3, \quad (3.6)$$

$$\frac{dx_4}{dt} = K_f x_3 - d_3 x_4. \quad (3.7)$$

3.0.6 Experimental observation of dynamics properties of PFL and NOPFL networks

In order to observe the dynamics of the PFL and NOPFL networks, we performed time-series experiments in which stably-integrated CHO-PFL cells and CHO-NOPFL cells were imaged using time-lapse fluorescence microscopy. The experimental design consisted in treating both PFL and NOPFL cells with different amounts of Doxycycline in order to "switch off" the circuit, by preventing the tTA protein from binding the *CMV-TET* promoter. We tested the following Doxycyline concentrations: 1ng/mL, 10ng/mL, 100ng/mL and 1μg/mL and followed the dynamic behaviour of both the PFL and NOPFL cells for 43h, collecting images every 15 min, and quantifying the average fluorescence intensity of the cell population. In this way, we averaged out cell-to-cell variability in the response, since at the beginning of each experiment the tracked microscopy field contained at least 15 cells.

Experiments were carried out at a temperature of 32 °C in order to limit

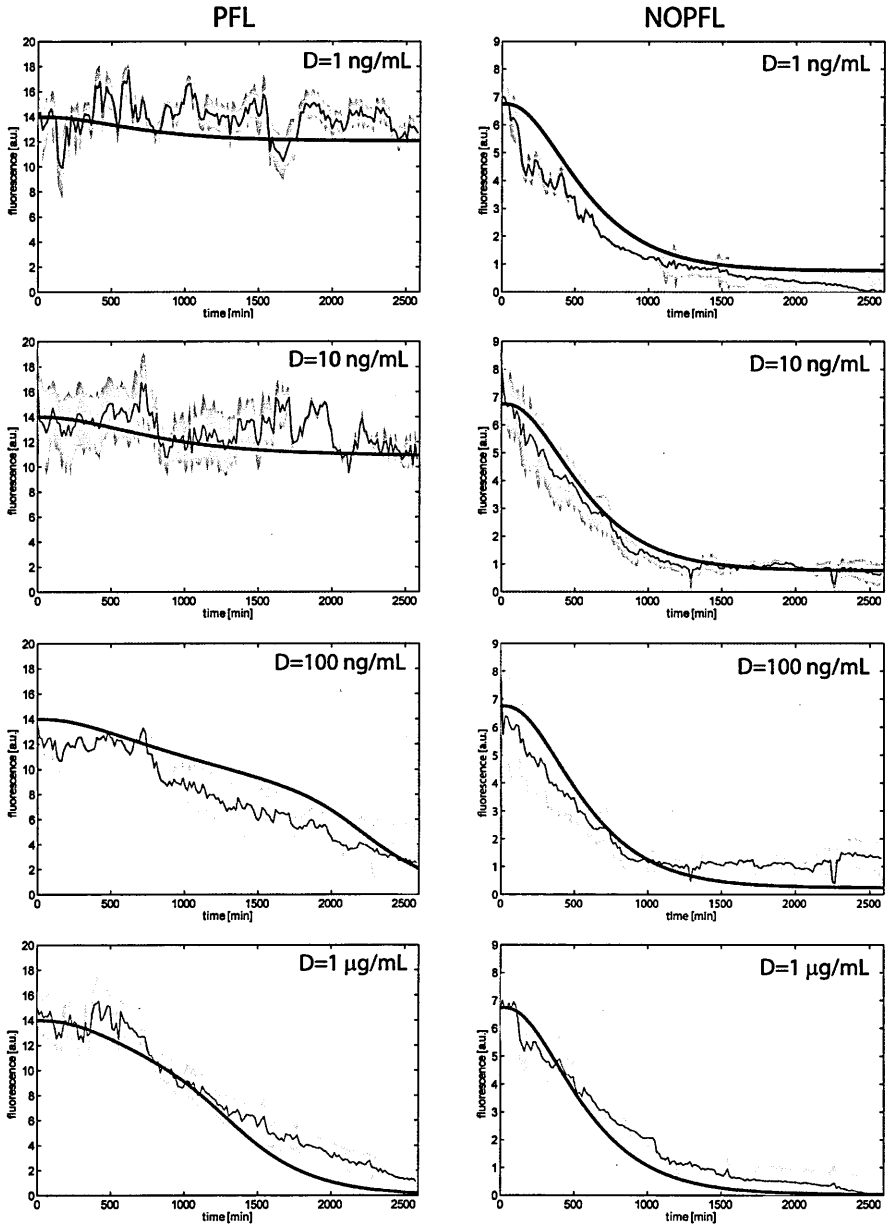

Figure 3.3: **Experimental and simulated switch off time-course across the PFL and NPFL cell population.** Experimental data (thin lines) and model simulations (thick lines) were reported for the PFL (left) and NOPFL (right) cells. Shaded areas represent standard deviations from replicate experiments.

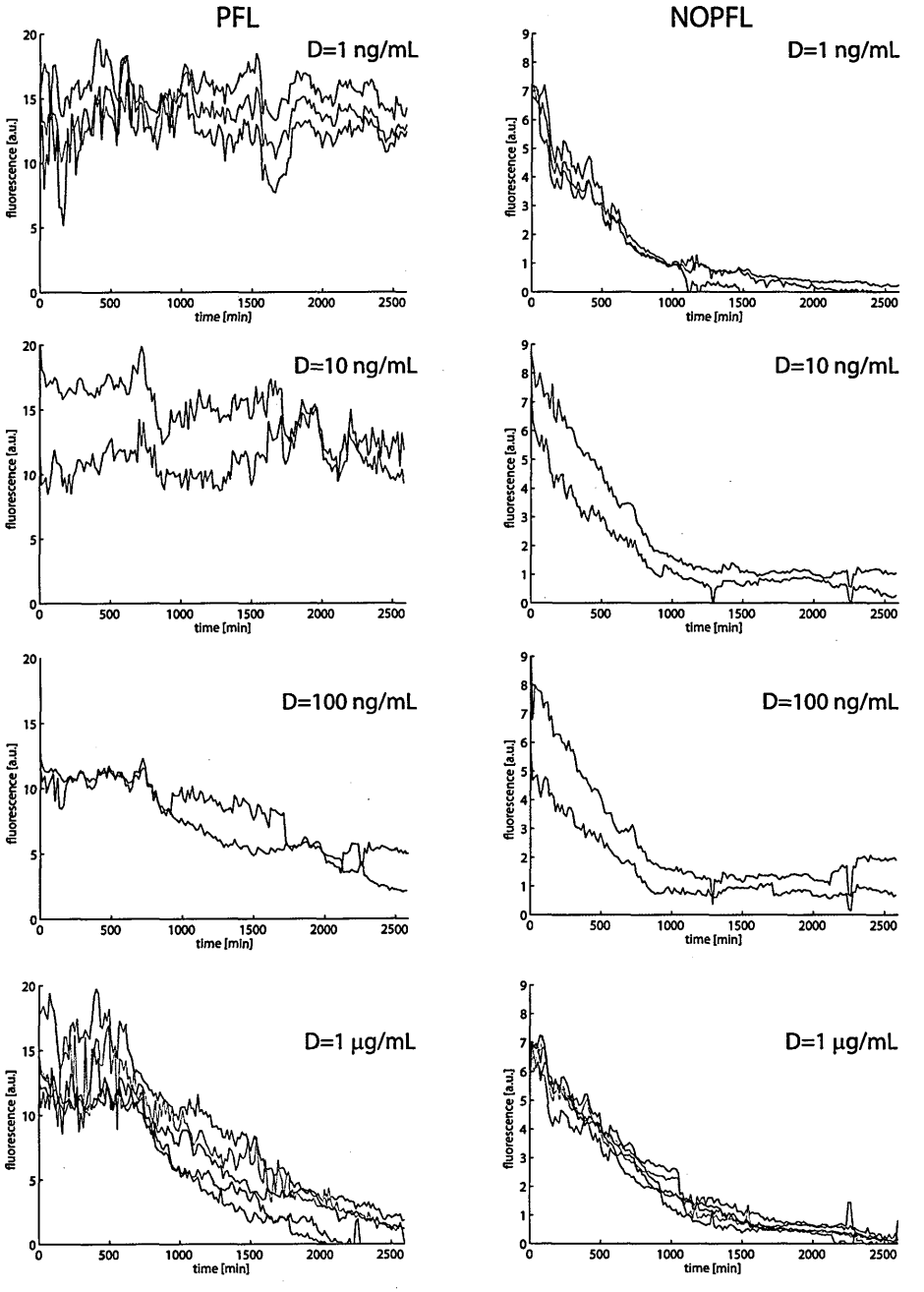

Figure 3.4: **Replicates of the experimental time-courses across the PFL and NPFL cell population.** Replicates of the experimental time-courses for the PFL (left) and NOPFL (right) cells. Each line in each panel represent the average fluorescence intensity across the cell population in one switch-off experiment.

cell motility and reduce the risk associated to data loss occurring when cells exit the tracked field [46]. The average fluorescence intensity of the reporter gene across the cell population for both the PFL and NOPFL networks is shown in Fig. 3.3 for the different concentrations of Doxycycline indicated. In Fig.3.4 replicate time-course experiments are shown for each of the Doxycycline concetrations used.

3.0.7 Derivation of the model parameters

The parameter estimation problem is defined as an optimisation problem, where the goal is to minimise a performance measure defined as the error between model predictions and observations, which in our case are the time-series during the "switch off" experiments in Fig. 3.3and 3.4. Several alternative strategies can be pursued in order to obtain best estimates of the model parameters, ranging from Newton's method to Genetic Algorithms (GAs). Here we employed the Trust Region method (TRM) implemented in PottersWheel [58]; thanks to the multi−experimental capabilities of this tool, we were able to identify the 12 parameters by simultaneously fitting Eqs. (1) to (7) to all of the experimental time-courses at once. These time-courses include all of the different Doxycycline concentration for both the PFL and NOPFL cells for a total of 24 time-courses, when taking experimental replicates into account. Moreover, in order to obtain confidence intervals for each parameter, we run multiple identification procedures on the time-series data, using parameter perturbation routines to allow extensive sampling in the pa-

rameters' space. In order not to introduce any specific bias in our search, we only set admissible ranges for the parameter values to be identified, which reflected physical and biological feasibility, either obtained from literature [51] or directly estimated (degradation rate $d2EYFP = d_3$).

The result of the parameter fitting procedure are reported in Table 1 along with the estimated standard deviation, which are in most of the cases one order of magnitude less than the corresponding parameter value, or at most of the same order of magnitude.

We observed that the parameter h_2 in Table 1, which affects the strength of Doxycycline repression on the tTA protein activity, is much smaller than 1. Usually Hill coefficients are greater than 1, therefore we wondered what could be a possible biological explanation for this behaviour. We observed that for the range of Doxycycline concetration used in the experiment (1ng/mL to 1μg/mL), using the parameters' values in Table 1, the function: $\frac{\theta^{h_2}}{\theta^{h_2}+D^{h_2}}$ can be approximated by the function: $\alpha + \beta\frac{\theta}{\theta+D}$ ($\alpha \approx 0.4$ and $\beta \approx 0.6$). This means that a Michealis-Menten function can also describe the effect of Doxycyline on tTA acitivity, but a certain level of leakiness (α) must be taken into account; that is even for large concentrations of Doxycycline, the activity of the tTA protein cannot be completely shut down.

Table 3.1: Parameters identified after the fitting procedures: parameters values as well as standard deviation are reported for each parameter.

Parameter	Description	Fitted value	STD
K_1 [nM]	Activation coefficient	4.81	1.06
α_1	Basal activity CMV-TET promoter	1.13E-05	3.62E-05
v_1 [nM min^{-1}]	Maximal transcription rate CMV-TET promoter	7.54E-02	1.97E-02
v_2 [min^{-1}]	General translation rate	2.71E-02	1.22E-02
d_1 [min^{-1}]	Degradation rate tTA mRNA	1.01E-02	1.22E-03
d_2 [min^{-1}]	Degradation rate tTA protein	1.00E-02	3.42E-03
d_3 [min^{-1}]	Degradation rate d2EYFP protein	3.24E-03	2.66E-04
h_1	Hill coefficient of the CMV-TET promoter	3.16	1.40E-01
θ [nM]	Affinity Doxycycline - CMV-TET promoter interaction	1.00	8.85E-03
K_f [min^{-1}]	Folding rate d2EYFP	1.24E-03	1.41E-02
$\bar{x}_2$ [nM]	Steady state tTA in NOPFL	13.69	7.63E-01
h_2	Hill coefficient for Doxycyline	6.03E-02	7.19E-03

The "switch off" time-series experiments were simulated with both the PFL and NOPFL models using the fitted parameters as shown in Fig. 3.3. The inferred models are able to recapitulate the observed dynamics in response to different inducer concentrations and experimental settings. Observe that the parameters for both the PFL and NOPFL models are identical, except for $\bar{x}_2$ in the NOPFL equations, which is not present in the PFL model. Hence, the observed differences in the dynamical behaviour of the PFL and NOPFL networks are due to the intrinsic differences in their topology.

Dynamic properties of the PFL and NOPFL networks

In order to further investigate the relationship between topology and dynamical properties, we first observed that the NOPFL model described by Eq.3.5-3.7 is a system of linear time-invariant ODEs, for which the theory of liner dynamical systems applies. From the theory, we know that changes in Doxycycline concentration in Eq.3.5 will not affect the dynamic behaviour of the model, which is governed by the smallest among three degradation terms $d_1, d_3, (d_3 + K_f)$ (Model simulations and parameter identification). The concentration of Doxycycline affects only the steady-state values, i.e. how much the network will switch off, but not its dynamics, i.e. how fast it will switch off. Therefore, independently of the values of the parameters, the model of the NOPFL network predicts that for any concentration of Doxycycline, the network will switch off with the same dynamics, albeit possibly reaching different steady-state levels. Fig. 3.6 reports the "switch off" time, τ_{off}, for both

the PFL (dashed) and the NOPFL (solid) networks as a function of Doxycycline concentration, computed via numerical simulations of the two models with the parameters estimated in Table 1. τ_{off} is defined as the time taken by the fluorescence intensity to reach 50% of its final steady-state value (OFF), following treatment with Doxycycline at a given concentration (Material and Methods). In agreement with the experimental observations, the τ_{off} for the NOPFL network is constant and does not change with Doxycyline. In Fig. 3.6, the switch off time for the NOPFL network for the different concentrations of Doxycycline was estimated from the experimental time-series data ($\times$ in Fig. 3.6) (Model simulations and parameter identification).

On the other hand, the PFL network has a very different behaviour, as can be seen in Fig. 3.6. Specifically, for a range of Doxycycline concentrations, the PFL τ_{off} is considerably longer (+ in Fig. 3.6) that the NOPFL counterpart, which again is in agreement with the experimentally observed behaviour (Material and Methods). In order to investigate the origin of the observed dynamical behavior of the PFL circuit, we analysed the phase portrait associated to the *d2EYFP–tTA mRNA* and the tTA protein, which allows to directly observe trajectories of two state variables at once. Moreover, by imposing the steady-state conditions (i.e. $\dot{x}_i = 0$), we can derive nullclines, as well as, the their intersection points, which correspond to the equilibrium points of the network. In Fig. 3.6 the nullclines for different Doxycycline concentrations are shown. When no Doxycycline is present, two stable points (OFF and ON) and one unstable equilibrium points coexist

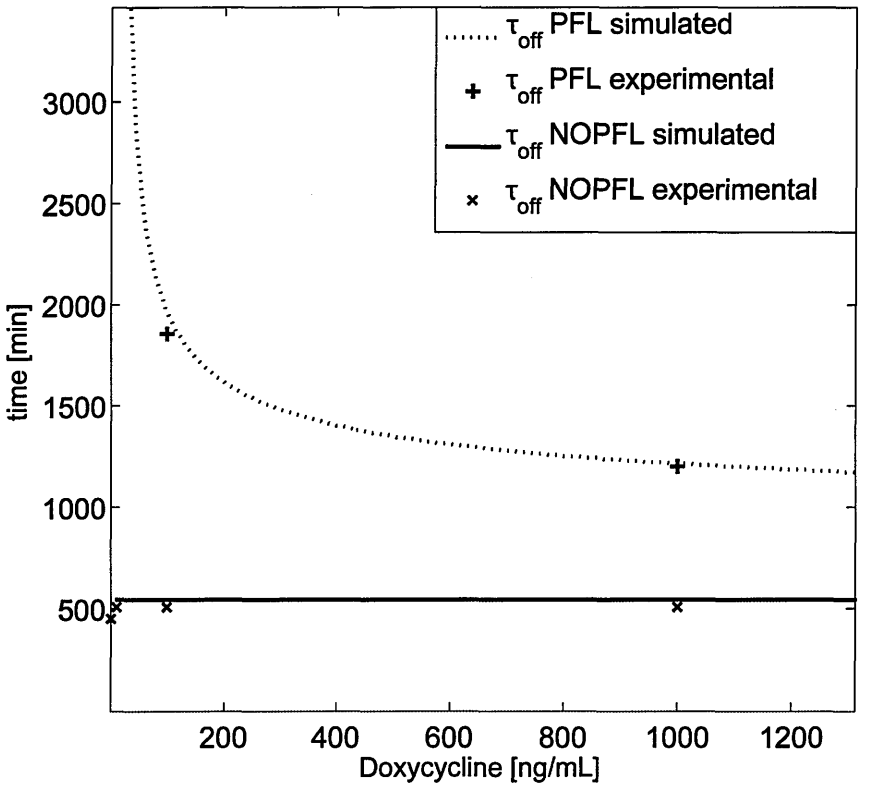

Figure 3.5: **Switch off time τ_{off} for varying Doxycycline concentrations from experimental data and model predictions.** The model predictions for the switch off times τ_{off} are shown for PFL (dashed thick line) and NOPFL (solid line). Experimental quantification of the τ_{off} for PFL and NOPFL models have been reported for comparison with $+$ and $\times$ respectively. Observe that the experimental τ_{off} for the PFL at 1ng/mL and 10ng/mL could not be estimated since the PFL is not switching off in the experimental observation time (43h).

in the same phase portrait, thus providing evidence for the bistability of the PFL network, a shared property among positive feedback loops[12]. However, as Doxycycline concentration increases, the bistability is lost (Fig. 3.6), and the only possible equilibrium point is the "OFF" state.

3.0.8 Conclusions

The most striking feature is the slow down in the switch off time of the PFL cells as compared to the NOPFL cells; moreover the switch off time of the PFL is affected by Doxycycline concentrations, whereas NOPFL cells always switch off with approximately the same dynamics.

The reason for a cell to "choose" a PFL control strategy for transcriptional regulation, rather than the NOPFL strategy, could be due to the intrinsic robustness of this approach to transient activation of the network. For example, in a signalling pathway, a ligand (equivalent to Doxycycline in our PFL) could cause a transcription factor to stop transcribing itself, as well as, a set of target genes, to initiate a specific response. However, in order for the pathway not to respond to a transient concentration of the ligand, the PFL strategy has to be chosen, otherwise the response would start immediately (NOPFL case). Moreover, the response time of the PFL network can be modulated by the ligand concentration, if this is really high, the system will switch off as quickly as possible, alternatively the ligand can be present at low, or medium, concentration, but it should persist for a long time, in order for the pathway to respond. This kind of behaviour has been

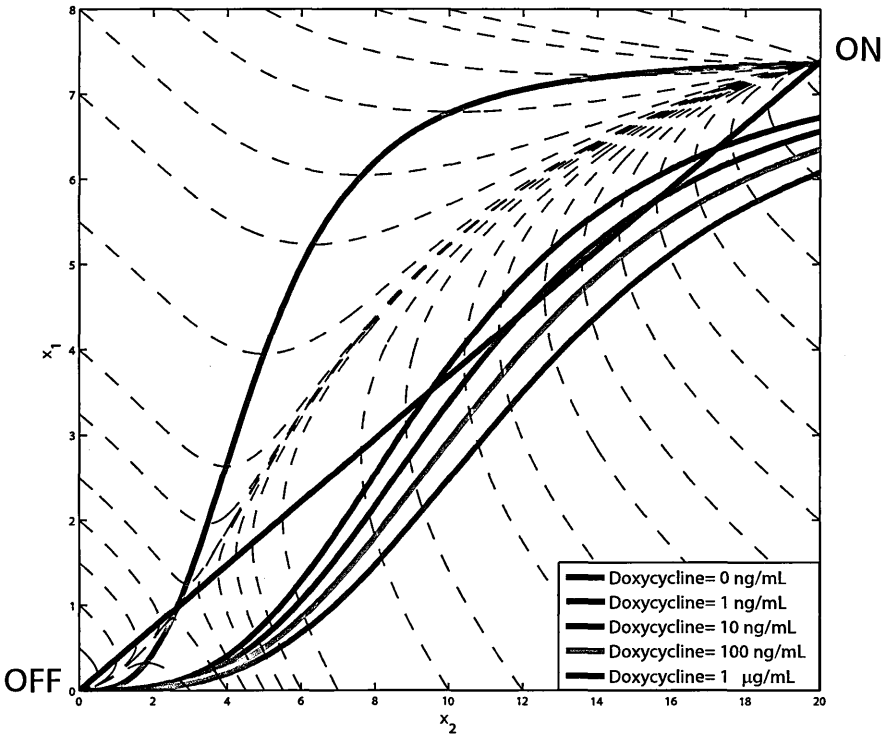

Figure 3.6: **Phase portrait of the PFL model.**The tTA-d2EYFP mRNA concentration (y axis) has been plotted against tTA protein concentration (x axis). Varying Doxycycline concentrations (1 ng/mL through 1 μg/mL) were used to investigate the dependence of the two stable equilibria ("ON" and "OFF" in the graph) on the amplitude of the input. The shape and dimensions of the two basins of attraction (the set of initial conditions ending up in one of the two stable steady states) can be studied with the same technique: in this figure the grey shaded area represents the basin of attraction of the "OFF" equilibrium for Doxycycline= 0 nM.

recently described as "persistence detection" in cellular signal processing to indicate the ability of the genetic circuit to distinguish between transient and persistent signals [89].

Moreover we verified that the PFL can exhibit bistability for zero or low concentrations of Doxycyline (Fig. 3.6). A bistable genetic network will cause a population of cells to divide in two sub-populations, each in one of the two possible states (OFF or ON). In yeast, this has been experimentally verified using a simple PFL based on the rtTA system[12]. In our system, the PFL model is bistable but the basin of attraction of the OFF equilibrium point (Fig. 3.6) is much smaller as compared to that of the ON state, when no Doxycycline is present. Therefore, just few cells will be in the OFF state and these will not be enough to be significantly detected experimentally. We predict however that for intermediate concentration of Doxycycline (100ng/mL in Fig. 3.6) the basin of attraction will be comparable and bistability should be detected experimentally.

Materials and Methods

Experimental procedure: construction of the circuit

To implement the gene circuit in a lentiviral vector, we used the ViraPower Promoterless Lentiviral Gateway Expression System (Invitrogen) which takes advantage of the site-specific recombination properties of bacteriophage lambda, making the transfer of single DNA sequences faster than the usual cloning

strategies.

The pMAtTA-IRES-EGFP vector containing the transactivator tTA, the IRES element and the enhanced green fluorescent protein (EGFP) was synthesised by GENEART together with the recombination sites. The d2EYFP was amplified from pd2EYFP-1 (Clontech) by PCR with a forward primer containing a NheI recognition sequence (5'-CATGGCTAGCATGGTGAGCAAGGGCGAGGAG-3') and a reverse primer containing an EcoRV recognition sequence (5'- ATTCGATATCAGTCGCGGCCGCATCTACA-3'). The PCR product and pMAtTA-IRES-EGFP were then digested with NheI-EcoRV restriction enzymes and the d2EYFP ligated in place of EGFP, generating a new vector termed pMAtTA-IRES-d2EYFP. The pMAtTA-IRES-d2EYFP was then linearised with the AseI restriction enzyme and recombined with the pDONR221 (Invitrogen) following the manufacturer instruction. In this way we generated pENTRtTA-IRES-d2EYFP vector with specific recombination sites.

The *CMV-TET* promoter was amplified from pTRE2 (Clontech) by PCR. The PCR was performed with the Taq polymerase provided by Invitrogen that adds a single deoxyadenosine (A) to the 3' ends of PCR products. This allows PCR inserts to ligate efficiently with the pENTR5'-TOPO vector which is supplied linearised with single 3'-deoxythymidine (T) overhangs, obtaining the pENTR5'-TOPO-*CMV-TET* with specific recombination sites.

Finally we performed a recombination reaction between the pENTRtTA-IRES-d2EYFP, pENTR5'-TOPO-*CMV-TET* and the pLenti/R4R2/V5-DEST

according to manufacturer instructions.

To generate the lentiviral vector containing the gene expression cassette lacking the positive feedback loop (NOPFL), the d2EYFP was amplified from pd2EYFP-1 with the High-Fidelity Taq Phusion (Fynnzimes) which gives blunted-end PCR product. The forward primer (5'-CACCGCCACCATGGTGAGCAAGGGCGAGGAG-3') was designed to allow the direct cloning in pENTR directional TOPO vector (Invitrogen), generating the pENTR d2EYFP vector. Then we performed a recombination reaction between the pENTR d2EYFP, pENTR5'-TOPO-*CMV-TET* and the pLenti/R4R2/V5-DEST according to manufacturer instructions.

As suggested by the manufacturer, the lentivirus was then produced in 293FTcells.

Experimental procedure: cell culture, lentiviral transduction, switch-off experiment

293FT cells were maintained at 37 °C in a 5% CO2-humidified incubator, and cultured in DMEM (GIBCO BRL) supplemented with 10% heat-inactivated fetal bovine serum (FBS) (Invitrogen), 1% L-glutamine, 1% MEM Non-Essential Amino Acids, 1% MEM Sodium pyruvate and 1% antibiotic/antimycotic solution (GIBCO BRL). CHO cells were maintained at 37 °C in a 5% CO2-humidified incubator, and cultured in α-MEM (Sigma) supplemented with 10% heat-inactivated fetal bovine serum (FBS) (Invitrogen), 1%

L-glutammine and 1% antibiotic/antimycotic solution (GIBCO BRL). CHO AA8 TET-OFF cell line (Clontech) was maintained α-MEM (Sigma)supplemented with 10% TET system approved FBS (Invitrogen), 4mM L-glutamine, 100μg/mL G418 (Sigma), and 1% antibiotic/antimycotic solution (GIBCO BRL).

To transduce cells with the virus produced, 500,000 CHO and CHO AA8 TET-OFF cells were plated and incubated overnight. On the day of transduction the medium was removed and 1mL of the virus was added to the cells together with polybrene (Invitrogen) to a final concentration of 6ug/mL. After an overnight incubation the medium containing the virus was removed and replaced with complete culture medium containing Blasticidin (Sigma) to a final concentration of 5μg/mL to select for stably transduced cells.

Cells were sorted for fluorescence intensity using a BD FACSAria Cell Sorting System (Becton Dickinson). d2EYFP was excited at 488 nm, and emission was detected using a 525 nm bandpass filter. Serial dilutions of stably transduced cells (up to 0.05 cells/mL) were plated in 96-well microtitre plates, and dilutions containing only one cell per well were selected. Monoclonal colonies were cultured and amplified as described, to obtain monoclonal populations.

For the switch off experiment, 500 stably-integrated-CHO and CHO AA8 TET-OFF cells were plated in chamber slide (lab-Tek) and treated with Doxycycline (Clontech) to a final concentration ranging from 1ng/mL to 1μg/mL). The switch off experiments were repeated twice for the 10ng/mL and 100ng/mL Doxycycline concentrations, while 3 and 5 replicates were

obtained for 1ng/mL and 1μg/mL. Experiments were performed in parallel for both the PFL and NOPFL cells.

Experimental procedure: DNA extraction, RealTime PCR

1,000,000 PFL and NOPFL cells were plated in a 6-well multiwell plate to reach a confluence of 80% at the moment of the DNA extraction. The day after cells were collected and resuspended in 200μL of PBS after centrifugation for five minutes at 300 x g . Then the DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen). We compared the DNA levels of *tTA* and *d2EYFP* in NOPFL cells and PFL cells by RealTime PCR following DNA extraction, proving that the both cell populations carry a unique copy of the networks in their genome. Quantitative RealTime PCR reaction were set up in duplicates using the LightCycler 480 SYBR green master mix (Roche) and the amplification was performed using a LightCycler 480 RealTime PCR instrument(Roche). The PCR were carried out using the following primers: d2EYFP forward (5'-acgacggcactcaagacc-3'); d2EYFP reverse (5'-gtcctccttgaagtcgatgc-3'); PFL tTA forward (5'-aaagcagctgaagtgcgagag-3'); PFL tTA reverse (5'-gatggtgctgccgtagttgtt-3'); NO PFL tTA forward (5'-acagcgcattagagctgctt-3'); NO PFL tTA reverse (5'-acctagcttctgggcgagtt-3'). Data analyses were performed using the LightCycler 480 Software(Roche). *GAPDH* DNA levels were used to normalise the amount of DNA and ΔCts were calculated as the difference between the average *GAPDH* Ct and the average *tTA* and *d2EYFP*.

Image acquisition and processing

Images were acquired using an inverted epifluorescence microscope (Nikon Eclipse TI-E, Nikon Instruments) equipped with an incubation chamber (H201-OP R2 ,Okolab), a digital camera (Andor Clara, Andor), a 20X objective (Obj. CFI PF DLL 20X Ph1, Nikon) and a 512-nm/529-nm (B/G/R) d2EYFP-specific excitation/emission filter set. Temperature was maintained at a constant level as the experimental setup required, while CO2 concentration was set to be 5% of the total air volume injected in the incubation chamber. Both phase-contrast images and fluorescent fields were acquired at intervals of 15 minutes. Exposure times for the phase-contrast field was set to 2 ms (transmitted light lamp voltage was set to 4.5 V) while 300 ms (Intensilight lamp set at 10% of the maximum power) was chosen as exposure time for the fluorescent images: this choice was meant to prevent photobleaching while optimising the ratio between the quality of the images and reflected-light-induced stress on the cells. Experiments were carried out using NIS-Elements AR v.3.10 644 (Nikon Instruments) software package and the Perfect Focus System (Nikon Instruments) to maintain the same focal plane during the whole duration of the experiment. At the end of the acquisition process, images were extracted as raw data for the fluorescence quantification procedures.

The experiments were set up so that at the beginning of each experiment the first image contained at least 15 cells and no more than 30 cells, to avoid cells exiting the image during the time-lapse experiment due to cell replica-

tion and "over-crowding". Image segmentation was carried out in Mathworks Matlab R2010b (Mathworks Inc.); the algorithm we implemented to quantify fluorescence was meant to distinguish the foreground (living cells) from the background in each image of the bright field. We used morphological operators such as erosion and dilation (*imerode* and *imdilate* functions from the MATLAB image processing toolbox). Thus two binary masks were built in order to compute separately the mean d2EYFP fluorescence of the foreground and the background using an element by element matrix multiplication between the binary images and the fluorescent one. The average fluorescence intensity across the cell population was then computed as the difference between the foreground and the background for each image at each time point (i.e. no single cell fluorescence quantification is performed).

Experimental procedure: determination of d2EYFP half-life

To evaluate d2EYFP degradation rate, 500 stably integrated CHO tetOFF cells were plated in chamber slide (lab-Tek) and, after cell adhesion, Cyclohexamide (Sigma, stock dilution 10 mg/ml in sterile water) was added to the medium to a final concentration of 10μg/mL, 50μg/mL, 100μg/mL or 500 500μg/mL. Temperature was maintained at 32 °C. Image acquisition and analysis was performed as described above. The experimental data were fitted into an exponential curve using the curve fitting tool (cftool) from Matlab

2010b, and the degradation coefficient d_3 was used to obtain the half-life (τ) of the d2EYFP protein: τ=log(2)/ d_3

Model simulations and parameter identification

Numerical simulations were run using Matlab 2010b (Mathworks Inc.). We used *ode23s* solver (a detailed discussion of the numerical methods used by *ode23* can be found in [14]). For the parameter identification, we used the PottersWheel toolbox [58] implemented in MATLAB. Two sets of parameters were identified: the dynamical parameters governing the model and a scaling factor meant to approximate the transduction contribution of the microscopy equipment. Since Doxycycline has been only added at time $t = 0$ min in our experiment we forced the fitting procedures to start from the model predicted ON steady state.

We defined the following objective function:

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_{Exp}(i) - y_{Model}(i))^2}{\sigma_{Exp}(i)^2} \tag{3.8}$$

where N is the number of experimental data points, y_{Model} are the predicted values of the mathematical model (using the inferred parameters), y_{exp} are the experimental data points and $\sigma_{Exp}(i)^2$ is the sample variance computed over the experimental replicates.

As optimisation algorithm we used Trust Region Method (TRM) in a logarithmic parameter space: at the k^{th} iteration of the optimisation procedure

the TRM approximates the shape of the function f to be minimised with the model m_k thus trying to solve the following problem:

$$\min_p m_k(x_k + p_k) \tag{3.9}$$

being $x_k + p_k$ the new parameter vector considered as solution at the k^{th} iteration. If the model m_k has quadratic form the vector p_k can be obtained by observing that:

$$m_k(x_k + p_k) = f_k + p_k^T + \nabla f_k + \frac{1}{2} p_k^T B_k p_k \tag{3.10}$$

and therefore

$$p_k = -\nabla_k \frac{\nabla f_k}{\| \nabla f_k \|} \tag{3.11}$$

In order to allow an extensive exploration of the parameters' space, and to avoid local minima, we used a quasi-random number generator routine in PottersWheel [58] to select an initial guess of the parameters' values, and then launched the TRM procedure M times (M=100 in our settings), requiring the cost function in eq. 3.8 to be $\chi^2/N \leq 0.5$ [58].

The values in Table 1 represent parameters for which the cost function (eq. 3.8) is the smallest across the M runs; whereas the standard devation of each parameter in Table 1 is evaluated by considering all of the M runs.

Moreover, in order to compare switch off times among the different experiments, we computed the τ_{off} defined as the time the circuit needed to

achieve the 50% of the mean initial fluorescence MIF calculated for each experiment as follows:

$$MIF = \frac{\sum_{i=1}^{3} f_i}{3} \tag{3.12}$$

with f_i fluorescence of the i^{th} frame in the sequence smoothed by moving average filtering.

Chapter 4

Construction of a post-transcriptional negative feedback loop in mammalian cells

This Chapter describes the implementation of a microRNA-based Negative Feedback Loop (NFL) in mammalian cells. In the first part, I present a work that we recently published in [29], to test experimentally mathematical models, so far described in literature, that formalize the effect of silencing RNAs on gene expression. The best performing model, was then used for outlining of the mathematical model of the NFL which will be detailed in Chapter 5. In the second part of the Chapter I will describe the construction

of the negative feedback loop and the experiments we performed to test its efficacy.

4.1 Introduction

In the last decade, 20-30 nucleotide RNAs have emerged as critical regulators of gene expression in eukaryotic genomes [20], involved in a process known as RNA interference (RNAi). RNAi is a cellular mechanism by which cells regulate gene expression at the post-transcriptional level [20]. It occurs through small RNAs, namely small interfering RNA (siRNA) and microRNA (miRNA) that, at the end of a cascade, direct sequence specific cleavage of perfectly complementary mRNAs, or translationally repress and degrade transcripts for imperfectly complementary targets [49]; nevertheless it has been highlighted that RNAi pathways can also induce transcriptional gene silencing in the nucleus, although the mechanism is not yet well understood [60, 86]. In this chapter I will describe the design, mathematical modeling, construction and characterization of a post-transcriptional negative feedback loop (NFL) in mammalian cells. The NFL described in this chapter will then be used together with the PFL described in Chapter 3, in order to consturct the oscillator described in Chapter 5.

4.2 Mechanism of RNAi-based gene silencing

In nematodes, insects and plants, the post transcription gene silencing functions as an innate defense mechanism. The RNA interference is based on the perfect or near perfect Watson-Crick base pairing between the guide (or antisense) strand of the siRNA and the mRNA target, ending up with the cleavage of the target by the RNA-induced silencing complex (RISC). In the cleavage process, the endonuclease Argonaute-2 (AGO-2) has a catalytic activity also in mammalian cells. AGO-2 seems to be involved first in the cleavage of the passenger (or sense) strand of the double-stranded siRNA, thus forming the single-stranded RNA that is used by the RISC complex as the guide strand to bind the target mRNA; then RISC can undergo multiple rounds of mRNA cleavage, mediating a robust silencing effect on the target gene.

In mammals, the endogenous miRNA pathway is involved in the fine-tuning gene expression during development and differentiation[11]. miRNAs bind sequences in the 3′ UTR of target genes they share partial sequence complementarity with; thus the silencing effect comes through translation repression, followed by mRNA degradation, which occurs in cytoplasmic compartments known as processing bodies (P-bodies)[55]. When instead the match between the miRNA and the 3′UTR of the target mRNA is perfect, again the silencing is mediated by RISC, which binds the target inducing its cleavage.

miRNA are generally transcribed by RNA polymerase II as long primary

transcript (pri-miRNA) and are processed by an RNase III enzyme Drosha, that works together with a dsRNA-binding protein of DiGeorge syndrome critical region gene 8 (DGCR8), into a 70 nt stem-loop structures known as precursor miRNAs (pre-miRNA) [53]. The pre-miRNA is then transported into the cytoplasm by the exportin 5; here Dicer and its dsRNA binding protein partners TRBP [22] and PACT [54] process the pre-miRNA, and the 22nt mature product is loaded into the RISC and eventually the silencing of target mRNA occurs (Figure 4.1).

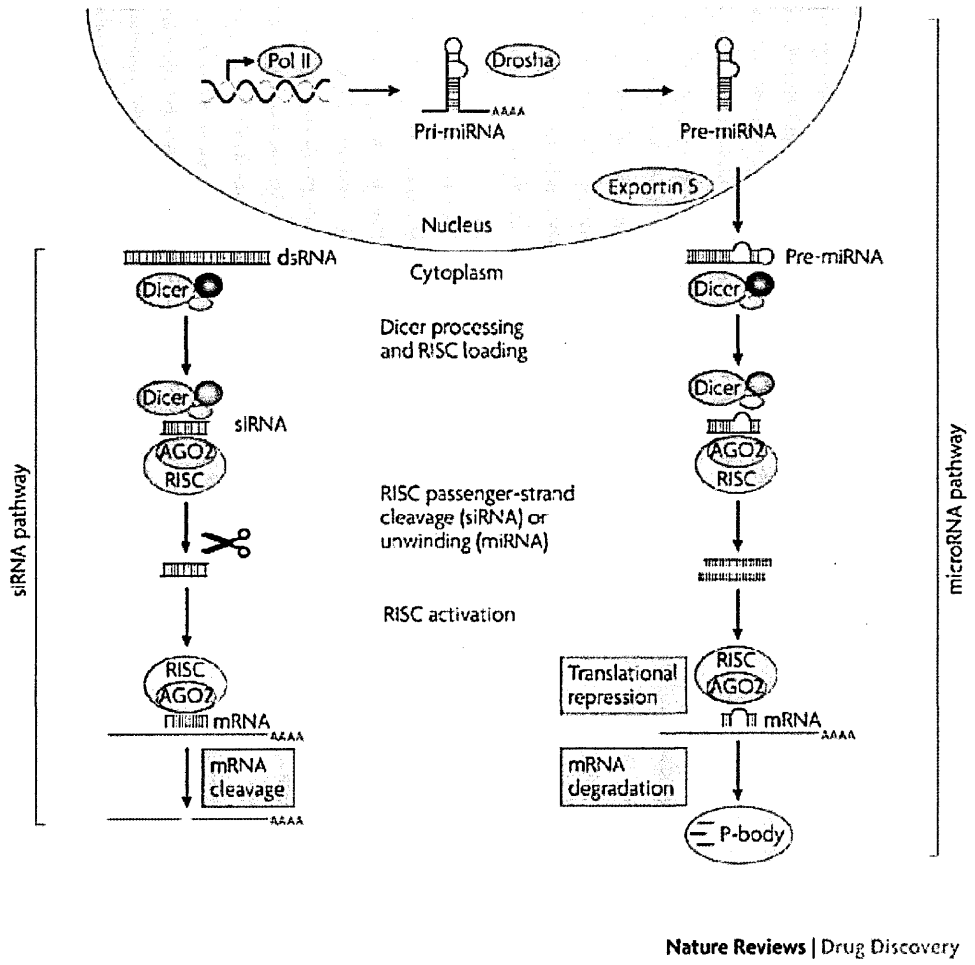

Figure 4.1: **Schematic representation of RNA interference in a mammalian cell.**

4.3 Modeling RNA interference in mammalian cells

Since the discovery of silencing pathway, RNAi has become not only a powerful tool for investigating gene function through the systematic knock down of the mRNA of interest using synthetic RNA oligomers, but it has been also used as "biological part" to create new synthetic biological circuits in synthetic biology, which needs mathematical models to carry out in silico predictions of biological processes [18, 37, 32, 80].

In Cuccato et al [29], four different mathematical models that could quantitatively describe the effect of different concentrations of siRNA on the mRNA target degradation, were compared in their ability to fit the experimental data with in silico predictions. I contributed to the experimental design of the RNAi experiments; the experiments were carried out on two different cell lines stably expressing the tTA protein or the EGFP protein respectively. We performed three sets of experiments. In the first set we transfected siRNA directed against the EGFP stably integrated in Human Embryonic Kidney (HEK) cells, in the range of 0 to 200 pmol, and evaluated the mRNA levels (Fig. 4.2).

In the second set of experiments, we measured the level of EGFP protein via FACS analysis 60 h after transfection (Fig.4.3).

Lastly, in the third set we performed the same silencing experiment but this time the siRNA was directed against the tetracycline regulated transac-

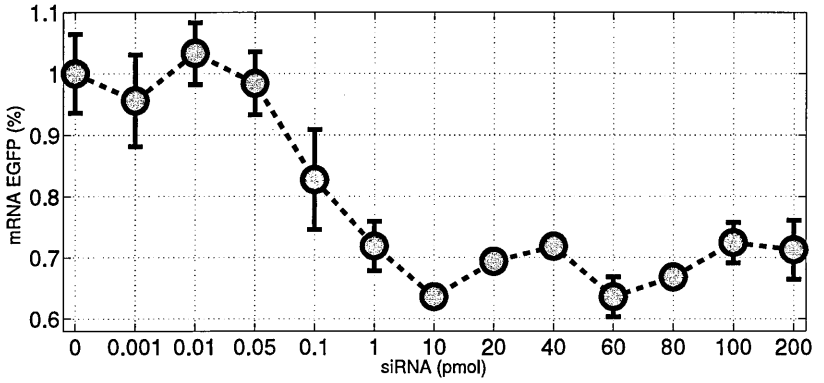

Figure 4.2: **Ratio of *EGFP* mRNA levels between cells transfected with the siRNAs specific for EGFP, and negative control cells, transfected with a non-specific siRNAs, measured 48 hours after transfection.** Error-bars represent the standard-error from three biological replicates for each point. The x-axis reports the different quantities of siRNA oligomers tested. mRNA levels were measured using real-time PCR. The error-bars have the length of one standard error.

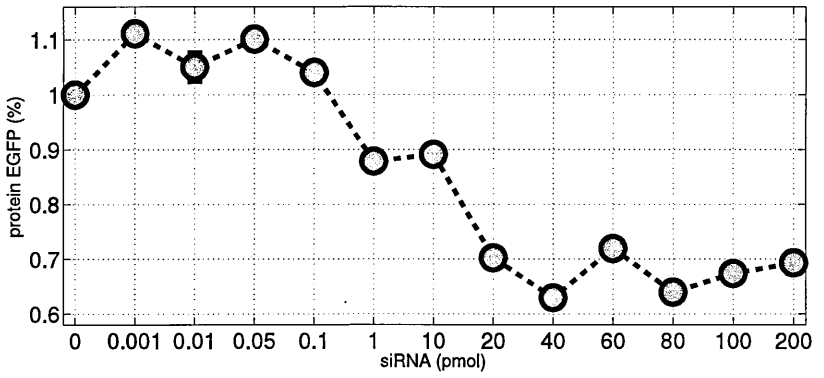

Figure 4.3: **Ratio of EGFP protein levels between cells transfected with the siRNAs specific for EGFP, and negative control cells, transfected with a non-specific siRNAs, measured 60 hours after transfection.** Error-bars represent the standard-error from three biological replicates for each point. The x-axis reports the different quantities of siRNA oligomers tested. Protein levels were measured using FACS analysis quantifying EGFP protein fluorescence. The error-bars have the length of one standard error.

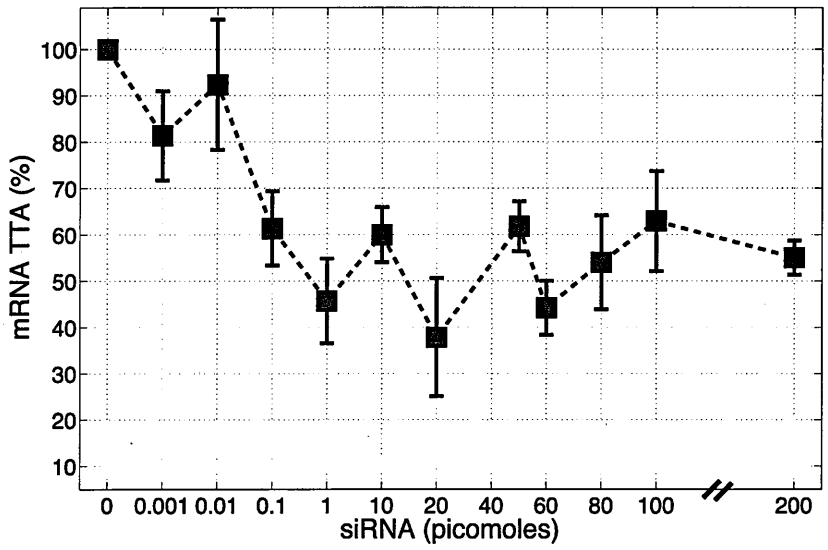

Figure 4.4: **Ratio of *tTA* mRNA levels between cells transfected with the siRNAs specific for tTA, and negative control cells, transfected with a non-specific siRNAs, measured 48 hours after transfection.** Error-bars represent the standard-error from three biological replicates for each point. The x-axis reports the different quantities of siRNA oligomers tested. mRNA levels were measured using real-time PCR. The error-bars have the length of one standard error.

tivator tTA stably expressed at low level in CHO cells (CHO AA8)(Fig. 4.4).

The four mathematical models are based on the general approach described by ordinary differential equations (ODEs), which model the evolution over time of the transcriptional and translational processes, as also described in Chapter 3:

$$\begin{aligned} \text{[mRNA]:} \quad & \frac{dX_m}{dt} = k_m - d_m X_m - \delta(X_m, X_s) \\ \text{[protein]:} \quad & \frac{dX_p}{dt} = k_T X_m - d_p X_p, \end{aligned} \tag{4.1}$$

Where X_m,X_p and X_s are the mRNA, protein and siRNA concentrations, respectively. The parameter k_m, represents the transcription rate from the promoter that transcribes the mRNA species targeted by the siRNA oligomer; d_m represents the basal degradation rate of the mRNA species. RNAi can be considered as a mechanism that enhances the degradation of the targeted mRNA, therefore the function $\delta(X_m, X_s)$ is an extra degradation term that represents the rate at which mRNAs are degraded due to RNAi.

This function, $\delta(X_m, X_s)$, depends on both the mRNA and siRNA levels, X_m and X_s respectively. The parameter k_T is the protein translation rate, whereas d_p represents the basal protein degradation rate. We then fitted the four models to the experimental data, by searching for the parameter values for which the fitting was the best according to a squared error measure. We found that the best model proposed in literature was a phenomenological one, in which the post-transcriptional effect of the microRNA on the gene expression is described with a standard Hill-kinetic model[48]. In particular, considering a Hill-type enzymatic model with a Hill coefficient h_1, the model can also be extended to account for multiple binding sites of the siRNA on the target mRNA, or for cooperativity of protein complexes involved in RNAi.

This model is described by the following equation:

$$\delta(X_s, X_m) = d_4 \frac{X_s^{h_4}}{{\theta_4}^{h_4} + X_s^{h_4}} X_m \tag{4.2}$$

The kinetic parameters d_4 and θ_4 depend on the efficiency of siRNA binding to its sites on the target mRNA [48]: d_4 represents the maximal degradation rate of the mRNA due to RNA interference; θ_4 the concentration of siRNA oligomers needed to achieve half of the maximal degradation rate. The above equation implies that for $X_s << \theta_4$, the increase in the RNAi mediated degradation is linear with $X_s^{h_4}$, while it saturates at higher levels of X_s, reaching the maximal degradation rate d_4. The fitting results for the mRNA and protein levels of EGFP are shown in Fig. 4.5 and Fig. 4.6; finally we observed that the experimental error for the third set, was larger compared to the EGFP experiment. This is probably due to the relative low expression of the tTA in the CHO TET-OFF cell lines, which made the measurements more noisy (figure 4.7).

The superior performance of the phenomenological model has been proven also by the "leave-one-out" cross validation strategy, where for each model and for each dataset, the parameter identification procedure was repeated each time removing one of the experimental points and then predicting the missing point with the identified parameters; the phenomenological model (Model 4) is again the one with the smallest prediction error. One significant feature of Model 4 is that it can predict the saturation effect of the

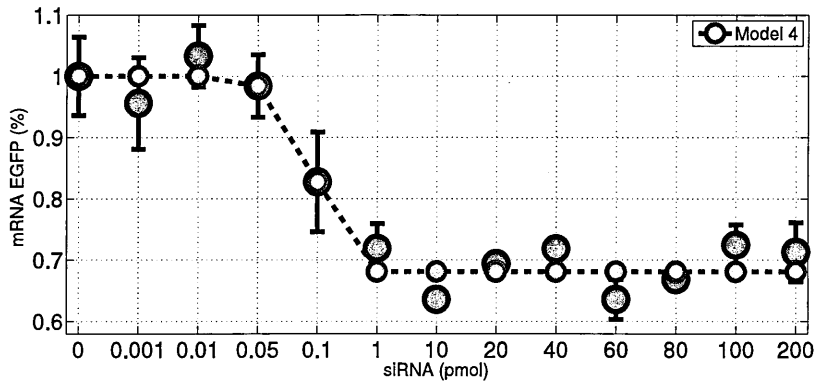

Figure 4.5: **Numerical fitting of the model on the in vitro experimental results on mRNA *EGFP*.**

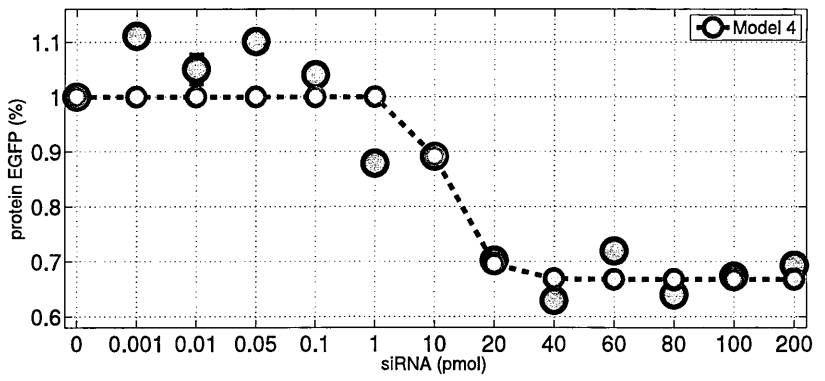

Figure 4.6: **Numerical fitting of the model on the in vitro experimental results on protein EGFP.**

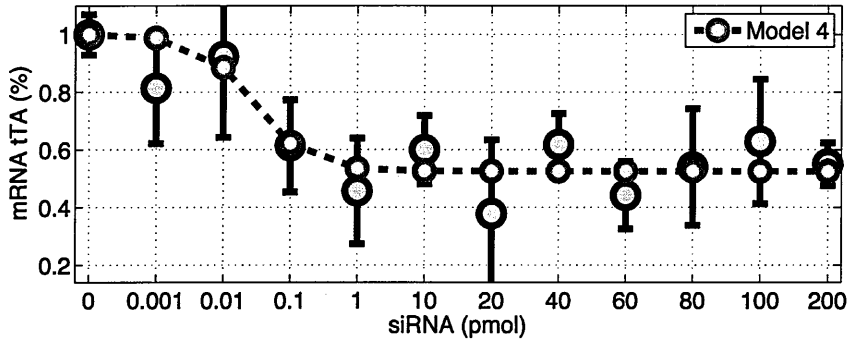

Figure 4.7: **Numerical fitting of the model on the in vitro experimental results on mRNA *tTA*.**

RNAi process that we observed experimentally. We considered the possibility that this saturation could be in fact due to the inability of the cell to uptake high concentration of siRNA oligomers, however the experiments in [67], prove that uptake of siRNA oligomers in cells is linear with the concentration of siRNA oligomers transfected, in the concentration range we used. Additionally, Khan *et al* in [47], observed upregulation of mRNA targets of endogenous micro-RNA when transfecting siRNA oligomers in mammalian cells. As an explanation of this effect, they suggested a saturation of the RISC complex (or other components necessary for small RNA processing or transport machinery).

4.4 Design of a microRNA-based negative feedback loop

Commonly, siRNA are transiently transfected within the cells, which is a dramatic limit for the long term knock-down of a gene of interest. Hence other strategies have been developed; one is the use of shRNAs that mimic the pre-miR structure and are processed by Drosha to be functional. The stable expression of shRNAs is driven by polymerase-III promoters[17] which have also been engineered for inducible expression[79], but they lack the tissue specificity which is instead possible to achieve with polymerase-II promoters. Moreover it has been reported the potential toxicity of this approach, possibly due to the saturation of miR processing steps, thus interfering with the endogenous miR pathway [61, 42, 21].

In order to create a post-transcriptional negative feedback loop (NFL), we used a natural microRNA (miR223), which is expressed to high levels only in myelomonocytic cells, and almost undetectable in other cell lines[16]. Furthermore, as shown in Naldini *et al* [2] the miR223 was inserted in the *EF1α* intron such that the microRNA is spliced out, and so the cleavage of pre-miRNA by Drosha does not affect the integrity of the whole mRNA that otherwise would lose the 5′-cap or the poly-A tail thus compromising the protein production [2].

The lentiviral platform we developed is therefore composed by the tTA-induced promoter *CMV-TET* driving the expression of a unique mRNA con-

sisting of the *EF1α* intron containing the miR223, of the *mCherry* coding sequence codifying a stabilised fluorescent protein as readout of microRNA expression and of a woodchuck hepatitis virus post-transcriptional element (WPRE) sequence needed to stabilize the viral mRNA, as shown in Fig. 4.8). We named this inducible microRNA construct, the Negative Feedback Loop (NFL) because of its ability to repress a target mRNA in an inducible way, thanks to the presence of the CMV-TET inducible promoter.

Following the same experimental procedure as for the construction of the PFL described in Chapter 3, we took advantage of the Gateway Technology (Invitrogen) optimized for lentivirus production (Materials and Methods).

Specifically, I produced two different lentiviral vectors: one containing the inducible NFL and another in which the miR223 expression is controlled by the constitutive *CMV* promoter(*CMV*-miR223 plasmid). The latter plasmid was built to test the efficacy of the miR223.

I performed two set of experiments; the first was a co-transfection in HEK cells in which, together with the lentiviral plasmid, I transfected a vector containing the *EGFP* coding sequence fused with four tags responsive to the miR223, or a vector containing only the *EGFP*, respectively. In the second set, I transfected PFL HEK293 cells, described in Chapter 3, with the *CMV*-miR223 plasmid. The miR223 was able to knock down the expression of the EGFP, as well as the d2EYFP, by more than 80% as shown following quantitative RealTime PCR, whose results are in Fig. 4.9, and in Fig. 4.10).

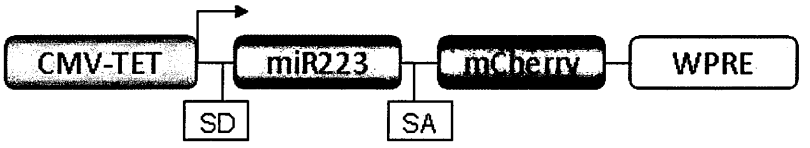

Figure 4.8: **Schematic representation of the miR223-based negative feedback loop.** The expression of the *EF1α* intron including miR223 and of the *mCherry* is driven by the *CMV-TET* promoter. WPRE, woodchuck hepatitis virus post-transcriptional element; SD, splice donor site; SA, splice acceptor site.

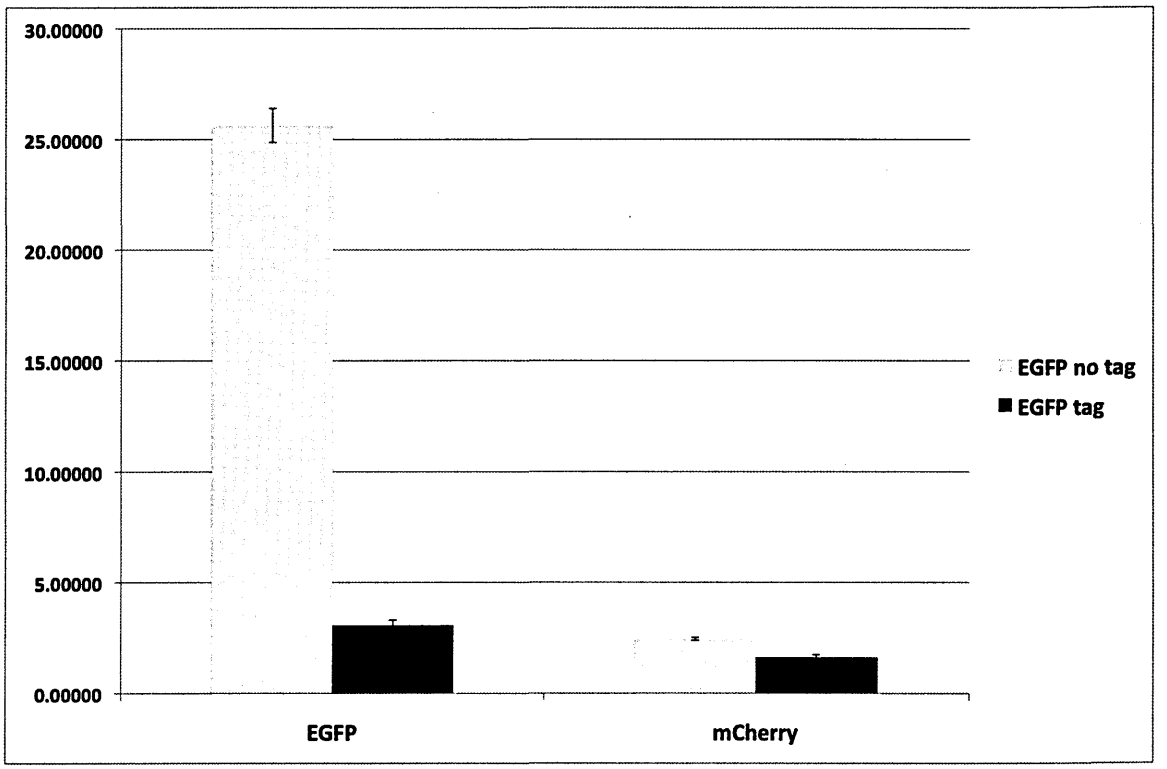

Figure 4.9: **HEK cells were co-transfected with the lentiviral vector carrying the miR223 controlled by a *CMV* promoter and a plasmid containing an *EGFP* with four tags for the binding of the miR223 (EGFP tag).** As negative control we used the same plasmid expressing the EGFP but lacking the tags (EGFP no tag). The efficiency of the knocking down due to the microRNA effect is shown in the real-time PCR.

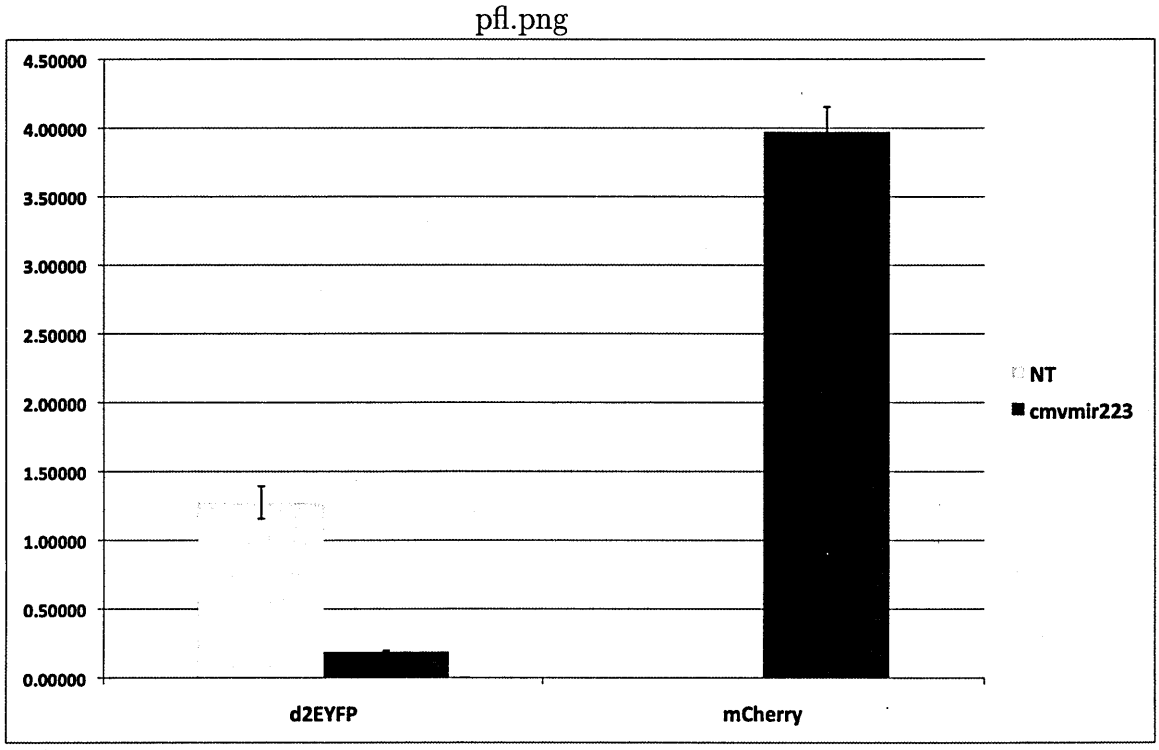

Figure 4.10: **miR223 plasmid transfection in PFL-HEK cells.** PFL-HEK cells were transfected with the *CMV*-miR223 plasmid. The mRNA level of the d2EYFP were compared to a mock control (same lentiviral plasmid not carrying the *miR223-mCherry* gene expression cassette), as shown in the real-time PCR.

4.4.1 Conclusion

In this chapter I proved that the miR223 is suitable for our synthetic network, and that it is able to strongly repress the transcription of a target mRNA where four repeats of a perfectly complementary microRNA seed-sequence are present.

I also contributed to derive a mathematical model to quantitatively described the effects of RNAi on mRNA degradation.

In Chapter 5, I will describe how the positive and negative feedback loops were integrated in mammalian cells to generate the synthetic clock.

Materials and Methods

RNA interference by small interfering oligonucleotides (siRNA)

The sequence of the 21-mer siRNA double-stranded oligomers targeting EGFP was identical to the one reported in[?]. This siRNA targets the coding sequence of the EGFP gene starting at position 237 from the ATG, on the target sequence AAGCAGCACGACTTCTTCAAG. The siRNA HPLC purified, with sequence GCAGCACGACUUCUUCAAGtt (concentration 100 μM) was synthesized by Ambion. As a negative control we used, in all experiments a shuffled sequence non targeting siRNA from Dharmacon.

Cell culture and transfection

HEK 293 stably expressing EGFP (kindly provided by Mara Alfieri) were maintained at 37°C in a 5% CO2-humidified incubator. HEK 293 cells were cultured in Dulbecco's modified Eagle's medium (DMEM, GIBCO BRL) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Invitrogen) and 1% antibiotic/antimycotic solution (GIBCO BRL). Cells were seeded at a density of 300.000 per well in a 6 wells multi-well and transfected 1 day after seeding using Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions with siRNA (Silencer Custom siRNA, 100μM, Ambion) in a range of quantities from 0.001 pmol to 200 pmol

(total concentration). The amounts of transfected siRNA oligomers were: 0, 0.001, 0.01, 0.05, 0.1, 0.5, 1.0, 10.0, 20.0, 40.0, 60.0, 80.0, 100.0 and 200.0 pmol in a total of 2 mL of medium (so the final concentrations of siRNA oligomers were $5x10^{-4}$, $5x10^{-3}$, $2.5x10^{-2}$, $5x10^{-2}$, $2.5x10^{-1}$, $5x10^{-1}$, 5.0, 10.0, 20.0, 30.0, 40.0, 50.0, and 100 nM respectively). Each experiment was performed in biological triplicates, and the resulting standard deviations were computed and reported in each graph. One day post-transfection, the media and ligand were replaced. Transfected cells were collected 48 hours post-transfection for RNA extraction and subsequent analysis. FACS analysis was performed 60 hours after transfection. Plasmid transfections were performed in 6-well plates at the same conditions described above using Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions.

RNA extraction and Real-time PCR

Total RNA extraction from 35mm culture plates was performed using the Qiagen RNeasy Kit (Qiagen) according to manufacturers instructions.

Retro-transcription of 1 μg of the total RNA extracted was performed using the QuantiTect®Reverse Transcription Kit (Qiagen), according to manufacturers instructions.

Quantitative real-time PCR was performed using a LightCycler (Roche Molecular Biochemicals, Mannheim, Germany) to analyze the amplification status of EGFP and tTA. Amplification of the genes was performed from the cDNA obtained from the total RNA and using the LightCycler DNA Mas-

ter SYRB Green I kit (Roche Molecular Biochemicals). Primer sequences for Human GAPDH used as reference genes, were designed by Primer 3.0 http://frodo.wi.mit.edu/ (Forward primer Human: GAAGGTGAAGGTCG-GAGTC; Reverse primer Human: GAAGATGGTGATGGGATTTC). Primer sequences for EGFP were also designed with Primer 3.0 (Forward primer EGFP: ACGACGGCAACTACAAGACC; Reverse primer EGFP: GCATC-GACTTCAAGGAGGAC); d2EYFP forward (5'-acgacggcactcaagacc-3');
d2EYFP reverse (5'-gtcctccttgaagtcgatgc-3');
mCherry forward (5'-cactacgacgctgaggtcaag-3');
mCherry reverse (5'-gtagtcctcgttgtgggaggt-3'). The relative amounts of mR-NAs were compared with the reference gene GAPDH and calculated using the Principle of Relative Quantification Analysis according to the standard formula 2^{-DCt}. To confirm the specificity of the amplification signal, we considered the primer dissociation curve in each case.

FACS analysis

Cells from 35mm culture plates were trypsinized, filtered and subjected to Fluorescence-Activated Cell Sorting (FACS) analysis 60 hours posttransfection in a Becton Dickinson FACSAria.

Experimental procedure: construction of the Amplified Negative Feedback loop

The generation of the Negative feedback loop is described in chapter 5. The control in which the miR223-mCherry is under control of a *CMV* promoter has been developed following the same procedures. To integrate the *CMV* promoter in the pENTR5'-TOPO vector, the sequence was amplified by a pCMV-Myc vector (Clontech). The The PCR was performed with the Taq polymerase provided by Invitrogen that adds a single deoxyadenosine (A) to the 3' ends of PCR products. This allows PCR inserts to ligate efficiently with the pENTR5'-TOPO vector which is supplied linearised with single 3'-deoxythymidine (T) overhangs, obtaining the pENTR5'-TOPO-*CMV* with specific recombination sites.

Models

In the context of the specific in vivo experiments we carried out, we can make the following assumptions to derive the mathematical model:

1. The HEK293 cells express mRNA EGFP at a constant rate k_m which corresponds to the maximum transcription rate of the promoter.

2. We assume that the siRNA oligomers will be quickly loaded into the RISC and that this step takes place in much shorter time scale than targeting and degrading the cognate mRNA.

Therefore, the entire RNA interference cascade can be described by Equations 4.1.

The negative control experiments involved the addition of non-specific siRNA oligomers, which are not complementary to the mRNA EGFP and therefore are not able to trigger the RNA interference mechanism. Namely, $\delta(X_m, X_s) = 0$. Therefore, the equations corresponding to the negative control experiments are:

$$\begin{aligned} \text{[mRNA]:} \quad & \frac{dX_m}{dt} = k_m - d_m X_m, \\ \text{[protein]:} \quad & \frac{dX_p}{dt} = k_T X_m - d_p X_p, \end{aligned} \tag{4.3}$$

Steady-state equations

For the numerical fitting of the in vivo experiments we used the steady state equations for the mRNAs or proteins. For example, for the in vivo experiments on RNA levels, the experimental period of 48 hours before extracting the RNA is considered long enough for the mRNAs to approach their equilibrium value. In order to solve for the mRNA or protein steady state we assume that siRNA concentration remains constant through the 48 hours of the in vivo experiments. In general, the siRNA-RISC complex, is considered very stable and one can assume that the degradation of siRNA is so slow that it does not have any effect on the overall dynamics.

The steady state equations for the mRNA concentrations of the four

models are:

$$\begin{aligned}
&\text{Model 1:} \quad \tilde{X}_m = \frac{k_m}{d_m + k_1 X_s} \\
&\text{Model 2:} \quad \tilde{X}_m = \frac{k_m}{d_m + k_2^{h_2} X_s^{h_2}} \\
&\text{Model 3:} \quad \tilde{X}_m = \frac{-B + \sqrt{B^2 + 4c_3 k_3 h_3 k_m d_m}}{2k_3 h_3 d_m} \qquad \text{where } B = (c_3 d_m + c_3 k_3 h_3 X_s - k_3 h_3 k_m) \\
&\text{Model 4:} \quad \tilde{X}_m = \frac{k_m(\theta^{h_4} + X_s^{h_4})}{\theta^{h_4} d_m + (d_4 + d_m) X_s^{h_4}}
\end{aligned} \tag{4.4}$$

The corresponding mRNA equilibrium of the negative control experiments is simply $X_m = k_m/dm$ (for all the models since $\delta(X_m, X_s) = 0$). Therefore, when fitting the ratio of the mRNA levels between positive and negative control, we multiply equations (4.4) by d_m/k_m.

Through numerical optimization, the degradation rate of mRNA EGFP was fixed at the value of $d_m = 0.0173\text{min}^{-1}$, which corresponds to a half-life of 40 minutes. This value was taken from a recent paper on synthetic biology [80] in which the degradation rate of GFP mRNA was experimentally determined. In order to optimize Model 3 with the smallest possible number of parameters, we clustered its 4 different parameters (k_3, h_3, c_3, k_m) in order to have only two optimized quantities: $k_3 \cdot h_3$ and c_3/k_m.

For the in vivo experiments in protein levels, we fitted numerically the protein steady-state equations. The equilibrium concentration of protein is given by:

$$X_p = \frac{k_T}{d_p} \tilde{X}_m \tag{4.5}$$

where $\tilde{X}_m$ is the mRNA equilibrium, which is different for each model (equations 4.4). Additionally, the protein steady-state of the negative control model is:

$$X_p = \frac{k_T}{d_p}\frac{k_m}{d_m}. \tag{4.6}$$

For the numerical fitting of the ratio of protein levels between negative and positive control, one needs to divide equation (4.5) by equation (4.6).

Cost error functions

For the numerical fitting of the mRNA levels from in vivo experiments, we used the following cost error function:

$$\sum_{i}^{N} \left(\frac{Y^i_{model} - Y^i_{data}}{SE^i} \right)^2 \tag{4.7}$$

where N is the number of experimental points, Y^i_{data} is the experimental measurement of experiment i, Y^i_{model} is the model prediction for experiment i and SE^i is the standard error of experiment i. The absolute value errors of each model were then normalized against the largest error.

Chapter 5

Synthesis and Analysis of an amplified negative feedback loop oscillator in mammalian cells

In this chapter I will portray the construction of the genetic oscillator via systematic integrations of the positive and negative feedback loop in CHO cells and HEK 293 cells. First, I will give an overview of the attempts reported in literature to achieve an oscillatory behavior, based on different types of repressions. Then, I will detail the rational implementation of our amplified negative feedback loop oscillator.

5.1 Construction of the oscillator in mammalian cells.

Over the past few years numerous designs to achieve oscillatory behavior have been proposed and analyzed *in silico* and *in vivo* both for prokaryotic and eukaryotic systems [41, 38, 10, 25, 81]. Although natural genetic oscillators, involved in the regulation of cell cycle, circadian rhythms, and signaling pathways are based on the interaction of many components forming complex regulatory networks [43], also simpler oscillator architectures have been found in different biological processes; in *Xenopus laevis* embryos i.e. it has been shown that the regulation of cell cycle is given by an oscillator consisting of negative and positive feedback loops; the combining of a two-component negative feedback loop between *cdc2* and the anaphase promoting complex, and a positive feedback centered on *cdc2*, leads to robust oscillations [68]. Further, it has been recently shown that oscillation of a single monomolecular CDK module is sufficient to trigger cell cycle events even in absence of many of known regulatory inputs and feedbacks in yeast [26]

Among the variety of oscillators described the most common architecture is based on an amplified negative feedback loop, as shown in Fig. 5.2.

So far, three different possible kind of repressions have been considered for the amplified negative feedback loop: repression at transcriptional level [7], repression by dimerization [10] and repression due to proteolytic degradation [43, 25].

5.1.1 Repression by transcriptional control

In Atkinson *et al* a genetic clock, shown in Fig.5.1, has been obtained in E.Coli by combining two modules: the Ntr and Lac systems. Both these systems originate from *Escherichia coli.*

The Ntr regulon of *Escherichia coli* consists of genes and operons that are regulated by the availability of ammonia and whose products facilitate survival under nitrogen limiting growth conditions. The most important enzyme for the assimilation of ammonia under nitrogen-limiting conditions is glutamine synthetase (GS), the product of the glnA gene. Under nitrogen-excess conditions, a low intracellular concentration of this enzyme results, regulated by the *glnAp1* promoter. Under nitrogen-limiting conditions, a much higher intracellular concentration of this enzyme results from activation of the *glnAp2* nitrogen-regulated promoter from the phosphorylated form of NRI (NRI P); the same NRI P controls the *glnK* promoter whose protein product GlnK also is required to regulate expression of Ntr genes [6, 5, 39]. The lac operon comprises three genes required for the uptake andmetabolism of lactose and related sugars : *lacZ*, *lacY* and *lacA*. Lactose in converted into allolactose by an enzyme, b-galactosidase, encoded by *lacZ*. *lacY* encodes for the lactose permease (LacY), which facilitates the uptake of lactose and similar molecules. *lacA* codify for an acetyltransferase, which is involved in sugar metabolism. The operon has two transcriptional regulators: a repressor

(LacI) and an activator (the cyclic AMP receptor protein, CRP). Inducers, among them allolac- tose, bind to and inhibit repression by LacI, whereas cAMP binds to and triggers activation by CRP.

The operons were engineered in order to obtain an oscilatory network. In particular, the activator module consists of the control region of the *glnA* promoter (*glnAP2*) upstream the *glnG* gene encoding the NRI thus creating an auto-activated circuit; additionally two "perfect" LacI operators, one downstream of the promoter and one immediately upstream of the enhancer were placed in order to be responsive to the LacI repressor.

The repressor module was constructed by placing the *lacI* repressor gene under the control of the *glnK* promoter. Therefore, the *glnK* promoter is fully activated at high NRIp levels. The oscillatory behavior is achieved by the antagonism between the activator and repressor to form, in a mutual exclusive way, a DNA loop; during activation NRIp binds the enhancer and interacts with the RNA polymerase by means of an activation of a DNA loop; when instead LacI is bound to the two operators it also forms a loop repressing the system [7] (Figure2.3).

Mathematical analysis

In [7], a Hopf bifurcation, is proposed to explain the onset of oscillatory behaviour. The Hopf bifurcation can arise from significantly faster activator than repressor dynamics, causing the activator concentration to reach a significant amplitude before the repressive effect becomes too great, ending up

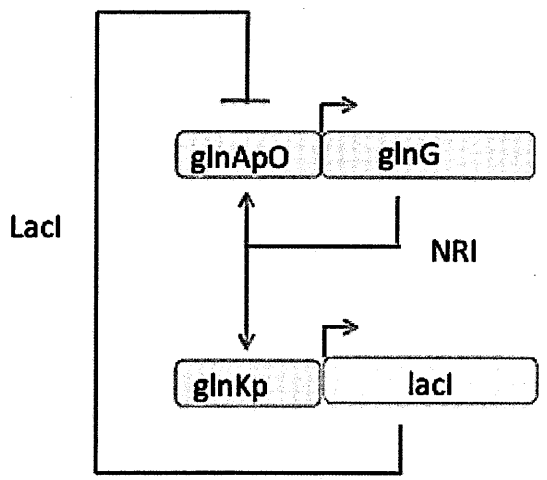

Figure 5.1: **Genetic clock by combination of activator and repressor modules in *E.coli*.**

with sustained oscillations. For the same topology, in [43], the oscillations are mathematically explained in terms of a saddle-node bifurcation on an invariant circle (SNIC), while under different parameters value oscillatory behavior can originates from a SNIC or a subcritical Hopf bifurcation [25].

5.1.2 Repression by dimerization

In this second type of repression, the gene x encoding a transcriptional activator, is self-regulated, and also induces the expression of the repressor Y. Once the repressor is produced, by dimerizing with the activator X, it prevents it by binding its own promoter, thus inhibiting the transcription (Fig. 5.2) [10]. For this topology, a mathematical analysis has been proposed together with *in silico* simulations, but it has never been implemented *in vivo*.

Mathematical analysis

Although this network has never been implemented *in vivo*, mathematical modelling was studied by using ODEs. In order to justify the existence of oscillations, the Poincare'-Bendixon theorem was used on a simplified model that describes only the evolution of the repressor and the activator-repressor complex. This model was able to capture qualitatively the main features of a more complete model, and the oscillatory behaviour was detected over a broad range of parameters value, demonstrating its robustness, but it requires an intermediate repressor degradation rate.

5.1.3 Repression by proteolysis

The third kind of repression is *repression by proteolysis* (5.2); the repression occurs through the degradation of the activator (gene$x$), by a protease (gene y), whose expression is driven by the activator X. As in the previous case this concept has been studied mathematically and *in silico* experiments have been carried out, but it has never implemented *in vivo*.

Mathematical analysis

As for the repression by transcriptional control, oscillations occur through a subcritical Hopf bifurcation; this requires faster dynamics of the activator compared to those of repressor (faster activator degradation and translation rates)[43]. In [25] it was shown that in this topology oscillation arise through a SNIC, and that these oscillations exist for a significant range of values of the ratio between the activator and repressor degradation rate rates [43].

5.2 A microRNA based oscillator.

The genetic circuit, which we developed, is an amplified negative feedback oscillator where repression happens at the post-transcriptional level; it consists of two modules: an auto-catalitic loop with a transcriptional activator amplifying its own transcription, and a repressor whose expression is driven by the transcriptional activator, and represses it, as depicted in (Figure 5.2).

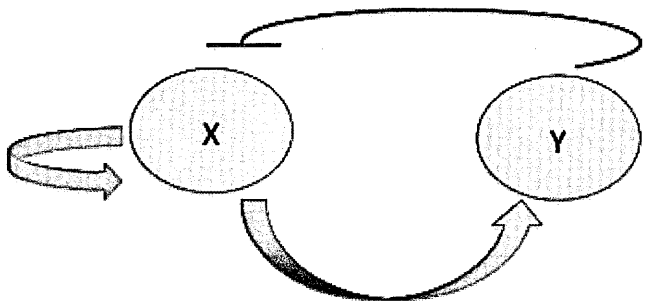

Figure 5.2: **Amplified negative feedback loop topology**. Gene X activates itself and geneY, while gene Y represses transcription from X. The topology is conserved for the respression by dimerization, and repression by proteolysis.

In order to construct the genetic oscillator with the features defined above, I combined the PFL and NFL previously described in Chapter 3 and 4.

The tTA transcriptional activator binds the *CMV-TET* promoter auto-regulating itself, together with the expression of the d2EYFP fluorescent protein; at the 3' end of the construct, I inserted four tags, 21 nt long, with perfect complementarity to the miR223 seed. A second *CMV-TET* promoter was placed upstream of the miR223 and *mCherry* reporter gene; following the mRNA production, the microRNA is spliced out of the intron, and once processed, the mature product is loaded into the RISC, and the *tTA-d2EYFP-223TAG* mRNA gets degraded. The entire cycle should induce an oscillatory dynamics Fig.5.3.

Human Embryonic Kidney 293 (HEK 293) and Chinese Hamster Ovary (CHO) cells were infected with the PFL virus and sorted according to the green fluorescence as described in Chapter 3.

As reported in Chapter 3, we generated monoclonal populations of the PFL-infected CHO and also of HEK 293 cells; the clones were then infected with the virus carrying the NFL.

To test all of the possible combinations, we also performed the same "cross-infection" but this time starting from a NFL-CHO (or NFL-HEK 293) clonal cells, and infecting them with the virus carrying the PFL.

A representative scheme of making clonal populations carrying both the PFL and NFL is shown in Fig.5.4

Since double infected cells could not be selected via antibiotic resistance

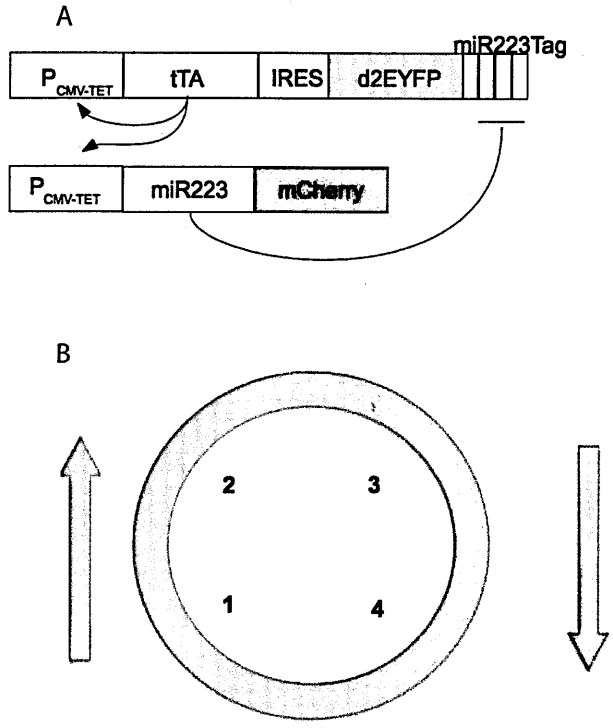

Figure 5.3: **Mammalian oscillator.** (A) Scheme of the genetic oscillator. The tTA transcriptional activator binds the *CMV-TET* promoter autoregulating itself, together with the expression of the d2EYFP fluorescent protein (positive feedback loop). The activation of the *CMV-TET* promoter triggers the miR223-mCherry production and the consequent degradation of the *tTA-d2EYFP* mRNA through the binding to four tags placed at the 3' end of the construct. (B) Autoregulated *CMV-TET* promoter-driven tTA expression triggers d2EYFP production (1). As the tTA and d2EYFP reach the peak (2) the miR223 is transcribed together with the mCherry (3) resulting in the downregulation of tTA and d2EYFP (4).

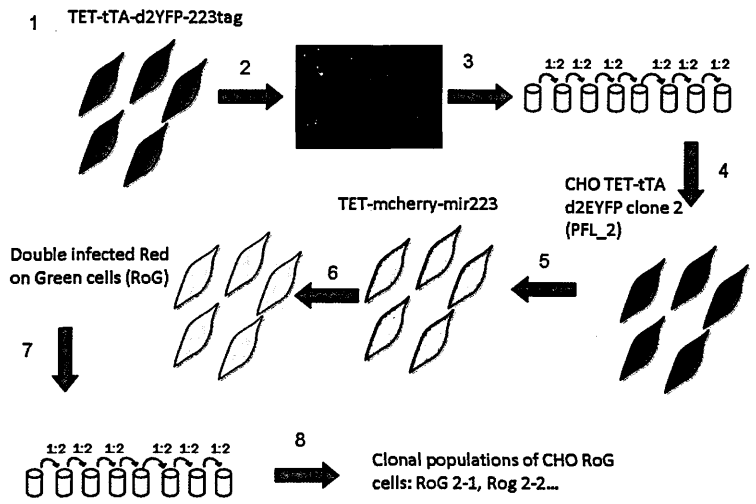

Figure 5.4: **Selection of CHO cells stably expressing the positive and negative feedback loops.** (1)CHO cells were infected with a virus carrying the positive feedback loop (*CMV-TET-tTA-d2EYFP*). (2) CHO cells positive to viral infection were selected either with blasticidine antibiotic or with FACS sorting. (3)Serial dilution of CHO PFL cells were made to generate monoclonal population (CHO-PFL cells). (4-5) CHO-PFL clonal population number 2 was infected with the lentivirus harboring the NFL; double infected cells were sorted via FACS according to the contemporary expression of green and red fluorescence . For simplicity they were called Red on Green cells (RoG cells). (7-8) Following serial dilutions, clonal populations of RoG cells were generated.

(both the PFL-carryng virus and the NFL-carrying one include the same blasticidine resistance gene), positive cells were sorted via FACS analysis according to their level of red and green fluorescence (CHO-Osc).

Monoclonal populations were then generated from the mixed population through single cell expansion. In order to check the quality of the resulting clones, we extracted genomic DNA and mRNA in order to evaluate the construct integration and its expression.

After clonal population selection, we then analyzed clones via fluorescence microscopy (Fig.5.5).

In order to keep track of the different clonal populations, we adopted a simple nomenclature, according to the sequence of infections. For example, the CHO-PFL monoclonal population number 2, infected with the NFL virus, were called RoG2 (Red on Green) cells; the clonal populations deriving from the double infection done on RoG2 cells were called RoG^2-1, RoG^2-2 and so on 5.4.

We obtained several clones of RoG2 CHO cells but no working clones of GoR (Green on Red) CHO cells, probably because despite the comparable number of integration events, the expression of miR223 from the NFL was much higher than the d2EYFP from PFL, thus completely repressing its expression.

Similarly, we were able to obtain GoR HEK 293 cells (Materials and Methods section). Experiments were performed, according to DNA integration, mRNA expression and fluorescence microscopy analysis, with HEK GoR-B

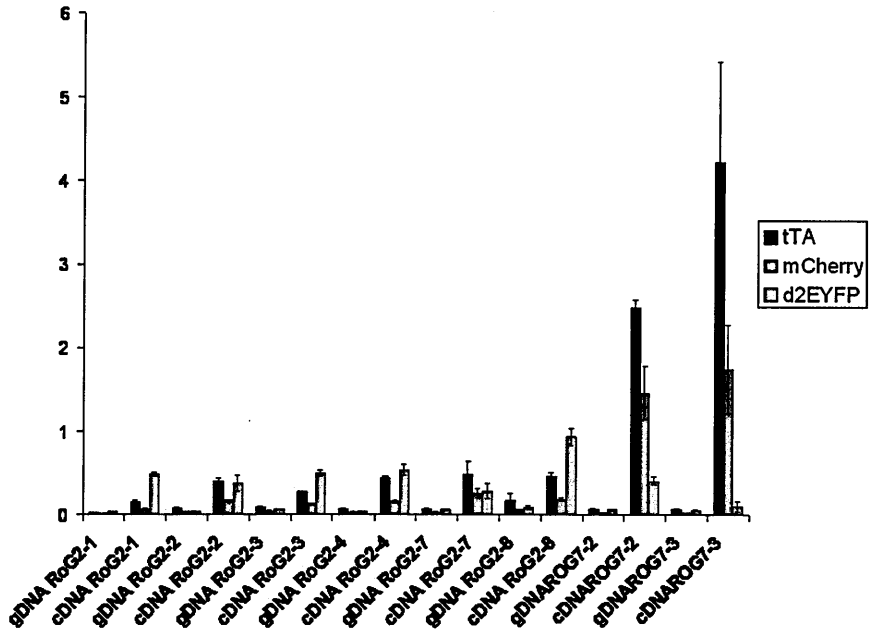

Figure 5.5: **Real-time PCR of genomic DNA (gDNA) and mRNA (cDNA) extracted from the monoclonal population of double infected RoG CHO cells.**

clonal population (figure 5.6).

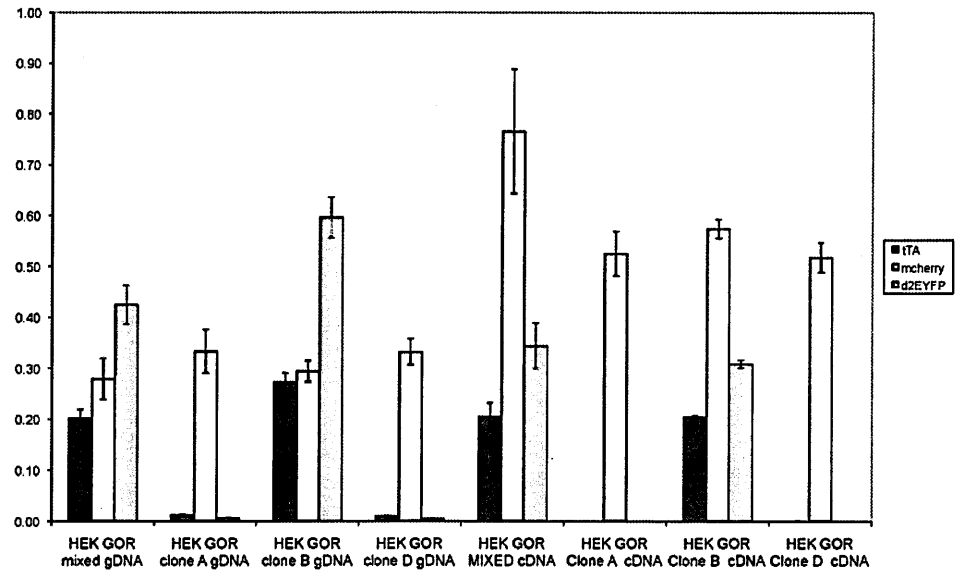

Figure 5.6: **Real-time PCR of genomic DNA (gDNA) and mRNA (cDNA) extracted from the monoclonal population of double infected GoR HEK 293 cells.**

5.3 Mathematical modeling of synthetic oscillator

As for the PFL motif, also for the genetic clock the mathematical model is based on ODEs, since we are considering the average behavior of a population of cells ([38, 40, 51, 81, 78]). We integrated in the previous model describing the dynamics of the auto-regulatory PFL, new equations to take into account the effect of the NFL on the whole system;

We assumed:

- Hill functions to model the rates of gene transcription, including basal

activity to describe the leakiness of the promoter (CMV-TET);

- linear degradation for all genes and proteins;
- Michaelis-Menten like function to model the miR silencing on tTA;
- linear dynamics for the translation process;
- Michaelis-Menten like modelling of the effect of the inducer (Doxycycline);
- distinct dynamics for inactive and active form of the microRNA to take into account maturation time. Thus, we used two separate variables to model the microRNA concentration: the first (x_3) represents the unprocessed miR , while the second (x_4) reflects the concentration of the active form after the miR has been cleaved by Dicer and is bounded to the RISC complex at rate K_D;
- distinct dynamics for the unfolded (inactive) and folded (active) forms of the reporter proteins.

Thus the resulting model looks like:

$$\frac{dx_1}{dt} = v_1\left(\alpha_1 + (1-\alpha_1)\frac{\left(\frac{\theta^{h_0}}{\theta^{h_0}+D^{h_0}}x_2\right)^{h_2}}{K_1^{h_2} + \left(\frac{\theta^{h_0}}{\theta^{h_0}+D^{h_0}}x_2\right)^{h_2}}\right) - d_1x_1 - \lambda\frac{{x_4}^{h_3}}{{K_3}^{h_3} + {x_4}^{h_3}}x_1, \tag{5.1}$$

$$\frac{dx_2}{dt} = v_2x_1 - d_2x_2, \tag{5.2}$$

$$\frac{dx_3}{dt} = v_1\left(\alpha_1 + (1-\alpha_1)\frac{\left(\frac{\theta^{h_0}}{\theta^{h_0}+D^{h_0}}x_2\right)^{h_2}}{K_1^{h_2} + \left(\frac{\theta^{h_0}}{\theta^{h_0}+D^{h_0}}x_2\right)^{h_2}}\right) - (d_3 + K_D)x_3, \tag{5.3}$$

$$\frac{dx_4}{dt} = K_Dx_3 - \delta x_4, \tag{5.4}$$

$$\frac{dx_5}{dt} = v_4x_1 - (K_{fg} + d_4)x_5, \tag{5.5}$$

$$\frac{dx_6}{dt} = K_{fg}x_5 - d_4x_6, \tag{5.6}$$

$$\frac{dx_7}{dt} = v_3x_3 - (K_{fr} + d_5)x_7, \tag{5.7}$$

$$\frac{dx_8}{dt} = K_{fr}x_7 - d_5x_8. \tag{5.8}$$

where

- x_1 is the tTA mRNA,
- x_2 is the tTA protein,
- x_3 is the miR223 mRNA,
- x_4 is the miR223 mRNA in the active form
- x_5 is the d2EYFP unfolded protein,
- x_6 is the d2EYFP folded protein,

- x_7 is the mcherry unfolded protein,
- x_8 is the mcherry folded protein.

Parameters estimation and Bifurcation Analysis

Table 5.1 reports parameter values for which *in silico* simulations give rise to undamped oscillations. Parameter values have been chosen based on: knowledge of the experimental system (i.e. the Hill coefficient describing the miR223 silencing strictly related to the number of tags); fitting of experimental results (Chapter 3); bifurcation analysis results (Fig.5.7, 5.8,5.9,5.10,5.11, 5.12).

Bifurcation analysis were carried out considering parameters for which certain ranges of values are physically and experimentally feasible, in order the system to behave with undamped oscillations.

In silico simulations are reported in Fig.5.13 and 5.14. For this parameters choice the model shows a period of 1000 minutes.

We also ran simulations using a similar mathematical model (described below) that uses a unique differential equation to describe the dynamics of the miR223. As shown in Fig.5.15, using the same set of parameters of the previous model, oscillations do not occur anymore, stressing the importance of a sufficient time delay to get the oscillatory behavior.

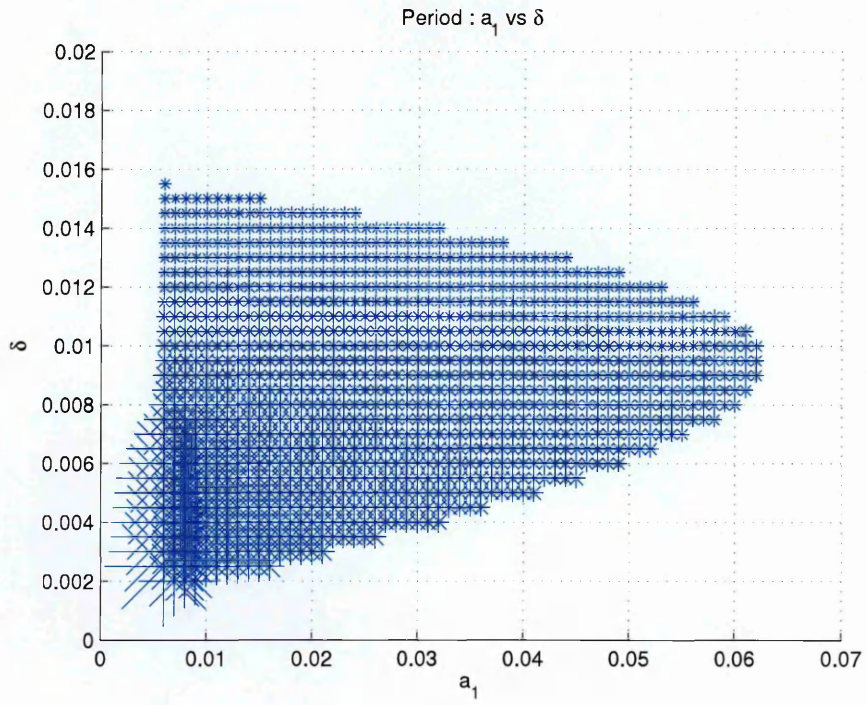

Figure 5.7: **Bifurcation diagram for basal activity of *CMV-TET* promoter (α_1) *versus* degradation rate of miR223 (δ).**
Big stars in the diagram correspond a longer period of the oscillations.

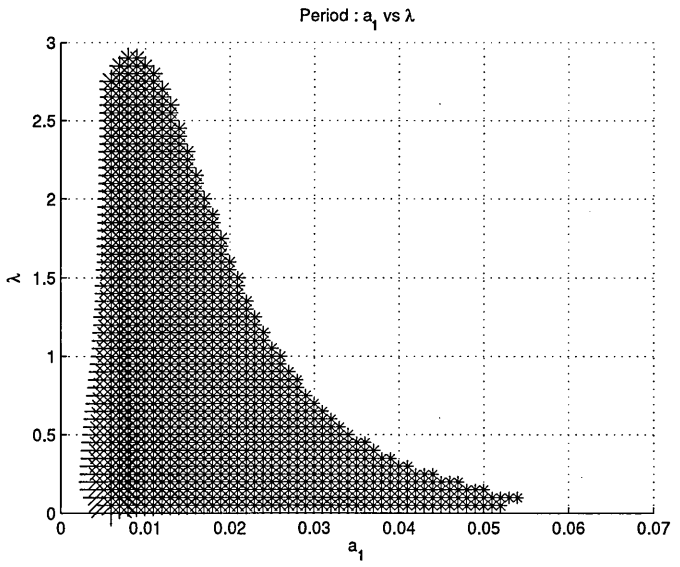

Figure 5.8: **Bifurcation diagram for basal activity of *CMV-TET* promoter (α_1) *versus* maximal rate of silencing of miR223 (λ)**.Big stars in the diagram correspond a longer period of the oscillations.

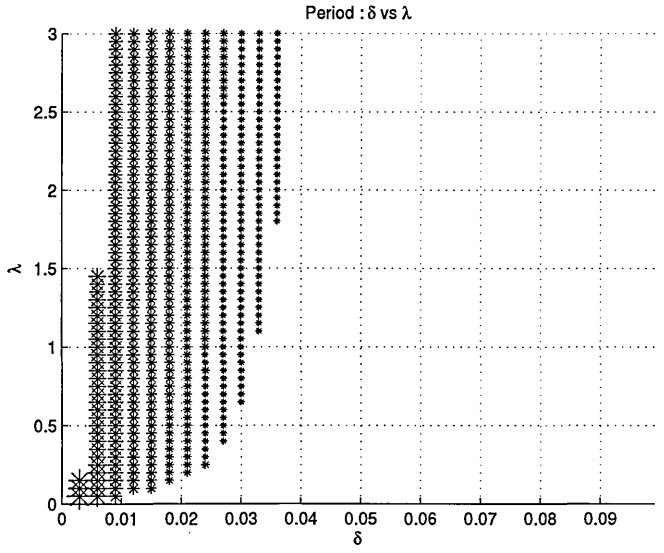

Figure 5.9: **Bifurcation diagram for degradation rate of miR223 (δ) *versus* maximal rate of silencing of miR223 (λ)**. Big stars in the diagram correspond a longer period of the oscillations.

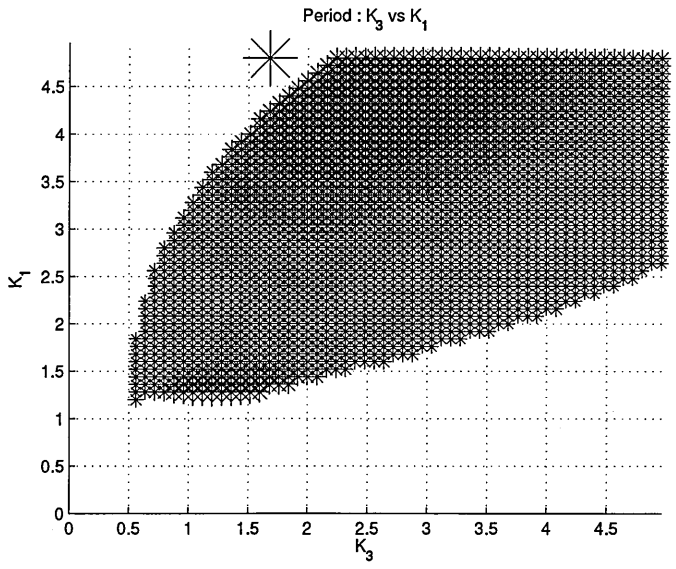

Figure 5.10: **Bifurcation diagram for hill constant for d2EYFP equation (K_3) *versus* hill costant for miR223 equation (K_1).** Big stars in the diagram correspond a longer period of the oscillations.

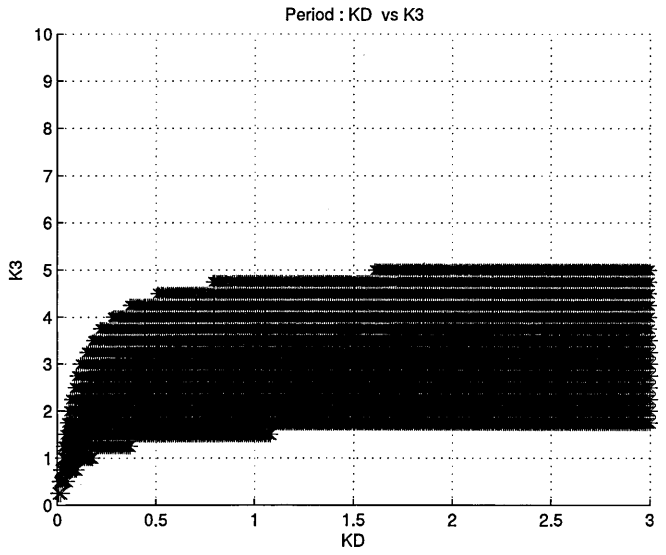

Figure 5.11: **Bifurcation diagram for hill costant for d2EYFP equation (K_3) *versus* folding rate for miR223 (K_D).** Big stars in the diagram correspond a longer period of the oscillations.

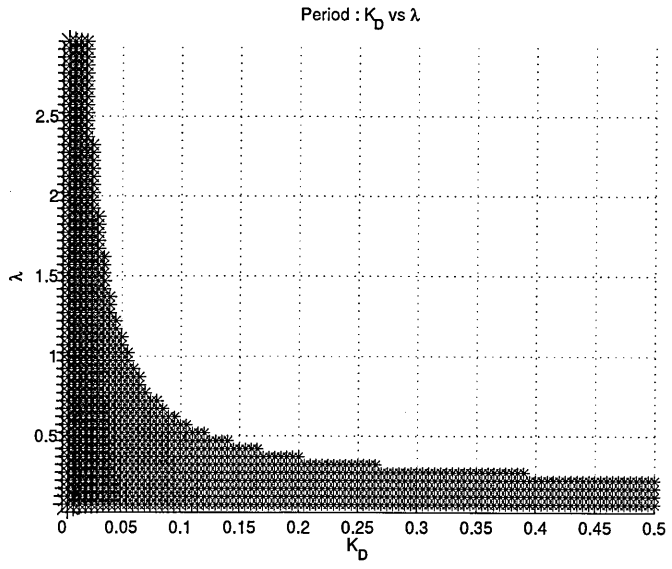

Figure 5.12: **Bifurcation diagram for folding rate for miR223 (K_D) *versus* maximal rate of silencing of miR223 (λ).** Big stars in the diagram correspond a longer period of the oscillations.

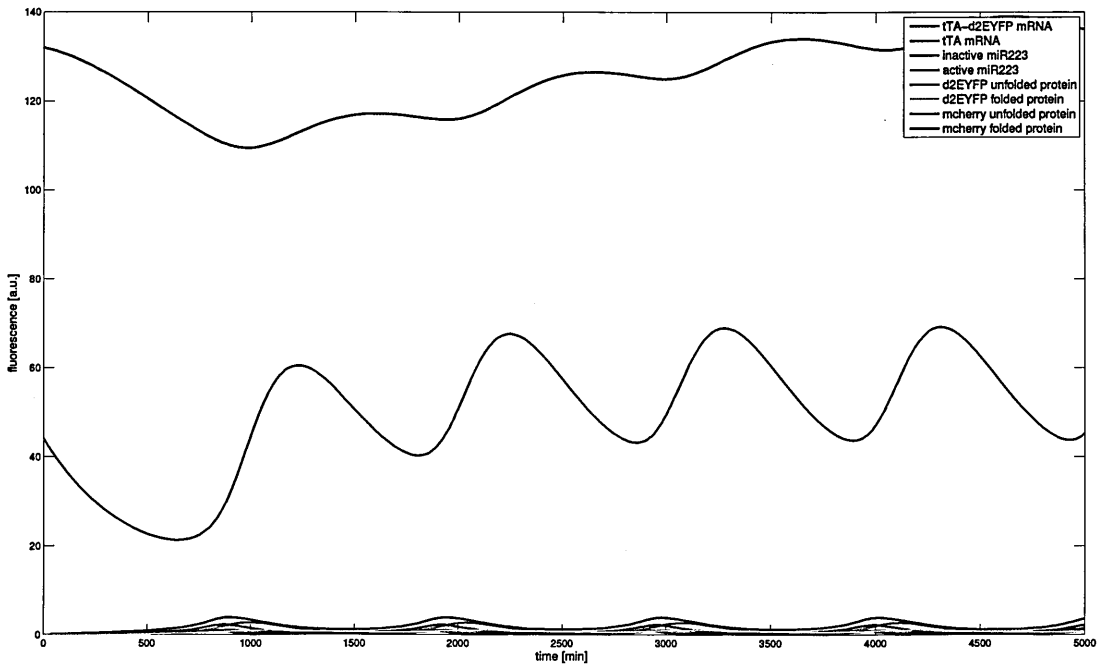

Figure 5.13: **Zoom of the folded and unfolded form of mCherry on the simulation of all equations for the second model with parameters of Table 5.1. Due to the stability of the protein, mCherry levels are up to 40 times more than d2EYFP protein levels.**

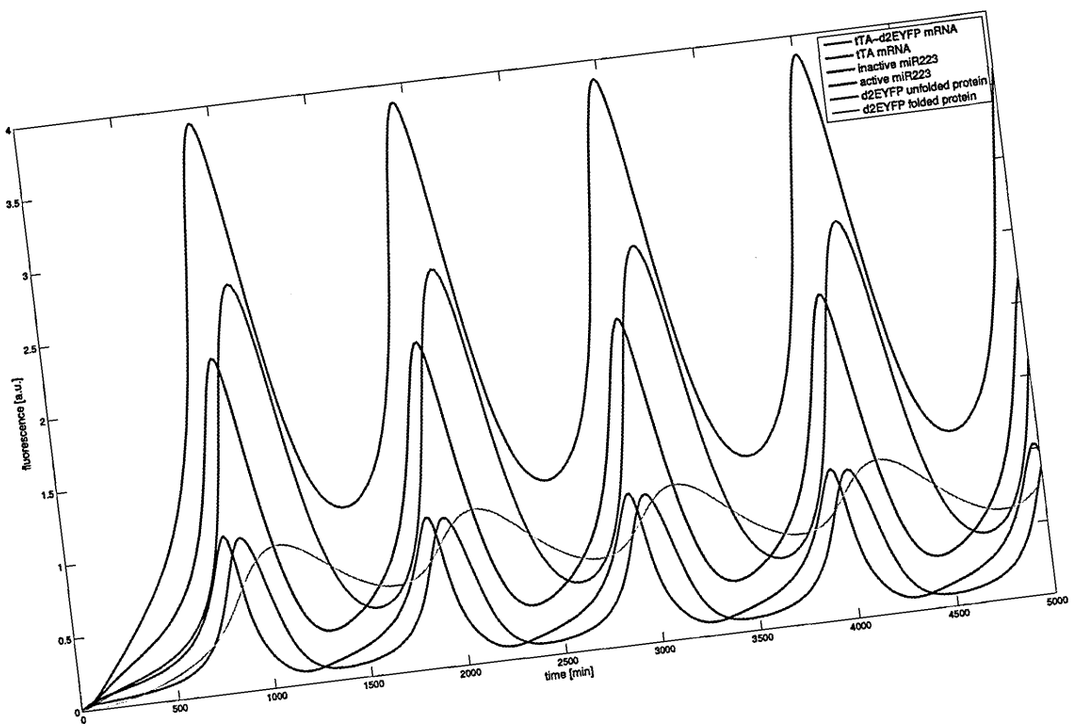

Figure 5.14: **Simulation for the second model with parameters of Table 5.1.**

Table 5.1: Parameter for the mathematical model; undamped oscillations.

Parameters	Definition	unit	Value
v_1	maximal transcription rate for CMV promoter	$[nMmin^{-1}]$	0.075432
v_2	translation rate for tTA protein	$[min^{-1}]$	0.027131449
v_3	maximal transcription rate for CMVTET promoter	$[nMmin^{-1}]$	0.075432026
v_4	translation rate for mcherry	$[min^{-1}]$	0.0271
d_1	degradation rate for tTA mRNA	$[min^{-1}]$	0.01012906
d_2	degradation rate for tTA protein	$[min^{-1}]$	0.010016646
d_3	degradation rate for miR223 mRNA	$[min^{-1}]$	0.0004814
d_4	degradation rate for mcherry protein	$[min^{-1}]$	0.003236
d_5	degradation rate for d2EYFP	$[min^{-1}]$	0.00048135
δ	degradation rate for miR223 mRNA	$[min^{-1}]$	0.007
α_1	basal activity for CMVTET promoter		0.015
λ	maximal rate of silencing	$[min^{-1}]$	0.073879
K_1	Hill constant for miR223 equation	$[nM]$	3
K_3	Hill constant for d2EYFP equation	$[nM]$	2
K_D	folding rate for miR223	$[nM]$	0.025
h_2	Hill constant for miR223 equation		2
h_3	Hill constant for d2EYFP equation		4

$$\frac{dx_1}{dt} = v_1\left(\alpha_1 + (1-\alpha_1)\frac{\left(\frac{\theta^{h_0}}{\theta^{h_0}+D^{h_0}}x_2\right)^{h_2}}{K_1^{h_2} + \left(\frac{\theta^{h_0}}{\theta^{h_0}+D^{h_0}}x_2\right)^{h_2}}\right) - d_1x_1 - \lambda\frac{{x_3}^{h_3}}{{K_3}^{h_3} + {x_3}^{h_3}}, \tag{5.9}$$

$$\frac{dx_2}{dt} = v_2x_1 - d_2x_2, \tag{5.10}$$

$$\frac{dx_3}{dt} = v_1\left(\alpha_1 + (1-\alpha_1)\frac{\left(\frac{\theta^{h_0}}{\theta^{h_0}+D^{h_0}}x_2\right)^{h_2}}{K_1^{h_2} + \left(\frac{\theta^{h_0}}{\theta^{h_0}+D^{h_0}}x_2\right)^{h_2}}\right) - d_3x_3, \tag{5.11}$$

$$\frac{dx_4}{dt} = v_4x_1 - (K_{fg} + d_4)x_4, \tag{5.12}$$

$$\frac{dx_5}{dt} = K_{fg}x_4 - d_4x_5, \tag{5.13}$$

$$\frac{dx_6}{dt} = v_3x_3 - (K_{fr} + d_5)x_6, \tag{5.14}$$

$$\frac{dx_7}{dt} = K_{fr}x_6 - d_5x_7. \tag{5.15}$$

where

- x_1 is the tTA mRNA,
- x_2 is the tTA protein,
- x_3 is the miR223 mRNA,
- x_4 is the d2EYFP unfolded protein,
- x_5 is the d2EYFP folded protein,
- x_6 is the mcherry unfolded protein,
- x_7 is the mcherry folded protein.

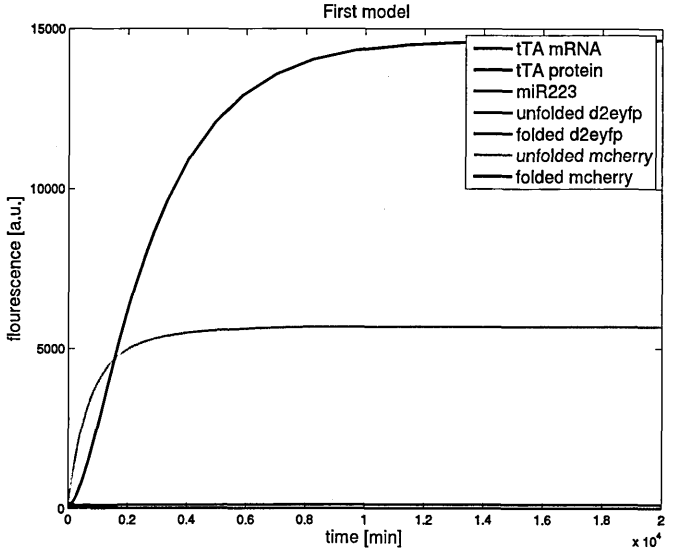

Figure 5.15: ***In silico* simulation of a mathematical model lacking of the equations that discriminate the inactive and active form of the miR223.** The network does no more behave as an oscillator.

5.4 Experimental investigation of dynamic properties of the network.

Since CHO cells are easy to track as single cells for image analysis, the first set of time lapse experiments were carried out with these cells.

To capture the dynamic properties of the circuit, and to distinguish oscillatory events due to cell cycle rather than the interaction of positive and negative feedback loops, we generated a negative control lacking the features typical of an oscillatory network. In particular we stably expressed via lentiviral infection the NFL in CHO NOPFL cells. Therefore the network lacks both of the autoregulatory loop because the tTA expression is controlled by the *CMV* constitutive promoter, and of the four miR223 tags (Fig.5.16) (CHO-nOsc). As described in Materials and Methods section the double infected cells were sorted by FACS and clonal population were generated by single cell expansion.

We then performed time-series experiments in which CHO-Osc e CHO-nOsc were imaged using time-lapse fluorescence microscopy every 15 minutes for up to four days, quantifying the average fluorescence intensity of the population of cells Experiments were carried out at a temperature of 32 °C in order to limit cell motility and reduce the risk associated to data loss occurring when cells exit the tracked field [46].

20% of the time lapse experiments showed oscillations with a period of 1000 minutes Fig.5.17; three out of thirteen clones of CHO cells exhibit oscil-

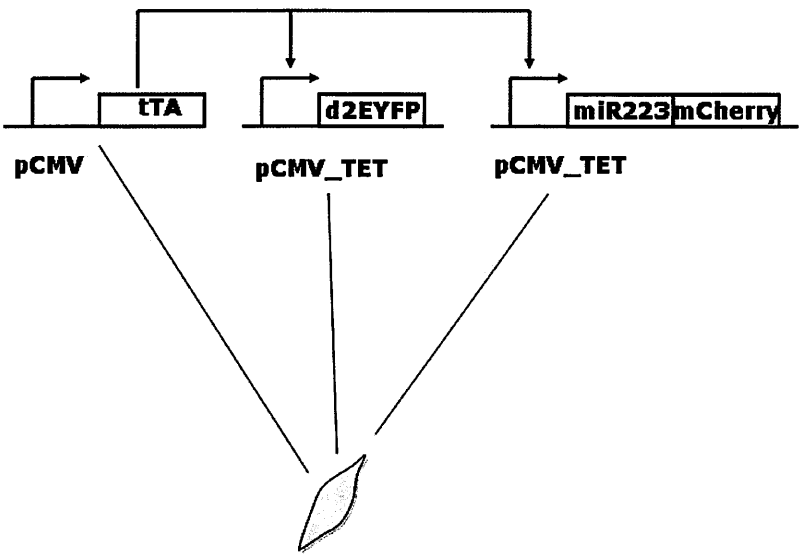

Figure 5.16: **Negative control of the oscillator in CHO cells.** CHO NO PFL cells were infected with the virus carrying the negative feedback loop. The resulting cell line was used as control in our experiments since cells lack of the tag for the miR223; therefore, variances in the level of fluorescence, are due only to the intrinsic cell variability.

latory behavior (RoG 7-1;RoG 7-2; RoG 2-8) compared to controls. Although the amplitude of oscillations is low, and the system is not as robust as expected, for the first time a complex synthetic network stably integrated in mammalian cells produces cyclic expression of a gene of interest.

We performed the same experiments using HEK GoR clones. Only the HEK GoR-B showed oscillations 6 out of 10 times. Differently from CHO cells, in this case we used as negative control HEK PFL cells. As is observed in fig 5.18 oscillations occur with a period of 1000 minutes.

In order to increase the robustness and the amplitude of the oscillations, we focused on the possibility to amplify the time delay between the activation of positive feedback loop and the response of the negative feedback loop. We ended up with a new design in which we added an intermediate step which comprises a new transcription factor. The design and implementation of the three steps genetic oscillator, will be discussed in Chapter 6.

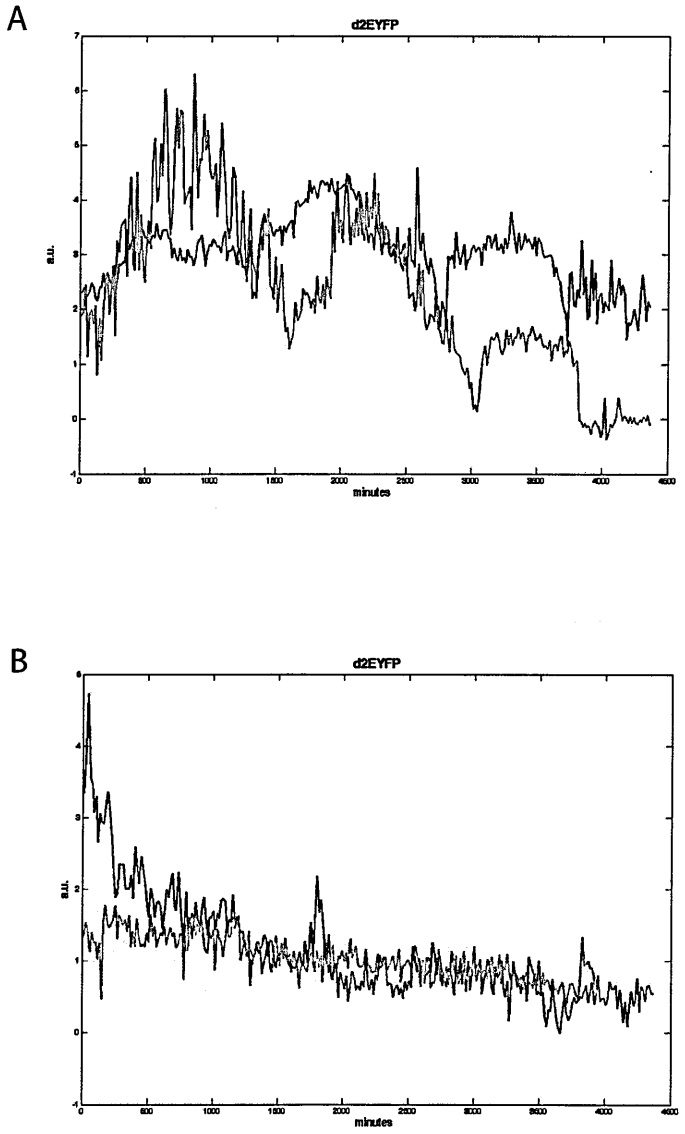

Figure 5.17: **CHO RoG cells show oscillatory behavior.** d2EYFP fluorescence was detected for up to 4000 minutes using a time lapse microscopy. Oscillations occur with a period of about 1000 minutes. (A) CHO-Osc. (B) CHO-nOsc.

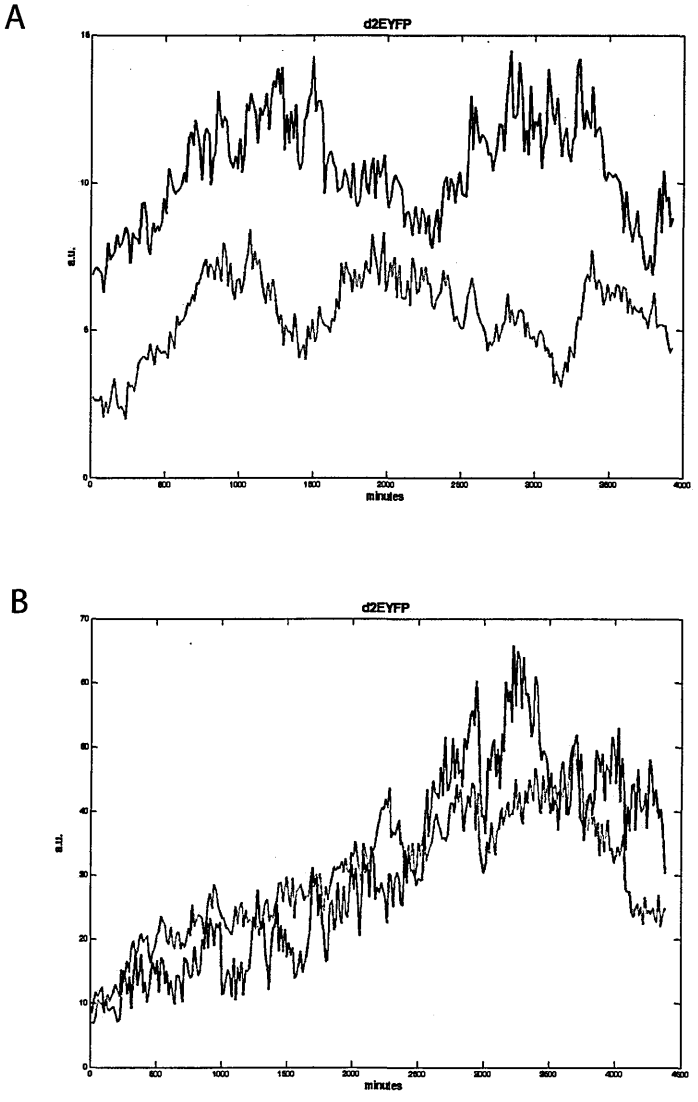

Figure 5.18: **HEK 293 GoR cells show oscillatory behavior.** d2EYFP fluorescence was detected for up to 4000 minutes using a time lapse microscopy. Oscillations occur with a period of about 1000 minutes. (A) HEK GoR cells. (B) HEK PFL cells.

Materials and Methods

Experimental procedure: construction of the Amplified Negative Feedback loop

As described in Chapter 3, to implement the negative feedback loop in a lentiviral vector, we used the ViraPower Promoterless Lentiviral Gateway Expression System (Invitrogen). The pMA-miR223-destRFP-WPRE vector containing the microRNA, and the destabilized form of RFP element was synthesised by GENEART together with the recombination sites. mCherry fluorescent protein coding sequence was amplified from the pmCherry (Clontech) by PCR with a forward primer containing a NheI recognition sequence (5'-catg GCTAGC atggtgagcaagggcgaggag-3'), and a reverse primer containing a NotI restriction site (5'-attc GCGGCCGC tta ctt gta cagctc gtc ca tgcc-3'). The PCR product and pMA-miR223-destRFP-WPRE were then digested with NheI-NotI restriction enzymes and the mCherry ligated in place of destRFP, generating a new vector termed pMA-miR223-mCherry. The pMA-miR223-mCherry was then linearised with the AseI restriction enzyme and recombined with the pDONR221 (Invitrogen) following the manufacturer instruction. In this way we generated pENTR-miR223-mCherry vector with specific recombination sites.

Finally we performed a recombination reaction between the pENTR-miR223-mCherry, pENTR5'-TOPO-*CMV-TET* (see Chapter 3 for details on construction) and the pLenti/R4R2/V5-DEST according to manufacturer

instructions.

As suggested by the manufacturer, the lentivirus was then produced in 293FTcells.

Experimental procedure: cell culture, lentiviral transduction, time lapse experiment

293FT, CHO, and CHO AA8 TET-OFF cells were maintained as described in Chapter 3. HEK 293 and HEK-PFL were maintained at 37°C in a 5% CO2-humidified incubator. HEK 293 cells were cultured in Dulbecco's modified Eagle's medium (DMEM, GIBCO BRL) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Invitrogen) and 1% antibiotic/antimycotic solution (GIBCO BRL). Clonal Populations of CHO-PFL, CHO-NoPFL and HEK-PFL cells were handled as described above, plus 6ug/mL of blasticidine to maintain the selection .

HEK 293 cells were infected with the virus containing the PFL as described for CHO cells in chapter 3, and then were sorted via BD FACSAria Cell Sorting System (Becton Dickinson). d2EYFP was excited at 488 nm, and emission was detected using a 525 nm bandpass filter. Then HEK-PFL cells were kept in blasticidine at a final concentration 8ug/mL. Serial dilutions of stably transduced cells (up to 0.05 cells/mL) were plated in 96-well microtitre plates, and dilutions containing only one cell per well were selected. Monoclonal colonies were cultured and amplified as described, to

obtain monoclonal populations of HEK.

To transduce cells with the virus produced, 500,000 of each clone of CHO-PFL, CHO-NoPFL, and HEK-PFL cell lines were plated and incubated overnight. On the day of transduction the medium was removed and 1mL of the virus was added to the cells together with polybrene (Invitrogen) to a final concentration of 6ug/mL. After an overnight incubation the medium containing the virus was removed and replaced with complete culture medium. HEK 293 cells were infected with the virus containing the PFL as described for CHO cells in chapter 3.

Cells were sorted for contemporary green and red fluorescence intensity using a BD FACSAria Cell Sorting System (Becton Dickinson).

Serial dilutions of stably transduced cells (up to 0.05 cells/mL) were plated in 96-well microtitre plates, and dilutions containing only one cell per well were selected. Monoclonal colonies were cultured and amplified as described, to obtain monoclonal populations of CHO-Oscillator cells (CHO-Osc), CHO-NoOscillator cells (CHO-NoOsc) and HEK-Oscillator (HEK-Osc).

For the time lapse experiment, 500 CHO-Osc, CHO-NoOsc, HEK-PFL and HEK-Osc cells, were plated in chamber slide (lab-Tek)and mantained in α-MEM/D-MEM (for CHO and HEK cells respectively), (Sigma)supplemented with 5% TET system approved FBS (Invitrogen), 4mM L-glutamine, and 1% antibiotic/antimycotic solution (GIBCO BRL).

Experimental procedure: DNA extraction, mRNA extraction RealTime PCR

1,000,000 CHO-Osc, CHO-NoOsc, HEK-PFL and HEK-Osc cells were plated in a 6-well multiwell plate to reach a confluence of 80% at the moment of the DNA/mRNA extraction. The day after cells were collected and resuspended in 200μL of PBS after centrifugation for five minutes at 300 x g . Then the DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen), the mRNA was extracted and retro-transcribed using the RNeasy mini kit and the Quantitec reverse transcrption kit (Qiagen), respectively. We compared the DNA levels of *tTA*, *d2EYFP* and *mCherry* all cell lines by RealTime PCR following DNA extraction, proving that the both cell populations carry a unique copy of the networks in their genome. We measured the mRNA levels to see the effect of the silencing due to the microRNA by Realtime PCR. Quantitative RealTime PCR reaction were set up in duplicates using the LightCycler 480 SYBR green master mix (Roche) and the amplification was performed using a LightCycler 480 RealTime PCR instrument(Roche). The PCR were carried out using the following primers: d2EYFP forward (5'-acgacggcactcaagacc-3'); d2EYFP reverse (5'-gtcctccttgaagtcgatgc-3'); PFL tTA forward (5'-aaagcagctgaagtgcgagag-3'); PFL tTA reverse (5'-gatggtgctgccgtagttgtt-3'); NO PFL tTA forward (5'-acagcgcattagagctgctt-3'); NO PFL tTA reverse (5'-acctagcttctgggcgagtt-3'); mCherry forward (5'-cactacgacgctgaggtcaag-3');

mCherry reverse (5'-gtagtcctcgttgtgggaggt-3').

Data analyses were performed using the LightCycler 480 Software(Roche). *GAPDH* DNA/mRNA levels were used to normalise the amount of DNA/mRNA and ΔCts were calculated as the difference between the average *GAPDH* Ct and the average *tTA*, *d2EYFP*, and *(*mCherry).

Image acquisition and processing

Please refer to Chapter 3

Model simulations

Numerical simulations and bifurcation diagrams were run using Matlab 2010b (Mathworks Inc.). We used *ode23s* solver (a detailed discussion of the numerical methods used by *ode23* can be found in [14]).

Chapter 6

Synthesis and Analysis of a delayed amplified negative feedback loop oscillator in mammalian cells

In this chapter I will describe an alternative topology of the network that we are currently developing in order the oscillation to be ensured. For this design we made new experimental analysis, mathematical model and simulations, which will be detailed in the following sections.

6.1 Design of a three stage genetic oscillator.

It has been shown that a time delay in the dynamics of negative feedback loops is required to get sustained and stable oscillations .

We designed a three stage genetic oscillator which allows us to increase the time elapsing between the auto-activation of the positive feedback loop and the repression of the microRNA on the tTA.

The new topology is shown in Figure 6.1. The PFL is conserved, but the tTA drives the expression of a gene expression cassette consisting of:

- destabilized artificial activator (dGAL4-VP16) whose expression has been optimized for mammalian cells [84].
- IRES sequence
- destabilized cyan reporter (dCyan).

The NFL was modified so that the miR223 can be regulated by a GAL4 responsive promoter (UAS). This promoter is composed by a minimal cytomegalovirus promoter (CMVmini) fused with four tandem repeats of the galactose upstream activating sequence [84]. Once the dGAL4-VP16 is produced, it binds the responsive promoter (UAS) driving the expression of miR223 , thus giving rise to a new Negative Feedback Loop (UAS-NFL), that will silence the tTA-d2EYFP mRNA expression.

In order to construct the network, we first constructed vectors carrying the three constructs and then we used the Gateway Technology (Invitrogen)

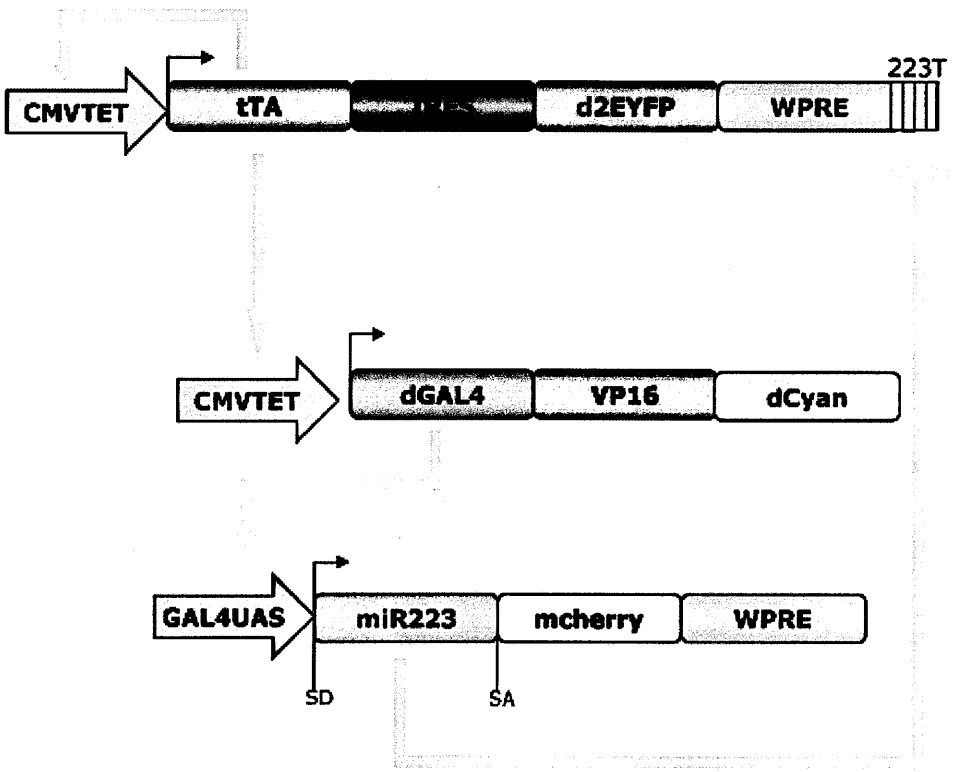

Figure 6.1: **Three stage genetic clock.**

to produce lentiviral vectors and finally the viruses (Materials and Methods).

We infected CHO and HEK 293 cells with the virus carrying the dGAL4-VP16 and we selected for blasticidine selection positive cells. Unfortunately only CHO cells survived the selection, therefore all the experiments were carried out on this cell line.

As shown in Fig.6.2, we extracted the genomic DNA to confirm the integration; to confirm that the promoter *CMV-TET* was activated by tTA, we transfected cells with a plasmid expressing constitutively the transcriptional factor, and compared the mRNA levels of *dGAL4-VP16* in absence and presence of tTA by Real Time PCR (Fig.6.2).

We then co-transfected CHO-GAL4 cells with the lentiviral vector carrying the PFL and the UAS-NFL to test whether the regulation of the complete system worked as expected. Since the activation of the system is induced by tTA, in the negative control, we transfected a plasmid carrying another gene not involved at all in our system (namely *grn*). Results are shown in Fig.6.3.

We then generated four clonal populations of CHO cells and performed the same transfection described above, by evaluating the expression of (dGal4-VP16) upon the induction by tTA activator (figure 6.4). The clonal population were called CHO-G1, CHO-G2, CHO-G3, CHO-G4.

In order to implement the entire genetic oscillator in CHO cells, we infected CHO-G clones with the PFL virus and we sorted positive cells via FACS analysis for the green fluorescence. Also in this case we adopted a

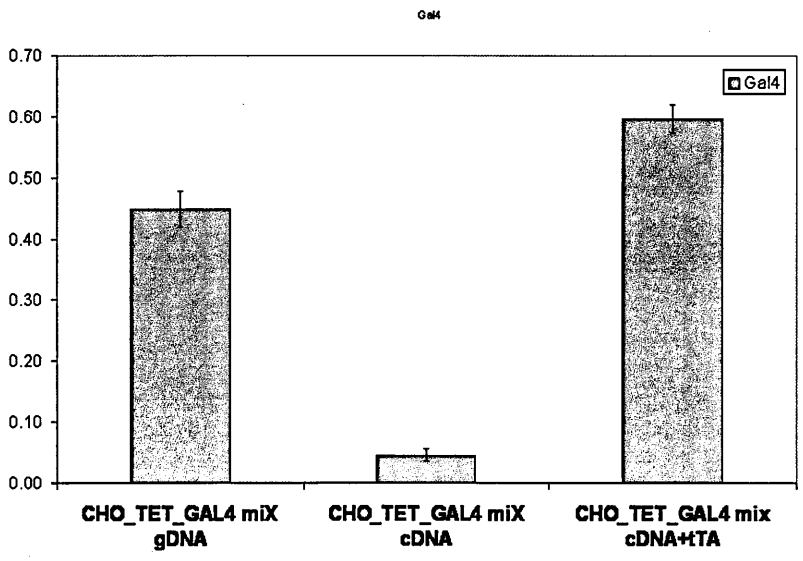

Figure 6.2: **RealTime PCR of genomic DNA (gDNA) and cDNA of CHO cells infected with the lentivirus carrying the dGAL4-VP16 under the control of *CMV-TET* promoter.** Third column: transfection of infected CHO cells with pTET-OFF vector expressing the tTA to test the activation of the system.

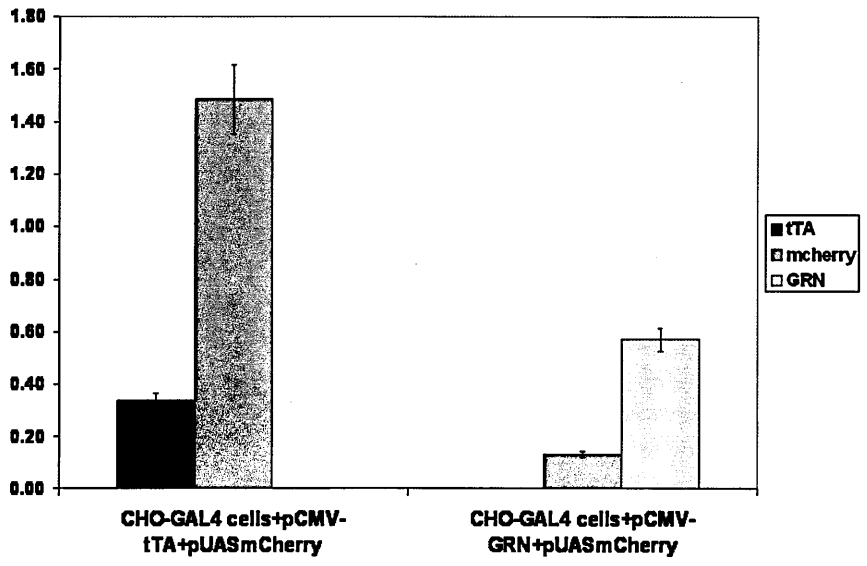

Figure 6.3: **Co-trasfections of CHO-GAL4 cells with a pTET-OFF vector expressing the tTA transactivator, and a vector where the miR223 is driven by the *UAS* promoter.** Upon the production of GAL4 induced by the tTA, also the mRNA levels of *mCherry* increase, compared to the negative control where we transfected a vector of comparable size, but lacking of the activator of GAL4 expression

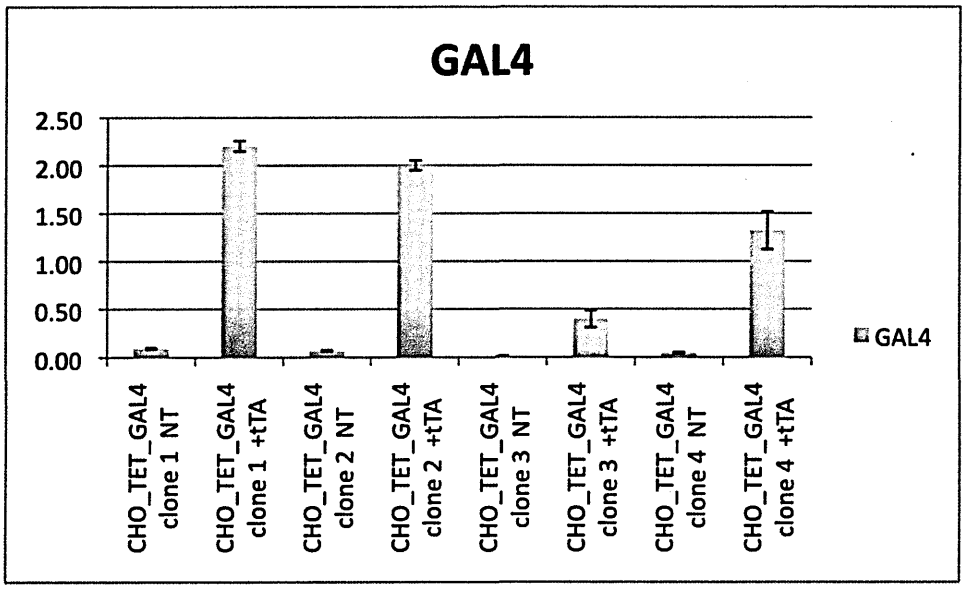

Figure 6.4: **RealTime PCR of clonal population of CHO cells infected with the lentivirus carrying the dGAL4-VP16 under the control of *CMV-TET* promoter.** Side by side are reported *GAL4-VP16* expression values in absence and presence of the activator tTA.

simple nomenclature; for example the clonal population 2 of CHO cells integrated with GAL4 infected with the PFL is called CHO-G2-P. Moreover from FACS analysis (data not shown) we recognized two different populations of CHO cells infected with the PFL, based on higher, or lower d2EYFP levels of expression. We isolated both populations and we checked from RealTime PCR the different mRNA levels of *tTA*, and *d2EYFP*, confirming a different expression level(Fig.6.5). These cell lines were termed CHO-G-P(+), and CHO-G-P(++). Of note, at DNA levels (Fig. 6.6), *tTA*, and *d2eyfp* are comparable in CHO-G-P(+), and CHO-G-P(++) cells; different expression levels could be explained by the random points of integration which are typical of lentiviral infections.

We are currently isolating clonal populations of CHO-G-P cells. Next, we will infect clones of CHO-G-P cells with the UAS-NFL virus to generate a stable integrated system that will be tested via time lapse microscopy as described in Chapter 5.

As described in the next section, also for this network a mathematical model, based on ordinary differential equations, has been developed.

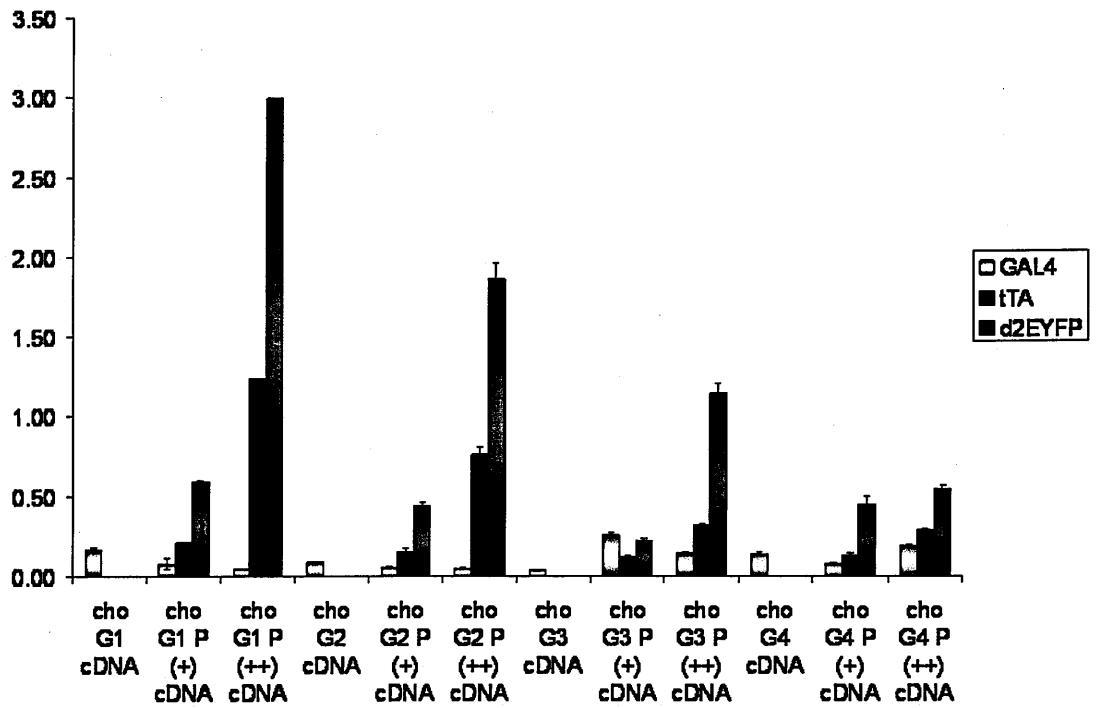

Figure 6.5: **RealTime PCR of cDNA of clonal population of CHO cells infected with the PFL.**

DNA.jpg

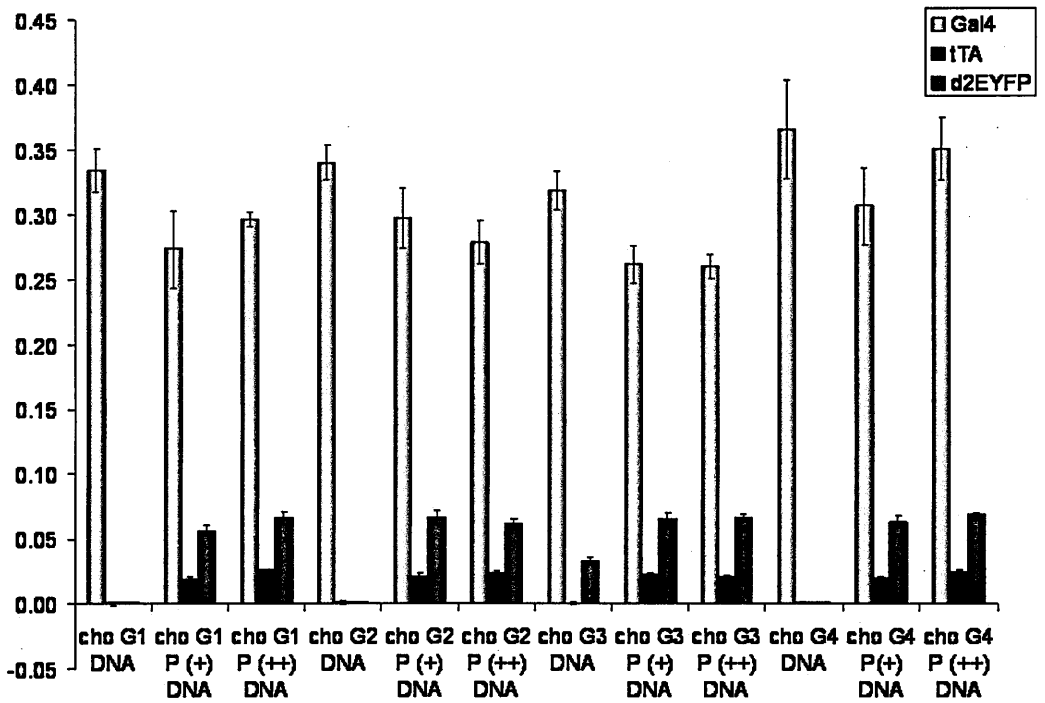

Figure 6.6: **RealTime PCR of genomic DNA of clonal population of CHO cells infected with the PFL.**

6.2 Mathematical modeling of three step oscillator.

In this section we propose a mathematical model close to the previous one used to describe the amplified feedback loop, but that takes into account the slowing down of the repression process due to the intermediate step. Since the delay obtained seems to be sufficient enough to get the oscillatory behavior, the inactive and active forms of miR223 are no longer considered separately, but we just describe the active form of the microRNA.

$$\frac{dx_1}{dt} = v_1\left(\alpha_1 + (1-\alpha_1)\frac{\left(\frac{\theta^{h_0}}{\theta^{h_0}+D^{h_0}}x_2\right)^{h_2}}{K_1^{h_2} + \left(\frac{\theta^{h_0}}{\theta^{h_0}+D^{h_0}}x_2\right)^{h_2}}\right) - d_1x_1 - \lambda\frac{{x_5}^{h_3}}{{K_3}^{h_3} + {x_5}^{h_3}}x_1, \tag{6.1}$$

$$\frac{dx_2}{dt} = v_2x_1 - d_2x_2, \tag{6.2}$$

$$\frac{dx_3}{dt} = v_1\left(\alpha_1 + (1-\alpha_1)\frac{\left(\frac{\theta^{h_0}}{\theta^{h_0}+D^{h_0}}x_2\right)^{h_2}}{K_1^{h_2} + \left(\frac{\theta^{h_0}}{\theta^{h_0}+D^{h_0}}x_2\right)^{h_2}}\right) - d_Gx_3, \tag{6.3}$$

$$\frac{dx_4}{dt} = v_{PG}x_3 - d_{PG}x_4, \tag{6.4}$$

$$\frac{dx_5}{dt} = v_{UAS}\left(\alpha_{UAS} + (1-\alpha_{UAS})\frac{\left(\frac{\theta^{h_0}}{\theta^{h_0}+D^{h_0}}x_4\right)^{h_{UAS}}}{K_{UAS}^{h_{UAS}} + \left(\frac{\theta^{h_0}}{\theta^{h_0}+D^{h_0}}x_4\right)^{h_{UAS}}}\right) - d_3x_5, \tag{6.5}$$

$$\frac{dx_6}{dt} = v_4x_1 - (K_{fg} + d_4)x_6, \tag{6.6}$$

$$\frac{dx_7}{dt} = K_{fg}x_6 - d_4x_7, \tag{6.7}$$

$$\frac{dx_8}{dt} = v_3x_5 - (K_{fr} + d_5)x_8, \tag{6.8}$$

$$\frac{dx_9}{dt} = K_{fr}x_8 - d_5x_9. \tag{6.9}$$

$$\frac{dx_{10}}{dt} = v_5x_3 - (K_{cr} + d_6)x_{10}, \tag{6.10}$$

$$\frac{dx_{11}}{dt} = K_{cr}x_{10} - d_6x_{11}. \tag{6.11}$$

where

- x_1 is the tTA mRNA,
- x_2 is the tTA protein,

- x_3 is the dGal4 mRNA,
- x_4 is the dGal4 protein,
- x_5 is the miR223,
- x_6 is the d2EYFP unfolded protein,
- x_7 is the d2EYFP folded protein,
- x_8 is the mcherry unfolded protein,
- x_9 is the mcherry folded protein,
- x_10 is the cyan unfolded protein,
- x_11 is the cyan folded protein,

All the parameters are the same as for previous model. Some of the new parameters have been chosen from literature, in particular the half-life of the protein dGAL4 is known to be of 3.76 hours [84], while the half-life of mir223 is supposed to be of 25 hours [9] (table.6.1).

From preliminary simulation (Fig.6.7) and qualitative analysis, a long oscillation period and a different amplitude can be predicted by this model; some diagrams of period are reported as function of the new parameters like the degradation of *gal4* mRNA (d3) (Fig.6.8) or the Hill constant for the *GAL4-UAS* promoter (K-UAS) (Fig.6.9).

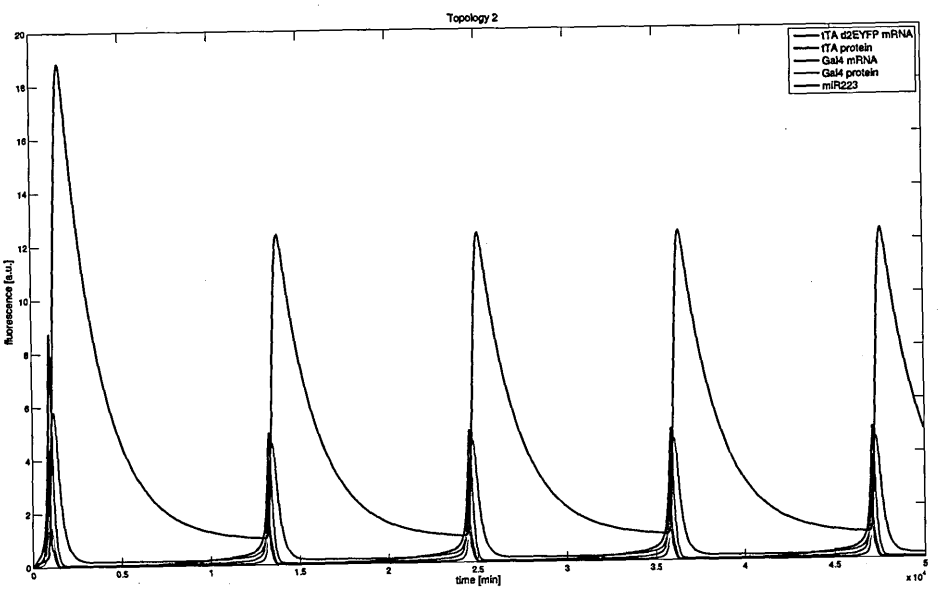

Figure 6.7: **Simulation for the three stage oscillator.**

Table 6.1: Parameter for the first model; undamped oscillations.

Parameters	Definition	unit	Value
v_1	maximal transcription rate for CMV promoter	$[nMmin^{-1}]$	0.075432
v_2	translation rate for tTA protein	$[min^{-1}]$	0.027131449
v_3	maximal transcription rate for CMVTET promoter	$[nMmin^{-1}]$	0.075432026
v_4	translation rate for mcherry	$[min^{-1}]$	0.0271
d_1	degradation rate for tTA mRNA	$[min^{-1}]$	0.01012906
d_2	degradation rate for tTA protein	$[min^{-1}]$	0.010016646
d_3	degradation rate for miR223 mRNA	$[min^{-1}]$	0.0004814
d_4	degradation rate for mcherry protein	$[min^{-1}]$	0.003236
d_5	degradation rate for d2EYFP	$[min^{-1}]$	0.00048135
α_1	basal activity for CMVTET promoter		0.015
λ	maximal rate of silencing	$[min^{-1}]$	0.073879
K_1	Hill constant for miR223 equation	$[nM]$	3
K_3	Hill constant for d2EYFP equation	$[nM]$	2
h_2	Hill coefficient for miR223 equation		2
h_3	Hill coefficient for d2EYFP equation		4
α_{UAS}	Basal activity for Gal4UAS promoter		0.008
K_{UAS}	Hill constant for Gal4UAS promoter		3
h_{UAS}	Hill coefficient for Gal4UAS promoter		4
d_{PG}	degradation rate for dGal4 protein		0.003086
d_G	degradation rate for dGal4 mRNA		0.0458
v_{UAS}	maximal transcription rate for Gal4UAS promoter		0.055
v_{PG}	translation rate for dGal4 protein		0.02

Materials and Methods

Experimental procedure: construction of the three step oscillator.

To implement this new topology of the network we took advantage of the ViraPower Promoterless Lentiviral Gateway Expression System (Invitrogen) as detailed in Chapter 3. The pDonR-dGAL4-VP16-dCyan, and the pMA-cPPT-UAS vectors containing specific recombination sites were synthesised by GENEART.

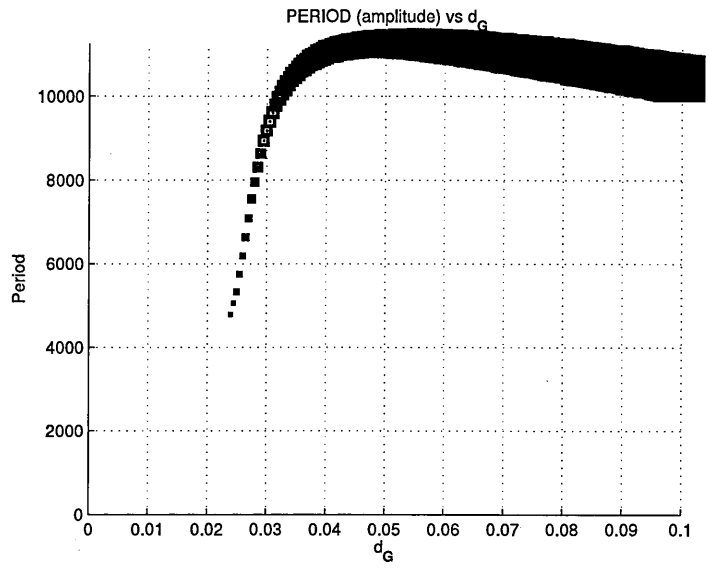

Figure 6.8: **Quantitative analysis of the period of the oscillations changing the degradation rate for *dGal4* mRNA (d_G).**

The *cPPt-UAS* cassette was amplified from pMA-cPPT-UAS by PCR. The PCR was performed with the Taq polymerase provided by Invitrogen that adds a single deoxyadenosine (A) to the 3' ends of PCR products. This allows PCR inserts to ligate efficiently with the pENTR5'-TOPO vector which is supplied linearised with single 3'-deoxythymidine (T) overhangs, obtaining the pENTR5'-TOPO-*UAS* with specific recombination sites.

Finally we performed a recombination reaction between the pDonR-dGAL4-VP16-dCyan, pENTR5'-TOPO-*CMV-TET* and the pLenti/R4R2/V5-DEST according to manufacturer instructions, to generate the *CMV-TET*-GAL4-VP16 lentiviral vector. We also performed the same recombination reac-

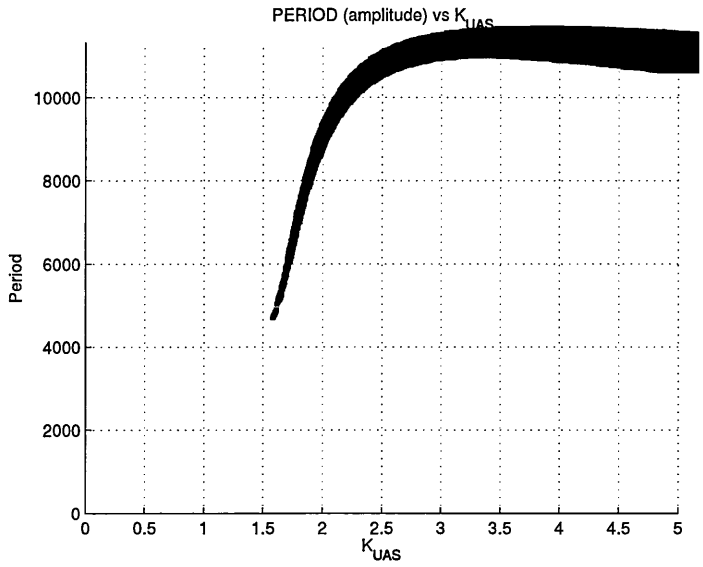

Figure 6.9: **Quantitative analysis of the period of the oscillations changing the hill costant for *Gal4-UAS* promoter (K_{UAS}).**

tion but using the pENTR5'-TOPO-*UAS*, pENTR-miR223-mCherry and the pLenti/R4R2/V5-DEST.

As suggested by the manufacturer, the lentivirus was then produced in 293FTcells.

Experimental procedure: cell culture, lentiviral transduction, transfections.

293FT, and CHO cells were maintained as described in Chapter 3. To transduce cells with the virus produced, 500,000 CHO cells were plated and incu-

bated overnight. On the day of transduction the medium was removed and 1mL of the virus carrying the *CMV-TET*-GAL4-VP16 was added to the cells together with polybrene (Invitrogen) to a final concentration of 6ug/mL. After an overnight incubation the medium containing the virus was removed and replaced with complete culture medium containing Blasticidin (Sigma) to a final concentration of 5μg/mL to select for stably transduced cells. Serial dilutions of stably transduced cells (up to 0.05 cells/mL) were plated in 96-well microtitre plates, and dilutions containing only one cell per well were selected. Monoclonal colonies were cultured and amplified as described, to obtain monoclonal populations. 500,000 CHO-G cells were transduced as described above with the PFL virus, and sorted according to green fluorescence using a BD FACSAria Cell Sorting System (Becton Dickinson). CHO-G cells were seeded at a density of 300.000 per well in a 6 wells multi-well and transfected 1 day after seeding using Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions

Experimental procedure: DNA extraction, mRNA extraction RealTime PCR

1,000,000 CHO-G, and CHO-G-P cells were plated in a 6-well multiwell plate to reach a confluence of 80% at the moment of the DNA/mRNA extraction. The day after cells were collected and resuspended in 200μL of PBS after centrifugation for five minutes at 300 x g . Then the DNA was extracted using the

DNeasy Blood and Tissue kit (Qiagen), the mRNA was extracted and retro-transcribed using the RNeasy mini kit and the Quantitec reverse transcrp-tion kit (Qiagen), respectively. The PCR were carried out using the following primers: d2EYFP forward (5'-acgacggcactcaagacc-3'); d2EYFP reverse (5'-gtcctccttgaagtcgatgc-3'); PFL tTA forward (5'-aaagcagctgaagtgcgagag-3'); PFL tTA reverse (5'-gatggtgctgccgtagttgtt-3'); NO PFL tTA forward (5'-acagcgcattagagctgctt-3'); NO PFL tTA reverse (5'-acctagcttctgggcgagtt-3'); GAL4 forward(5'-GGACGAGCTCCACTTAGACG-3')
GAL4 reverse (5'-GTCCCCCAACATGTCCAGAT3').

Data analyses were performed using the LightCycler 480 Software(Roche). *GAPDH* DNA/mRNA levels were used to normalise the amount of DNA/m-RNA and ΔCts were calculated as the difference between the average *GAPDH* Ct and the average *tTA*, *d2EYFP*, and *(*GAL4).

Model simulations

Numerical simulations and qualitative analysis diagrams were run using Mat-lab 2010b (Mathworks Inc.). We used *ode23s* solver (a detailed discussion of the numerical methods used by *ode23* can be found in [14]).

Chapter 7

Conclusion

The main challenge in synthetic biology is to develop a deeper understanding of the biological design principles using a bottom-up approach by constructing synthetic networks and studying their behavior in cells. Researchers have begun to design circuits using design principles derived from endogenous cellular processes. This thesis focused on the investigation of the biological mechanisms involved in cyclic gene expression (e.g. circadian clocks) by synthesizing a genetic oscillator and integrating it in mammalian cells. The oscillator creates a temporal structure that serves to anticipate what is needed by the cell and the organism, and when. Even though the molecular circadian network is described as interlocked loops formed by transcription factors, it has been recently shown that also non-trascriptional events can sustain circadian rhythms. For example the peroxiredoxins, a family of high conserved antioxidant proteins, under constant conditions (light, temperature),

are subjected to 24-hour redox cycles in human red blood cells.

We built a synthetic network in which classical motives, positive and negative feedback loop, are mutually connected to sustain periodic expression of a gene. We first investigated the cellular behavior upon the integration of a positive feedback loop (PFL). We have demonstrated that a transcriptional positive feedback loop can slow down the "switch off" times, as compared to an equivalent network without auto−regulation. The reason for a cell to "choose" an auto-regulation control strategy for transcriptional regulation, could be due to the intrinsic robustness of this approach to transient activation of the network. For example, in a signalling pathway, a ligand (equivalent to Doxycycline in our PFL) could cause a transcription factor to stop transcribing itself, as well as, a set of target genes, to initiate a specific response. However, in order for the pathway not to respond to a transient concentration of the ligand, the PFL strategy has to be chosen, otherwise the response would start immediately. Moreover, the response time of the PFL network can be modulated by the ligand concentration, if this is really high, the system will switch off as quickly as possible alternatively the ligand can be present at low, or medium, concentration, but it should persist for a long time, in order for the pathway to respond. This kind of behaviour has been recently described as "persistence detection" in cellular signal processing to indicate the ability of the genetic circuit to distinguish between transient and persistent signals.

Interestingly, it has been shown in E. coli, that the transcriptional nega-

tive feedback loop (NFL) has an opposite effect, that is, it can significantly speed up the rise-times of transcription, but has very little effect on the switch-off times. The duality between positive and negative feedback has been predicted in a biological setting, and it is a well established concept in "control engineering", a branch of engineering which deals with the design of automated mechanisms to control a variable of interest (the altitude of an airplane, or more simply, the temperature of a room via thermostat). Specifically, the negative feedback loop is a classic control engineering approach to speed up the response times of a sytem, thus quickly achieving a desired value of a variable of interest. Positive feedback loops, instead, can slow down the response of the system to external input, and are used by control engineers to build "memory" elements, also known as switches, which are able to be in one of two steady-states (OFF or ON), and which are robust against unwanted transient perturbations that may inadvertently switch off (or on) the system. We indeed verified that the PFL can exhibit bistability for zero or low concentrations of Doxycyline A bistable genetic network will cause a population of cells to divide in two sub-populations, each in one of the two possible states (OFF or ON). In our mammalian PFL, this behaviour was not detected experimentally. This can be easily explained by observing that the PFL model is bistable but the basin of attraction of the OFF equilibrium point is much smaller as compared to that of the ON state, when no Doxycycline is present. Therefore, just few cells will be in the OFF state and these will not be enough to be significantly detected experimentally. We pre-

dict however that for intermediate concentration of Doxycycline (100ng/mL) the basin of attraction will be comparable and bistability should be detected experimentally.

Moreover we showed that mathematics and biology are a 'winning combination' ; for example, from mathematical simulations, we found that to fit the experimental results, the half–life of d2EYFP should have been longer than the one previously estimated in literature. In fact at our experimental conditions (32 °C) we estimated this half–life to be around 4 hours instead of 2 hours.

In (Chapter4) we were able to model mathematically the effect of siRNA on gene expression. We evaluated the performance of 4 different models to fit the experiments we carried out in our lab. Then we integrated this mathematical approach to model the effect of microRNA in our system as described in (Chapter5).

Since obtaining sustained oscillations is no trivial, mainly because a consistent time delay is needed between the activation of positive and negative feedback loop, we are currently investigating an improved version of the synthetic network which is based on three stages, conserving the basic design of a auto–regulating loop, and of a negative one (Chapter6).

Compared to previous attempts [81], with this work we addressed the issue of the stably integration of mammalian cells, a critical point to study the long-term effect of the circuit. The only investigation on the effect of such

a synthetic network in a mammalian system is based on transient transfection of three different plasmids, where only the right combination of each DNA vector will give rise to oscillations. Furthermore, we realized a system as close as possible to the biological pathways within the cells, using as negative repressor a natural microRNA that enters in the endogenous interference cascade. Such a synthetic network will be useful in studies requiring the periodic perturbation of a target genes to evaluate the related-pathway alterations. In a future perspective , the improvement of the system to fine tune the period of the oscillations, will allow also to design therapeuthic approaches for diseases in which the mutated gene codifies for a protein whose expression should occur periodically (i.e. insulin in diabetes).

Oscillations have been observed in major signaling pathways; p53 protein levels have been shown to be crucial in controlling sensitivity and tolerance to DNA damage responses; hes1 expression in neural progenitor cells are essential for patterning of the vertebrate embryo during development [73, 33, 57]. One hypothesis is that one of the roles of cell signaling is to allow 'mutual synchronization' and 'entrainment' of these oscillations across a tissue for robust coordination of biological processes at the tissue-level. The next goal will be to use the cell autonomous oscillator, which we have contructed, to test the role of entrainment and mutual synchronization in cell-cell signalling, and to characterize them mathematically via dynamical and numerical simulations.

Appendix A

Vector Maps

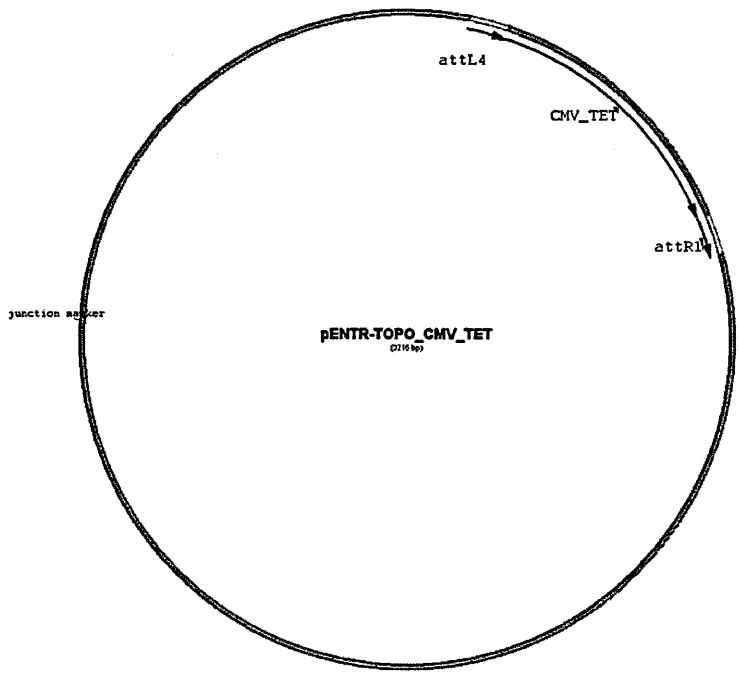

Figure A.1: **pENTR-CMV-TET vector.** Flanking are the attL4 and attR1 recombination sites.

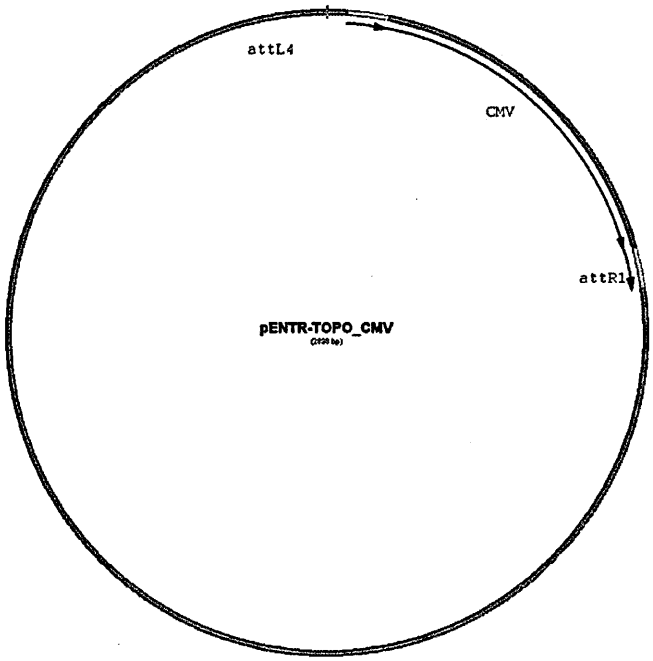

Figure A.2: **pENTR-CMV vector.** Flanking are the attL4 and attR1 recombination sites.

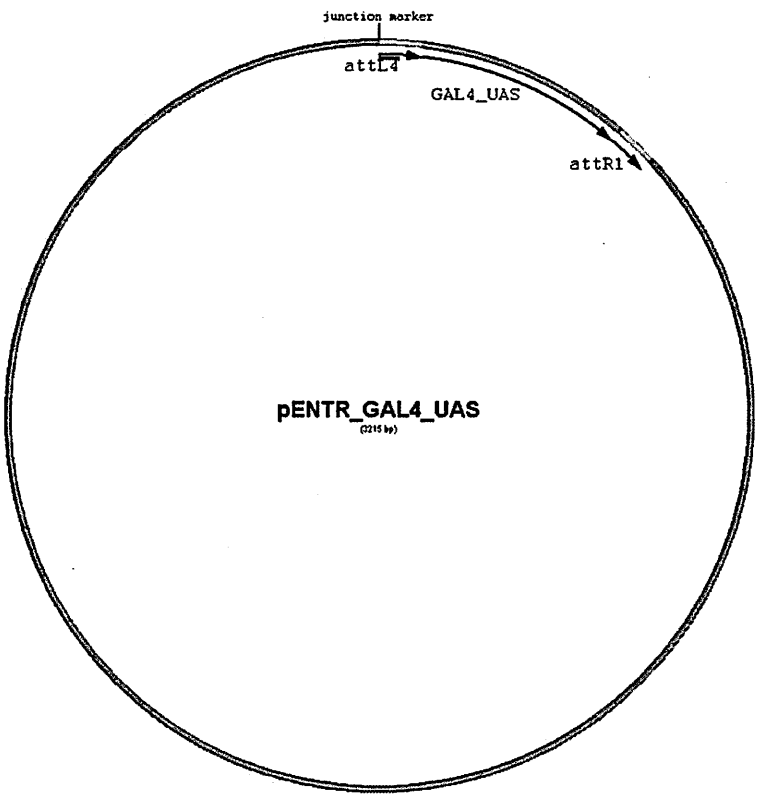

Figure A.3: **pENTR-UAS vector.** Flanking are the attL4 and attR1 recombination sites.

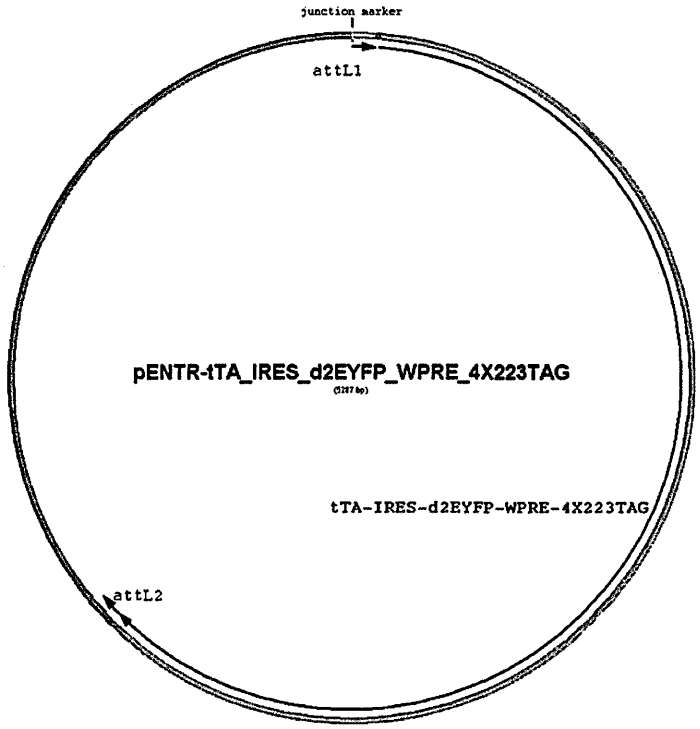

Figure A.4: **pENTR-PFL vector.** Flanking the tTA-IRES-d2EYFP-WPRE-223TAG are the attL1 and attL2 recombination sites.

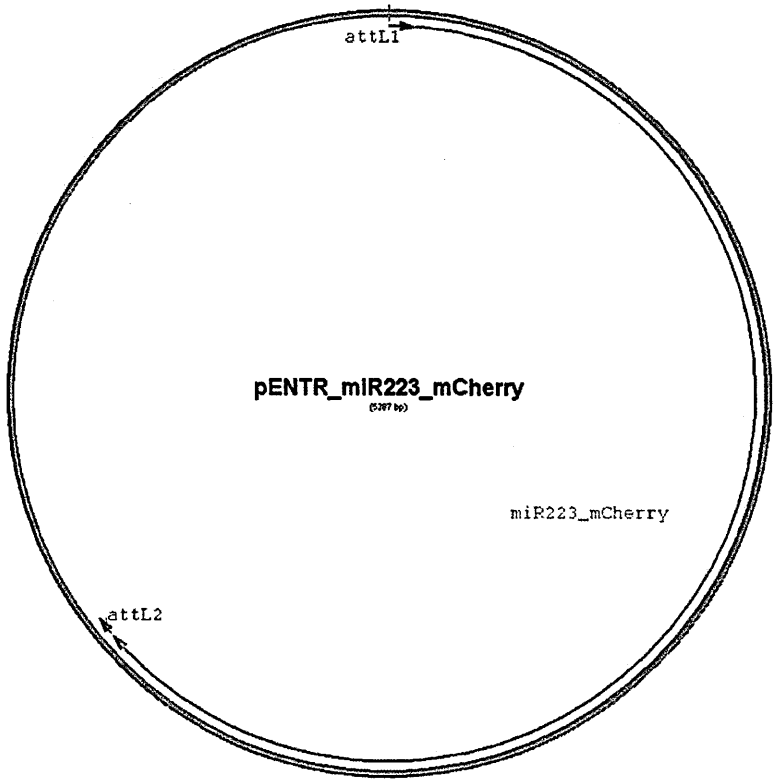

Figure A.5: **pENTR-NFL vector.** Flanking the miR223-mCherry are the attL1 and attL2 recombination sites.

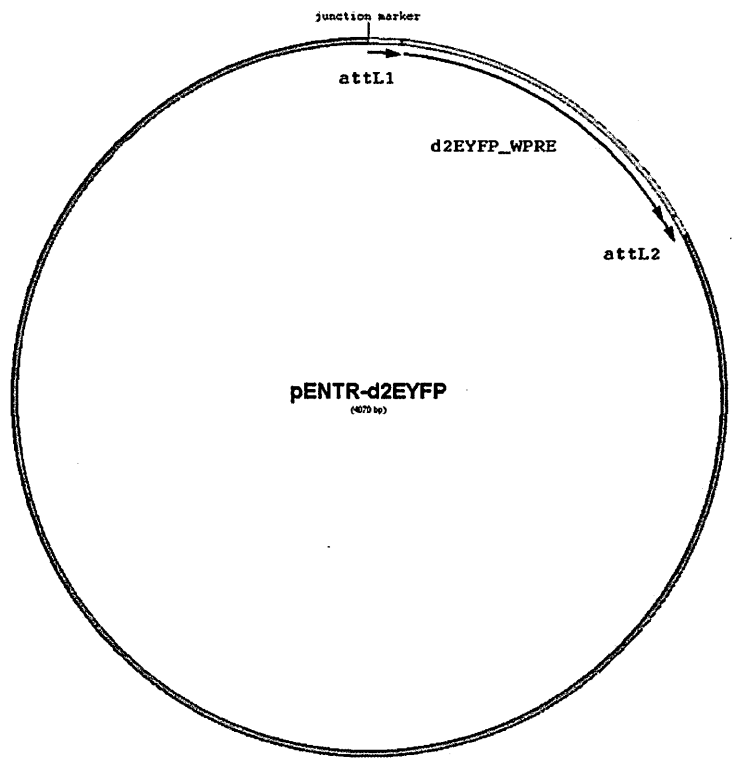

Figure A.6: **pENTR-d2EYFP vector.** Flanking the d2EYFP are the attL1 and attL2 recombination sites.

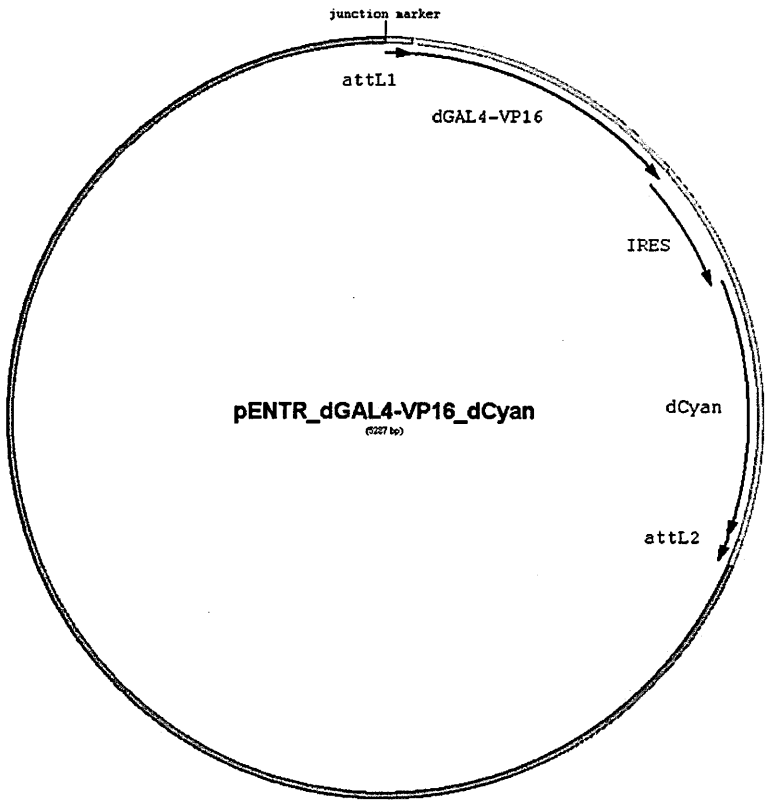

Figure A.7: **pENTR-GAL4-VP16 vector.** Flanking the gene cassette are the attL1 and attL2 recombination sites.

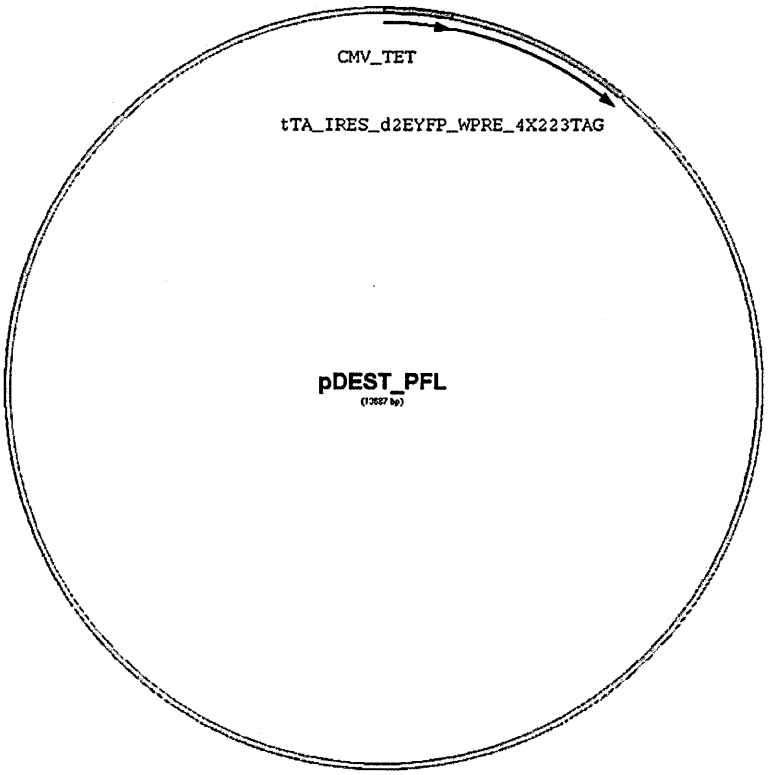

Figure A.8: **pDEST-PFL vector.** From the recombination reaction between three vectors we obtained the positive feedback loop

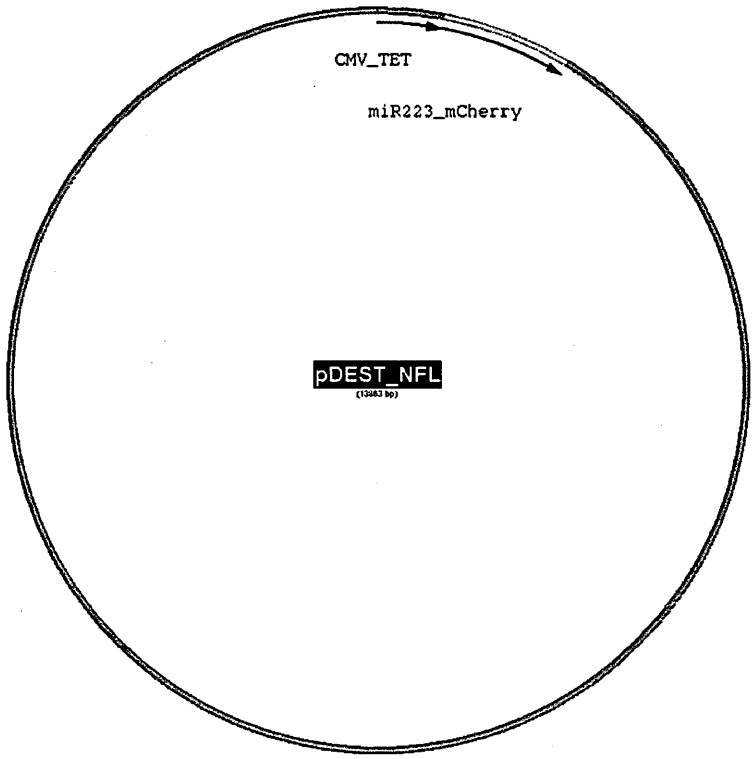

Figure A.9: **pDEST-NFL vector.** Obtained from recombination reaction between the plasmids carrying promoter, gene expression cassette e viral backbone.

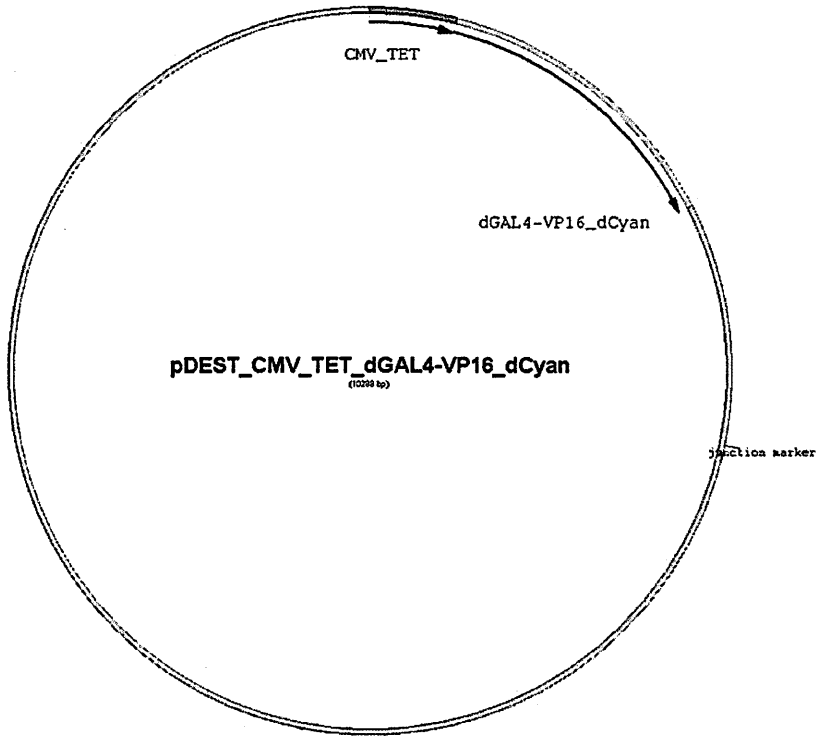

Figure A.10: **pDEST-GAL4 vector.** Obtained from recombination reaction between the plasmids carrying promoter, gene expression cassette e viral backbone.

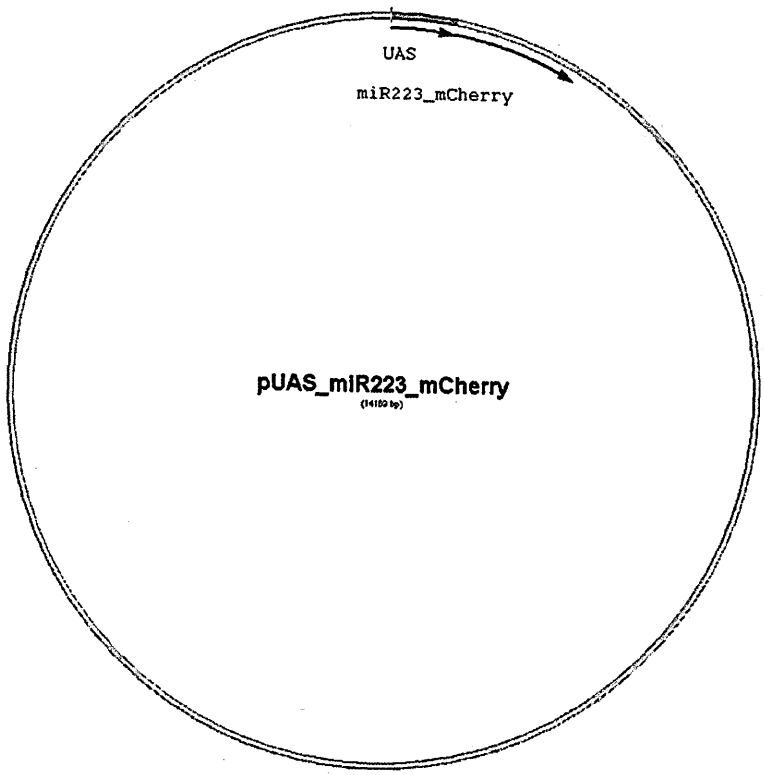

Figure A.11: **pDEST-UAS-miR223-mCherry vector.** Obtained from recombination reaction between the plasmids carrying promoter, gene expression cassette e viral backbone.

Appendix B

Publication

- **Siciliano V.**, Menolascina F., Marucci L., Fracassi C., Garzilli I., Moretti M.N., diBernardo D. Construction and modelling of an inducible positive feedback loop stably integrated in a mammalian cell-line. PLoS Computational Biology(2011, vol.7 (e1002074))

- BelcastroV., **Siciliano V.**, Gregoretti F., Pratibha M., Iorio F., Santoro M., DAngelo G., Oliva G., Brunetti-Pierri N., diBernardo D., Gene regulatory network inference from a massive dataset elucidates transcriptome organization and gene function. Nucleic Acid Research(2011, 39(20): 8677-8688)

- Cuccato, G.,Polynikis,A., **Siciliano V.**, Graziano, M., di Bernardo, M., diBernardo, D., Modeling RNA interference in mammalian cells. BMC System Biology. (2011,(1752-0509) 1-9)

- Gregoretti F., Belcastro V.,**Siciliano V.**, Santoro M.,D'Angelo G., Oliva G., di Bernardo D., Reverse-engineering and analysis of genome-wide gene regulatory networks from gene expression profiles using high-performance parallel computing .Transactions on Computational Biology and Bioinformatics (2010)

Bibliography

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

[2] Mario Amendola, Laura Passerini, Ferdinando Pucci, Bernhard Gentner, Rosa Bacchetta, and Luigi Naldini. Regulated and multiple miRNA and siRNA delivery into primary cells by a lentiviral platform. *Molecular therapy : the journal of the American Society of Gene Therapy*, 17(6):1039–1052, June 2009.

[3] James Anderson and Antonis Papachristodoulou. On validation and invalidation of biological models. *BMC Bioinformatics*, 10, 2009.

[4] Sylvain Arlot and Alain Celisse. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79, 2010.

[5] Mariette R. Atkinson, Timothy A. Blauwkamp, Vladamir Bondarenko, Vasily Studitsky, and Alexander J. Ninfa. Activation of the glna, glnk,

and nac promoters as escherichia coli undergoes the transition from nitrogen excess growth to nitrogen starvation. *J. Bacteriol.*, 184(19):5358–5363, 2002.

[6] Mariette R. Atkinson, Timothy A. Blauwkamp, and Alexander J. Ninfa. Context-dependent functions of the pii and glnk signal transduction proteins in escherichia coli. *J. Bacteriol.*, 184(19):5364–5375, 2002.

[7] Mariette R. Atkinson, Michael A. Savageau, Jesse T. Myers, and Alexander J. Ninfa. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in escherichia coli. *Cell*, 113(5):597–607, May 2003.

[8] Dominique Aubel and Martin Fussenegger. Watch the clock?engineering biological systems to be on time. *Current Opinion in Genetics & Development*, October 2010.

[9] Alessia Baccarini, Hemangini Chauhan, Thomas J. Gardner, Anitha D. Jayaprakash, Ravi Sachidanandam, and Brian D. Brown. Kinetic analysis reveals the fate of a MicroRNA following target regulation in mammalian cells. *Curr Biol*, 21(5):369–376, March 2011.

[10] Naama Barkai and Stanislas Leibler. Biological rhythms: Circadian clocks limited by noise. *Nature*, 403(6767):267–268, January 2000.

[11] David P. Bartel and Chang-Zheng Z. Chen. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nature reviews. Genetics*, 5(5):396–400, May 2004.

[12] A. Becskei, B. Séraphin, and L. Serrano. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *The EMBO journal*, 20(10):2528–2535, May 2001.

[13] Chase L. Beisel, Travis S. Bayer, Kevin G. Hoff, and Christina D. Smolke. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Mol Syst Biol*, 4, October 2008.

[14] P. Bogacki and L. F. Shampine. A 3(2) pair of runge-kutta formulas. *Appl. Numer. Math.*, 2:1–9, 1989.

[15] Olga Boubaker and Ada Fourati. Structural identifiability of non linear systems : An overview. *Industrial Technology, 2004, IEEE ICIT*, 3:1224–1248, 2004.

[16] Brian D Brown, Bernhard Gentner, Alessio Cantore, Silvia Colleoni, Mario Amendola, Anna Zingale, Alessia Baccarini, Giovanna Lazzari, Cesare Galli, and Luigi Naldini. Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state. *Nat Biotech*, 25(12):1457–1467, December 2007.

[17] T. R. Brummelkamp, R. Bernards, and R. Agami. A system for stable expression of short interfering RNAs in mammalian cells. *Science*, 296(5567):550–553, April 2002.

[18] Thijn R. Brummelkamp, Rene Bernards, and Reuven Agami. A System for Stable Expression of Short Interfering RNAs in Mammalian Cells. *Science*, 296(5567):550–553, 2002.

[19] L. Cardone, J. Hirayama, F. Giordano, T. Tamaru, J. J. Palvimo, and P. Sassone-Corsi. Circadian clock control by SUMOylation of BMAL1. *Science*, 309(5739):1390–1394, August 2005.

[20] Richard W. Carthew and Erik J. Sontheimer. Origins and mechanisms of miRNAs and siRNAs. 136(4):642–655, February 2009.

[21] Daniela Castanotto, Kumi Sakurai, Robert Lingeman, Haitang Li, Louise Shively, Lars Aagaard, Harris Soifer, Anne Gatignol, Arthur Riggs, and John J. Rossi. Combinatorial delivery of small interfering RNAs reduces RNAi efficacy by selective incorporation into RISC. *Nucleic acids research*, 35(15):5154–5164, July 2007.

[22] Thimmaiah P. Chendrimada, Richard I. Gregory, Easwari Kumaraswamy, Jessica Norman, Neil Cooch, Kazuko Nishikura, and Ramin Shiekhattar. TRBP recruits the dicer complex to ago2 for microRNA processing and gene silencing. *Nature*, 436(7051):740–744, August 2005.

[23] George M. Church. From systems biology to synthetic biology. *Molecular Systems Biology*, 1(1):msb4100007–E1–msb4100007–E2, March 2005.

[24] Claudio Cobelli and Joseph J. Distefano 3rd. Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *Am J Physiol.*, 239:R7–24, 1980.

[25] Emery Conrad, Avraham E. Mayo, Alexander J. Ninfa, and Daniel B. Forger. Rate constants rather than biochemical mechanism determine behaviour of genetic clocks. *Journal of The Royal Society Interface*, 5(Suppl 1):S9–S15, August 2008.

[26] Damien Coudreuse and Paul Nurse. Driving the cell cycle with a minimal CDK control network. *Nature*, 468(7327):1074–1079, December 2010.

[27] Markus W. Covert, Thomas H. Leung, Jahlionais E. Gaston, and David Baltimore. Achieving stability of lipopolysaccharide-induced nf-kappab activation. *Science*, 309(5742):1854–1857, September 2005.

[28] G. Cuccato, G. Della Gatta, and D. di Bernardo. Systems and synthetic biology: tackling genetic networks and complex diseases. *Heredity*, 102(6):527–532, June 2009.

[29] Giulia Cuccato, Athanasios Polynikis, Velia Siciliano, Mafalda Graziano, Mario di Bernardo, and Diego di Bernardo. Modeling RNA interference in mammalian cells. *BMC Systems Biology*, 5(1):1–9, 2011.

[30] Eric A. Davidson and Andrew D. Ellington. Synthetic RNA circuits. *Nature Chemical Biology*, 3(1):23–28, December 2006.

[31] Hidde De Jong. Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 9:67–103, 2002.

[32] T L Deans, C R Cantor, and Collins J J. A Tunable Genetic Switch Based on RNAi and Repressor Proteins for Regulating Gene Expression in Mammalian Cells. *Cell*, 130:363–372, 2007.

[33] M. L. Dequant, E. Glynn, K. Gaudenz, M. Wahl, J. Chen, A. Mushegian, and Pourqui. A complex oscillating network of signaling genes underlies the mouse segmentation clock. *Science*, 314(5805):1595–1598, December 2006.

[34] di Bernardo, Diego, Thompson, Michael J., Gardner, Timothy S., Chobot, Sarah E., Eastwood, Erin L., Wojtovich, Andrew P., Elliott, Sean J., Schaus, Scott E., and Collins, James J. Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. *Nature Biotechnology*, 23(3):377–383, March 2005.

[35] J. Dunlap. Molecular bases for circadian clocks. *Cell*, 96(2):271–290, January 1999.

[36] Erik J. Eide, Erica L. Vielhaber, William A. Hinz, and David M. Virshup. The circadian regulatory proteins bmal1 and cryptochromes

are substrates of casein kinase i. *Journal of Biological Chemistry*, 277(19):17248–17254, 2002.

[37] S.M. Elbashir, J Harborth, W Lendeckel, A Yalcin, K Weber, and T Tuschl. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, 411:494–498, 2001.

[38] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, January 2000.

[39] J Feng, M R Atkinson, W McCleary, J B Stock, B L Wanner, and A J Ninfa. Role of phosphorylated metabolic intermediates in the regulation of glutamine synthetase synthesis in escherichia coli. *J. Bacteriol.*, 174(19):6061–6070, 1992.

[40] Timothy S. Gardner, Charles R. Cantor, and James J. Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339–342, 2000.

[41] Brian C. Goodwin. *Temporal organization in cells; a dynamic theory of cellular control processes.* Academic Press, London, New York, 1963. ix, 163 p.

[42] Dirk Grimm, Konrad L. Streetz, Catherine L. Jopling, Theresa A. Storm, Kusum Pandey, Corrine R. Davis, Patricia Marion, Felix Salazar, and Mark A. Kay. Fatality in mice due to oversaturation of cellular mi-

croRNA/short hairpin RNA pathways. *Nature*, 441(7092):537–541, May 2006.

[43] Guantes and Poyatos. Dynamical principles of two-component genetic oscillators. *PLoS Comput Biol*, 2(3), March 2006.

[44] Karin Hammer, Ivan Mijakovic, and Peter Ruhdal Jensen. Synthetic promoter libraries tuning of gene expression. *Trends in Biotechnology*, 24(2):53 – 55, 2006.

[45] T. Kafri, H. van Praag, F. H. Gage, and I. M. Verma. Lentiviral vectors: regulated gene expression. *Mol. Ther.*, 1:516–521, 2000.

[46] H Kaufmann, X Mazur, M Fussenegger, and J E Bailey. Influence of low temperature on productivity, proteome and protein phosphorylation of CHO cells. *Biotechnol Bioeng*, 63(5):573–582, June 1999. PMID: 10397813.

[47] Aly A. Khan, Doron Betel, Martin L. Miller, Chris Sander, Christina S. Leslie, and Debora S. Marks. Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat Biotechnol*, 27:549555, 2009.

[48] R Khanin and V Vinciotti. Computational Modeling of Post-Transcriptional Gene Regulation by MicroRNAs. *J Comput Biol*, 15:305–316, 2008.

[49] Daniel H. Kim and John J. Rossi. Strategies for silencing human disease using RNA interference. *Nat Rev Genet*, 8(3):173–184, March 2007.

[50] C. H. Ko and J. S. Takahashi. Molecular components of the mammalian circadian clock. *Hum Mol Genet*, 15 Spec No 2, October 2006.

[51] Beat P. Kramer, Alessando U. Viretta, Marie Daoud-El-Baba, Dominique Aubel, Wilfried Weber, and Martin Fussenegger. An engineered epigenetic transgene switch in mammalian cells. *Nat Biotechnol*, 22(7):867–870, July 2004.

[52] G. Lahav. The strength of indecisiveness: oscillatory behavior for better cell fate determination. *Sci STKE*, 2004(264), December 2004.

[53] Yoontae Lee, Chiyoung Ahn, Jinju Han, Hyounjeong Choi, Jaekwang Kim, Jeongbin Yim, Junho Lee, Patrick Provost, Olof Radmark, Sunyoung Kim, and V. Narry Kim. The nuclear RNase III drosha initiates microRNA processing. *Nature*, 425(6956):415–419, September 2003.

[54] Yoontae Lee, Inha Hur, Seong-Yeon Park, Young-Kook Kim, Mi R. Suh, and Narry V. Kim. The role of PACT in the RNA silencing pathway. *The EMBO Journal*, aop(current), January 2006.

[55] Jidong Liu, Marco A. Valencia-Sanchez, Gregory J. Hannon, and Roy Parker. MicroRNA-dependent localization of targeted mRNAs to mammalian p-bodies. *Nature Cell Biology*, 7(7):719–723, June 2005.

[56] Locke, James C., Southern, Megan M., László Kozma-Bognár, Hibberd, Victoria, Brown, Paul E., Turner, Matthew S., and Millar, Andrew J. Extension of a genetic network model by iterative experimentation and mathematical analysis. *Molecular systems biology*, 1(1):msb4100018–E1–msb4100018–E9, June 2005.

[57] Alexander Loewer, Eric Batchelor, Giorgio Gaglia, and Galit Lahav. Basal dynamics of p53 reveal transcriptionally attenuated pulses in cycling cells. *Cell*, 142(1):89–100, July 2010.

[58] Thomas Maiwald and Jens Timmer. Dynamical modeling and Multi-Experiment fitting with PottersWheel. *Bioinformatics*, page btn350, July 2008.

[59] Yoshito Masamizu, Toshiyuki Ohtsuka, Yoshiki Takashima, Hiroki Nagahara, Yoshiko Takenaka, Kenichi Yoshikawa, Hitoshi Okamura, and Ryoichiro Kageyama. Real-time imaging of the somite segmentation clock: Revelation of unstable oscillators in the individual presomitic mesoderm cells. *Proceedings of the National Academy of Sciences of the United States of America*, 103(5):1313–1318, 2006.

[60] Marjori A. Matzke and James A. Birchler. RNAi-mediated pathways in the nucleus. *Nat Rev Genet*, 6(1):24–35, January 2005.

[61] J. L. McBride, R. L. Boudreau, S. Q. Harper, P. D. Staber, A. M. Monteys, I. Martins, B. L. Gilmore, H. Burstein, R. W. Peluso,

B. Polisky, B. J. Carter, and B. L. Davidson. Artificial miRNAs mitigate shRNA-mediated toxicity in the brain: implications for the therapeutic development of RNAi. *Proceedings of the National Academy of Sciences of the United States of America*, 105(15):5868–5873, April 2008.

[62] Melanie Mitchell. *An Introduction to Genetic Algorithms (Complex Adaptive Systems)*. The MIT Press, February 1998.

[63] Carmen G. Moles, Pedro Mendes, and Julio R. Banga. Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Genome Res.*, 13(11):2467–2474, 2003.

[64] Shankar Mukherji and Alexander van Oudenaarden. Synthetic biology: understanding biological design from synthetic circuits. *Nature reviews. Genetics*, 10(12):859–871, December 2009.

[65] Oliver Nelles. *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Springer, 1 edition, December 2000.

[66] John S. O/'Neill and Akhilesh B. Reddy. Circadian clocks in human red blood cells. *Nature*, 469(7331):498–503, January 2011.

[67] M Overhoff, W Wünsche, and G Sczakiel. Quantitative detection of siRNA and single-stranded oligonucleotides: relationship between uptake and biological activity of siRNA. *Nucleic Acids Res.*, 32:e170, 2004.

[68] J. R. Pomerening, S. Y. Kim, and J. E. Ferrell. Systems-level dissection of the cell-cycle oscillator: bypassing positive feedback produces damped oscillations. *Cell*, 122(4):565–578, August 2005.

[69] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmuller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.

[70] Dae-Kyun Ro, Eric M. Paradise, Mario Ouellet, Karl J. Fisher, Karyn L. Newman, John M. Ndungu, Kimberly A. Ho, Rachel A. Eachus, Timothy S. Ham, James Kirby, Michelle C. Y. Chang, Sydnor T. Withers, Yoichiro Shiba, Richmond Sarpong, and Jay D. Keasling. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, 440(7086):940–943, April 2006.

[71] Uwe Sauer, Matthias Heinemann, and Nicola Zamboni. Getting closer to the whole picture. *Science*, 316(5824):550–551, April 2007.

[72] Daniel Segrè, Vitkup, Dennis, and Church, George M. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(23):15112–15117, 2002.

[73] Hiromi Shimojo, Toshiyuki Ohtsuka, and Ryoichiro Kageyama. Oscillations in notch signaling regulate maintenance of neural progenitors. *Neuron*, 58(1):52–64, 2008.

[74] Velia Siciliano, Filippo Menolascina, Lucia Marucci, Chiara Fracassi, Immacolata Garzilli, Maria Nicoletta Moretti, and Diego di Bernardo. Construction and modelling of an inducible positive feedback loop stably integrated in a mammalian cell-line. *PLoS Comput Biol*, 7(6):e1002074, 06 2011.

[75] Sisler HD Siegel MR. Inhibition of protein synthesis in vitro by cycloheximide. *Nature*, 200:675–676, 1963.

[76] Roy S. Smith and John C. Doyle. Model Validation: A Connection Between Robust Control and Identification. *IEEE Transactions on Automatic Control*, 37:942–52, 1992.

[77] Jesse Stricker, Scott Cookson, Matthew R. Bennett, William H. Mather, Lev S. Tsimring, and Jeff Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456(7221):516–519, October 2008.

[78] Jesse Stricker, Scott Cookson, Matthew R R. Bennett, William H H. Mather, Lev S S. Tsimring, and Jeff Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456(7221):516–519, October 2008.

[79] Jolanta Szulc, Maciej Wiznerowicz, Marc-Olivier Sauvain, Didier Trono, and Patrick Aebischer. A versatile tool for conditional gene expression and knockdown. *Nature methods*, 3(2):109–116, January 2006.

[80] M Tigges, T.T Marquez-Lago, J Stelling, and M Fussenegger. A tunable synthetic mammalian oscillator. *Nature*, 457:309–312, 2009.

[81] Marcel Tigges, Tatiana T. Marquez-Lago, Jorg Stelling, and Martin Fussenegger. A tunable synthetic mammalian oscillator. *Nature*, 457(7227):309–312, January 2009.

[82] S. Topp and J. P. Gallivan. Riboswitches in unexpected places–a synthetic riboswitch in a protein coding region. *RNA (New York, N.Y.)*, 14(12):2498–2503, December 2008.

[83] Hideki Ukai, Tetsuya J. Kobayashi, Mamoru Nagano, Koh-Hei Masumoto, Mitsugu Sujino, Takao Kondo, Kazuhiro Yagita, Yasufumi Shigeyoshi, and Hiroki R. Ueda. Melanopsin-dependent photoperturbation reveals desynchronization underlying the singularity of mammalian circadian clocks. *Nature Cell Biology*, 9(11):1327–1334, October 2007.

[84] Maki Ukai-Tadenuma, Takeya Kasukawa, and Hiroki R. Ueda. Proof-by-synthesis of the transcriptional logic of mammalian circadian clocks. *Nature Cell Biology*, 10(10):1154–1163, September 2008.

[85] V.A. Living colors destabilized ecfp and eyfp vectors. *Clontechniques*, XIV(3):14–15, 1999.

[86] Michael Wassenegger. The role of the RNAi machinery in heterochromatin formation. *Cell*, 122(1):13–16, 2005.

[87] W Weber, J Stelling, M Rimann, B Keller, M Daoud-El Baba, CC Weber, D Aubel, and Fussenegger M. A synthetic time-delay circuit in mammalian cells and mice. *Proc Natl Acad Sci U S A.*, 104(8):2643–2648, 2007.

[88] Maung N. Win and Christina D. Smolke. Higher-Order cellular information processing with synthetic RNA devices. *Science*, 322(5900):456–460, October 2008.

[89] Nir Yosef and Aviv Regev. Impulse control: temporal dynamics in gene transcription. *Cell*, 144(6):886–896, March 2011. PMID: 21414481.