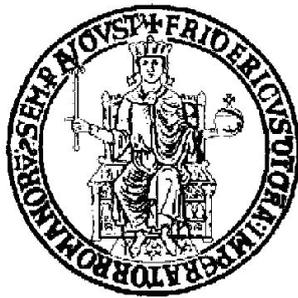


UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II

Degree of Doctor of Philosophy Department of Computer and
Systems Engineering



PhD Thesis

**Mathematical modelling and non-linear
analysis of Synthetic Gene Regulatory
Networks**

Advisors

Dr. Diego di Bernardo

Prof. Mario di Bernardo

Dr. Stefania Santini

Candidate

Lucia Marucci

YEAR 2007/2010

Dedicated to Oscar, because life is the most fascinating discovery.

The most beautiful experience we can have is the mysterious. It is the fundamental emotion which stands at the cradle of true art and true science.

Albert Einstein

Acknowledgements

I would like to thank my supervisors Dr. Diego di Bernardo, Prof. Mario di Bernardo and Dr. Stefania Santini whose expertise, knowledge and unique skills added greatly to my research. They became more of friends to me, and I want to fully express my appreciation and gratitude for their constant direction, support and encouragement. Diego, thanks for being a severe boss, and a great friend. You helped me a lot, especially in difficult moments, I'll never forget this. Mario and Stefania, thanks for having guided me into the engineering world, and for your constant help. Thanks to Irene Cantone, Maria Aurelia Ricci and Velia Siciliano, without you I would not have had anything to model! A very special thank goes to all the people in my two labs: Giulia, Velia, Mario, Francesco, Filippo, Vincenzo, Alda, Maria Aurelia, Stefania, Nicoletta, Imma, Chiara, Giusy, Alberto, Mukesh, Luisa, Rossella, Annamaria, Margherita, Santosh, Gopu, Umberto, Pietro, Achille, Giovanni, Davide, Dino, you were my family, thanks to all of you. Thanks again to Diego, Mario and Stefania for having chosen such wonderful people to work with me, in a stimulating and friendly atmosphere. Thanks to Barbara Zimbardi, Silvana Ruotolo, Mariolina Pepe, Valeria Rotoli, Gabriela Gagliardi,

Valeria Lamorgese and Roberta Ballabio. Thanks to Agostino and Dina, as their smiles were precious, and to all the guys of the Informatics Core.

Thanks to Dr. David Burton for giving me the opportunity to spend two weeks in Bristol University, and to Dr. Maria Pia Cosma and to her lab for letting me playi with the cells for three months in the Center of Genomic Regulation of Barcelona.

Now I want to mention people that walked with me during these three years. Giulia, you are my second sister, I'll miss you a lot. I really hope the best for you and your wonderful brand new family. Monica, Flavio, Otto, Paola, Gennaro, thanks for the wonderful nights and weekends spent together, for the deep love you have for me. Luca, Rosanna and Hugo, thanks for sharing the pleasure of playing music with me. Stefano, thanks for being close to me and for having shown me the beauty again.

And thanks Naples, the sunsets in Dazio, the smell of the pizzerias, the old women in the old part of the city, the charm of this incredible city. And I want to thank my town, L'Aquila, that is the only place in which I really feel home.

Now my family comes. Let's start with my parents, my example of how life can be lived in a real, deep and sincere way. Marina, Carmine, your strength will never end to surprise me. Your love is what I'll always keep as the most precious present life gave me. Nonna Maria, thanks for your smiles that, whatever happens, light your face. And finally thanks to my sister, Mariagrazia, a beautiful, pure child 32 years old, for her love and constant help. And for the gorgeous swedish landscapes we admired together.

Contents

1	Introduction	1
1.1	System and Synthetic biology	2
1.2	Motivation and thesis outline	4
2	Background	7
2.1	Physical mechanisms of genetic regulation	7
2.1.1	Gene regulatory networks and network motifs	9
2.2	Mathematical modelling of gene regulatory networks	14
2.2.1	Model derivation approaches	15
2.2.2	Modelling approaches	18
3	A novel synthetic network in yeast: IRMA	23
3.1	Introduction	24
3.2	Construction of the synthetic network	26
3.2.1	Choice of model organism	26
3.2.2	Choice of the network genes	27
3.2.3	Choice of network motifs and resulting topology	27
3.3	Testing <i>in vivo</i> the network response to the inducers.	30

3.3.1	Switching the network on by injecting galactose	30
3.3.2	Inducibility of the network using methionine	31
3.4	Discussion	34
4	Derivation, identification and validation of the mathematical model of IRMA network.	35
4.1	Derivation of model equations: step (i). Model A	37
4.2	Simplification of the complete model: Model B	40
4.2.1	Identification of model parameters: step (ii)	43
4.2.2	Validation of model predictive performance: step (iii)	44
4.3	Additional experimental investigation	48
4.4	Model refinement: Model C (step (i))	50
4.4.1	Identification of the model parameters and validation of its predictive performance: steps (ii) and (iii)	52
4.4.2	Experimental identification of the Hill function parameters	54
4.5	Further model refinement: Model D (back to step (i))	55
4.5.1	Identification of parameters and validation of model D: step (ii) and (iii)	58
4.6	Discussion	63
5	Turning IRMA into an autonomous oscillator or a bistable switch: non-linear analysis and continuation results	66
5.1	Turning IRMA into an oscillator	68

5.1.1	Scenario 1: stable oscillations keeping the activation of Swi5 on <i>CBF1</i> (DDEs model). Simulation and continuation results.	71
5.1.2	Scenario 2: stable oscillations by removing the activation of Swi5 on <i>CBF1</i> (ODEs model). Simulation and continuation results.	79
5.1.3	Scenario 3: stable oscillations by removing the activation of Swi5 on <i>CBF1</i> and by adding a positive auto-feedback loop on <i>SWI5</i> (ODEs model). Simulation and continuation results.	83
5.2	Turning IRMA into a bistable switch.	86
5.2.1	Experimental implementation of Scenario 4 <i>in vivo</i>	87
5.3	Discussion	89
6	Response of the synthetic network in yeast to an external periodic input	91
6.1	Entrainment to periodic input; <i>in silico</i> experiments	93
6.1.1	Numerical results of the oscillatory version of the network (Scenario 2)	93
6.1.2	Bifurcation analysis	97
6.1.3	Numerical results on the non-oscillatory version of the network (Scenario 2 modified)	97
6.2	Analytical results	99
6.2.1	Introduction to contraction theory	99
6.2.2	Contraction results on IRMA	101

6.3	Discussion	107
7	Mathematical model of a novel synthetic oscillator in mam- malian cells	109
7.1	Aim of the project	109
7.2	Description of the biological system: Oscillator topology 1 . .	110
7.3	Mathematical model of the network and continuation results .	111
7.3.1	Identification of the parameters oscillatory region and continuation results	114
7.4	Model guided re-engineering of the network: oscillator Topol- ogy 2	115
7.4.1	Mathematical model, identification of the parameters oscillatory region and continuation results	118
7.5	Preliminary <i>in vivo</i> data and parameter identification results .	119
7.5.1	Mathematical model of the inducible positive feedback loop and parameters fitting	121
7.6	Discussion	123
8	Concluding remarks and future work	127
8.1	Future work	130
8.1.1	Yeast synthetic network	130
8.1.2	Synthetic oscillator in mammalian cells	131
A	Parameters of the mathematical models	132
B	Hybrid Genetic Algorithm	138

B.1	Problem statement	138
B.2	Description of the algorithm	139
C	Matlab code of the mathematical models	146
C.1	Models of IRMA	146
C.2	Models of the re-engineered versions of IRMA	153
C.3	Models of the synthetic oscillator in mammalian cells	159
D	Experimental procedures	162
D.1	Experiments performed on the yeast synthetic network	162
D.1.1	Construction of <i>S. cerevisiae</i> strains	162
D.1.2	Data collection	164
D.2	Preliminary experiments performed on the synthetic oscillator in mammalian cells	166
D.2.1	Construction of the inducible positive feedback loop	166
D.2.2	Cell culture lentiviral transduction switch-off experiment	167
D.2.3	Data processing	169

Chapter 1

Introduction

The functioning and development of living organisms is controlled on the molecular level by networks of genes, proteins, small molecules, and their mutual interactions, the so-called *gene regulatory networks*. Common interactions in these systems are feedback loops, in which, as the name suggests, information from the output of a system transformation is sent back to the input of the system. The presence of interacting parts determines the *complexity* of biological networks.

Moreover, the interactions are non-linear, making gene regulatory networks *emergent systems* ([80], [52]). It means that, unlike systems which can be modelled by considering averaged effects, it is not possible to reduce the systems behaviour to the sum of its parts, or study one part at a time, with the expectation of understanding the emergent properties of the whole system [24].

The inherent complexity and irreducibility of biological phenomena show the

need of looking at the whole picture, in an holistic sense, and using proper mathematical tools. The cell must be studied at the systems level by unraveling the regulatory, signalling and metabolic interactions, and understanding their coordinated action, if we are to continue to make strides in our understanding of these phenomena. Prediction, control, and understanding arise mainly from modelling these systems using iterated computer simulations and non-linear mathematical analysis. Biotechnological advances in quantitative high-throughput technology, in combination with the growing interdisciplinarity between biology with engineering and natural sciences, have made this challenge achievable thanks to the emerging fields of Systems and Synthetic biology ([48] [56], [99], [23]).

1.1 System and Synthetic biology

Systems biology aims at developing a formal understanding of biological processes using computational techniques. Systems biology can proceed in two directions: a “bottom-up” approach, in which starting from detailed knowledge of a biological process of interest, mathematical language is used to quantitatively describe the biological information and the experimental data into a model of the process under study [23]. The biological process is thus represented as a network describing interactions between genes, proteins, metabolites and other molecules. A model is a mathematical formalism to describe changes in concentration of each gene transcript and protein as a function of their regulatory interactions (gene regulatory network). This net-

work can then be used to probe the behaviour of the biological process using computer simulations and mathematical analysis, to generate novel hypotheses to be then tested *in vivo*. The Systems biology “top-down” approach, on the contrary, aims at learning the network of gene-gene interactions for a biological process for which very limited knowledge is available. This approach is called “reverse engineering” and, typically, makes use of high-throughput gene expression profiling following a variety of perturbations to the cell to learn gene-gene interactions.

The emerging discipline of **Synthetic biology** can be defined as the engineering of biology. 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 [32] or genetic switching [39]), 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 [91]). Reconstructing simplified replicas of natural genetic circuits helps to understand the sufficient and essential biological features that drive a specific function. This approach is well-known in engineering, in which problems are often tackled via simplified empirical models of the process to be studied, where the complexity is reduced to facilitate its handling, but its key features are kept. For example, a jumbo-jet contains over six million parts and is complex enough to be incomprehensible to the human mind without appropriate simplifications. Nevertheless a simplified toy model of a flying airplane retains some of the most complex and relevant features of the jumbo-jet (fluidodynamics and control) and it is routinely

used to derive models and design principles for the full-scale plane [22].

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. In particular, quantitative models are needed for a precise and unambiguous description of synthetic circuits [58].

The usefulness of a model in both Systems and Synthetic biology lies in its ability to formalise the knowledge about the biological process at hand, to identify inconsistencies between hypotheses and observations, and to predict the behaviour of the biological process in yet untested conditions. The aim must be to develop holistic models which capture the essence of various interactions within the system and are able to analyse and give predictions of the system as a whole. From this perspective, the theory of dynamical systems and control can have a fundamental role in the analysis, modelling and design of synthetic biological circuits.

1.2 Motivation and thesis outline

The main aim of this thesis is to explore and solve some of the challenges faced by a mathematical modeller when studying gene regulatory networks. In particular, in Chapter 2 we will describe the systems of interest (gene regulatory networks), and the methodology used to model them (Ordinary Differential Equations).

In Chapters 3 we will introduce a synthetic network that we built *de novo* in the yeast *Saccharomyces cerevisiae*. The goal of our work was to provide the System biology community with an *in vivo* benchmark, which can be used as “ground of truth” to test and compare modelling approaches and reverse-engineering inference strategies as, at present, the usefulness and predictive ability of these computational approaches in the field of Synthetic biology cannot be assessed and compared rigorously.

In Chapter 4 the mathematical modelling of the yeast synthetic biological is derived. All the steps are reported: model derivation, experimental design, parameter identification and model validation.

In Chapter 5, we will show how to use novel tools from numerical bifurcation theory (e.g. DDE-BIFTOOL [35], able to deal with delayed systems), together with recent results on the link between the dynamics and topology of networks, in order to redesign the yeast synthetic circuit, turning it into an autonomous oscillator, or a bistable switch. The results presented in this Chapter were derived in collaboration with Dr David A. W. Barton (Applied Non-linear Mathematics Research Group in the Department of Engineering Mathematics of the University of Bristol).

In Chapter 6 we will analyse the response of the yeast synthetic network to an external periodic input. Such forcing can lead to entrainment, that means that the period of the forced oscillator is exactly the one of the external signal and that the phase of the oscillations are locked. We will analyse the entrainment both via simulation and analytically, using recent contraction theory results [93].

In Chapter 7 we will present preliminary results about the modelling and construction of a novel synthetic oscillator in mammalian cells is presented . A full mathematical analysis is developed for two possible topologies of the oscillator, and preliminary data are presented with the relative parameters fitting.

Finally, in Chapter 8, conclusions are drawn and possible directions for future work are suggested.

Finally, Part of the research performed in this Thesis was under the European Union funded project COBIOS [1].

Chapter 2

Background

This Chapter provides a background on the physical mechanisms involved in gene regulatory network, and network motifs. Moreover, we present an overview of different strategies to model gene regulatory network.

2.1 Physical mechanisms of genetic regulation

Living cells are the product of gene expression programs involving regulated transcription of thousands of genes.

The central dogma, schematically described in Figure 2.1, defines the paradigm of molecular biology. Genes are perpetuated as sequences of nucleic acid, but function by being expressed in the form of proteins [9]. *Transcription* and *translation* are responsible for their conversion from one form to the other. Transcription generates a messenger RNA (mRNA) which provides an inter-

mediate that carries the copy of a DNA sequence that represents a protein. It is a single-stranded RNA identical in sequence with one of the strands of the duplex DNA. In protein-coding genes, translation will convert the nucleotide sequence of mRNA into the sequence of amino acids comprising a protein [9]. This two-stage process is called gene expression. Gene expression is a complex process regulated at several stages in the synthesis of proteins [9]. Some proteins are structural and will accumulate at the cell-wall or within the cell to give it particular properties. Other proteins can be enzymes that catalyse certain reactions. A large group of proteins have an important role in the regulation of the genes, known as *transcription factors*. Gene regulation by transcription factors can be negative or positive. In negative regulation, an inhibitor protein binds to the promoter (a region of DNA that facilitates the transcription of a particular gene), and decreases the mRNA transcription of the gene. In positive regulation, a transcription factor is required to bind at the promoter in order to increase the mRNA transcription rate [2]. Several other steps in the gene expression process may be modulated [9].

Apart from DNA transcription regulation, the expression of a gene may be controlled during RNA processing and transport (in eukaryotes), RNA translation, and the post-translational modification of proteins [25]. The degradation of gene products can also be regulated in the cell. Recent work is forcing a rethink of the roles of RNA and proteins in cell control mechanisms. Until recently, RNA was not believed to have a role in regulation of gene expression. Now it is known that small RNA molecules can act, through RNA interference (RNAi) mechanism, to silence gene expression (see [37], [13]).

Two examples are small interfering RNAs (siRNAs) and microRNAs (miRNAs) [16]. RNA interference regulates transcription by inducing degradation of targeted mRNAs.

In this Thesis, we will present preliminary results about a synthetic oscillator we are currently building in mammalian cells (Chapter 7) that includes in its topology a microRNA towards the gene of interest.

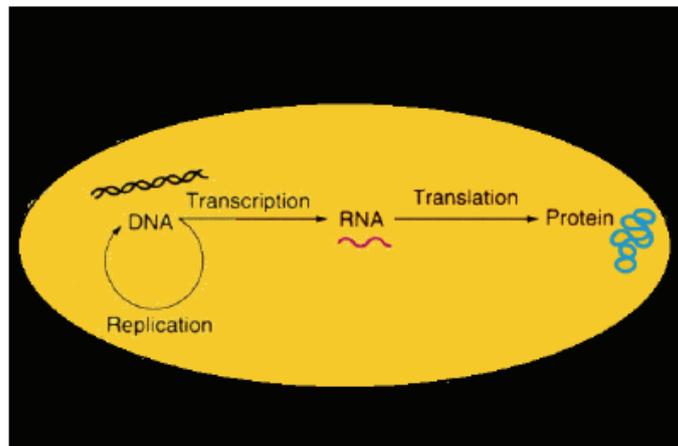


Figure 2.1: **Schematic diagram of the central dogma of molecular biology.** Figure reproduced from [45].

2.1.1 Gene regulatory networks and network motifs

Hence, a gene regulatory network is a collection of DNA, RNA, proteins, and other molecules which interact with each other. How a collection of regulatory proteins associates with genes across a genome can be described as a network in which the nodes are genes and the edges represent regulations among them. In the graph, directed edges with an arrow end represent activation (Figure 2.2 (A)), whereas a dash end represents inhibition (Figure 2.2

(B)).

Since the cell is not an isolated system, but it continuously responds to external stimuli in order to follow a specific developmental program or to adapt to changing environmental conditions, the transcriptional network is a dynamic system: after an input signal arrives, transcription factor activities change, leading to changes in the production rate of proteins.

In order to study the complex dynamics of cellular networks, during the last years several studies aimed to identify the basic building-blocks of transcriptional networks and to study the functional relevance of these modular components ([64], [74], [75], [117]). The approach is based on the identification of meaningful patterns on the basis of statistical significance. To define statistical significance, the real network is compared to an ensemble of randomized networks, which have the same number of nodes and edges as the real one, but where the connections are made at random. If a pattern occurs in the real network significantly more often than in the randomized networks, it is defined as a *network motif*. The basic idea is that network motifs that occur in the real network more often than in randomized networks must have been preserved over evolutionary timescales against mutations that randomly change edges. As a matter of fact, point mutations, which occur in a promoter sequence, can alter the binding of a specific transcription factor to the promoter thus resulting in the loss of an edge of the transcriptional network. Similarly, new edges can be added to the network by either point mutations or by duplication events in a promoter region, thus generating a new binding site for a transcription factor. Hence, conserved network motifs must have

been selected in order to survive during evolution because they provide some advantage to the organism. If a motif did not offer a selective advantage, it would be “washed out” and occur about as often as in randomized networks. A *feed-forward loop* (FFL) ([70], [2]) is defined by a transcription factor (X) that regulates a second transcription factor (Y), such that both X and Y jointly bind a common target (Z). Since each of the regulatory interactions may either be positive or negative there are possibly eight types of FFL motifs. Two of those are the most frequently found: the coherent type 1 showed in Figure 2.2 (C) (where all interactions are positive) and the incoherent type (X activates Z and also activates Y which represses Z). In addition to the structure of the circuit, the way in which the signals from X and Y are integrated by the Z promoter should also be considered. In most of the cases the FFL is either an AND gate (X and Y are required for Z activation) or OR gate (either X or Y are sufficient for Z activation), but other input function are also possible [2].

In Figure 2.3 we report the other most common network motifs. A *regulatory chain* consists of chains of three or more transcription factor in which one regulator binds the promoter for a second regulator, and the second binds the promoter for a third regulator and so forth.

An *auto-regulation motif* consists of a transcription factor that binds its own promoter. This motif is thought to reduce response time to environmental stimuli and increase stability of gene expression.

A *multi-component loop* motif consists of a regulatory circuit whose closure involves two or more factors. The closed loop structure provides the capacity

for the feedback control and offers the potential to produce bistable systems, which switch between two alternative states. This motif is peculiar of yeast and of developmental networks of higher eukaryotes since, with the exception of the auto-regulation, feedback loops composed only by direct transcriptional interactions have not been identified in bacteria.

The *single input* motif contains a single regulator that binds a set of target genes. In this way the expression of the target genes is coordinated under a specific condition. In the *multiple input* motif, there is the presence of multiple regulators acting on the metabolites of the system.

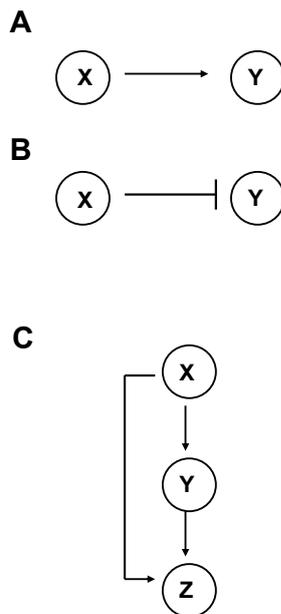


Figure 2.2: **Representation of gene interactions and coherent feed-forward loop.** (A) Directed edges with an arrow end represent activation. (B) Directed edges with a dash end represents inhibition. (C) Schematic representation of a coherent feed-forward loop.

In this Thesis, we will present two novel synthetic network built at Telethon

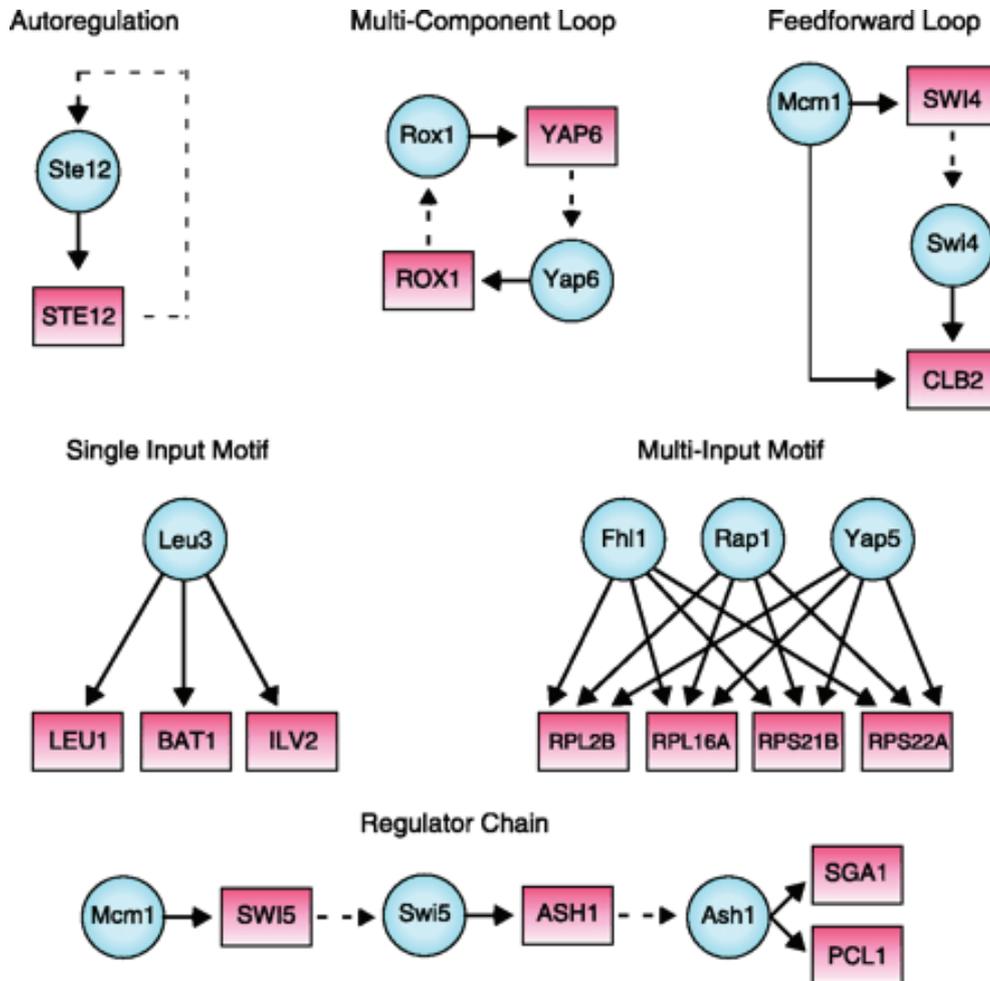


Figure 2.3: **Examples of network motifs in the yeast regulatory network.** Schematic representation of network motifs identified in [64]. Regulatory proteins are represented as blue circles, while their target promoters as red rectangles. A solid arrow indicates binding of a regulator to a promoter. The dashed arrow links the gene to its protein product, representing transcription and translation processes.

Institute of Genetics and Medicine (TIGEM) of Naples, one in yeast (Chapters 3, 4, 5 and 6) and one in mammalian cells (Chapter 7). The topologies of our circuits include various kinds of the mentioned motifs, and we will show how the presence of them is fundamental to understand and eventually tune the networks dynamics.

2.2 Mathematical modelling of gene regulatory networks

In the fields of Systems and Synthetic biology, theory and experiments need to be viewed as a 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 [66]. Modelling a cellular process can highlight which experiments are likely to be the most informative in testing model hypotheses, and allow testing for the effect of drugs [28] or mutant phenotypes [94] on cellular processes, thus paving the way for individualized medicine.

A mathematical model is a 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. Ideally, all information relevant to a system (not only concentrations and rates of events, but also spatial distribution, diffusion parameters, and so on) would be known to make

a maximally accurate *in silico* replica of the system. 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, 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).

2.2.1 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, while in *black-box* modelling the model is completely derived from input-output data. The third alternative, the so-called *gray-box* approach [83], combines the two above approaches. This is the approach we will use in this Thesis. Specifically, first principles are used to partially derive the model structure, while parameters or terms in the model are determined by measurement data.

In this case, modelling 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 [4]) of the model.

Step (i) requires introducing simplifying hypothesis and choosing a proper formal framework. A huge variety of mathematical formalisms have been proposed; we will discuss them in the next subsection.

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 [114]. The notion of identifiability addresses feasibility of estimating unknown parameters from data collected in well-defined stimulus-response experiments [18]. Structural non-identifiability is related to the model structure independently from experimental data. In contrast, practical non-identifiability 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 [90]. Unfortunately, while being well assessed in the case of linear dynamical systems, the identifiability analysis of highly non-linear systems remains an open problem [12]

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 [77], 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) [76] provide a very flexible approach to non-linear optimization. Their application showed good results in the parametrisation of synthetic networks [115], [107].

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 [98]. 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 [98]. To this aim, one possible approach is to use a cross-validation like procedure [6] 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 [4]. 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.

2.2.2 Modelling approaches

A huge variety of mathematical formalisms have been proposed in the literature, such as directed graphs, Bayesian networks, Boolean networks and their generalizations, ordinary and partial differential equations, qualitative differential equations, stochastic equations, and rule-based formalisms (see, for example, [26], [29], [102] and references therein).

A major distinction can be done between qualitative and quantitative modes.

Qualitative 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. As an example, computer science algorithms used to perform code checks can assess the logical consistency of a set of statements: that is, check that no subset of statements is in contradiction with any other [8]. Automated tools such as these and others used in qualitative reasoning approaches become indispensable if logical inferences are to be made on very large sets of experimental observations.

In qualitative modelling, kinetic processes are simulated by tracking over

discrete time the state of the system, defined in terms of a coarse range for each variable. The weak specification of such models conserves computer resources needed to explore the space of possible behaviours; moreover, it provides high-level predictions applicable to a whole family of systems. Although simulation of qualitative models can be fast, even a rough exploration of parameter space can become intractable as the size of the system increases, highlighting the need for increasing computer resources and methods to accelerate the parameters search. For genes that are naturally found in only two states, the trade-off in accuracy may not even be high. On the other hand, simple models can, in some cases, predict behaviours that are far away from reality [29].

Quantitative models

Compared with qualitative models, quantitative ones 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 models can be either deterministic or stochastic.

Deterministic formalisms are commonly used to describe the average behaviour of a population of cells [26]. They have been shown to be viable

for the analysis of synthetic networks in a number of works (e. g. [32], [39], [60], [107], [101]). 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) [2].

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. Usually, the function describing transcriptional interactions are non-linear Hill function or Michealis-Menten, the two differing for considering or not, respectively, the cooperativity of a protein on the gene of interest [2]. The Hill functions model transcriptional interaction in the following way:

- activation: $H^+(y; k, h) = \frac{y^h}{y^h + k^h}$;
- repression: $H^-(z; k, h) = \frac{k^h}{y^h + k^h}$;
- combination of activation or repression $H^{+-} = H^+(y; k, h)(\cdot, +)H^-(z; k_1, h_1)$,

where $(\cdot, +)$ indicates that we can either sum or multiply the Hill functions in the case of multiple regulation, depending on the AND or OR kind of interaction [2]; y and z represent transcription factor levels, h are the Hill coefficients (pure numbers that refer to the cooperativity of the activation binding reaction) and k are the Michaelis-Menten constants, equal to the amount of transcription factor needed to reach half maximal activation (or repression). In the case of Michalis-Menten term, the formalism is identical, but the Hill coefficient h is fixed equal to 1. For a complete derivation of Hill

and Michealis-Menten formalism please refer to [2].

The DEs modelling approach is based on the following biological assumptions: the quantified concentrations do not vary with respect to space and they 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. Of note, the models presented in this Thesis are all Differential Equations based, because the systems of interest satisfy the above assumptions.

As the number of molecular species and consequently of reaction events decreases, the probabilistic nature of biological events becomes more evident. In this case, the response of individuals within a population of genetically identical cells may be significantly different from the average population response. Population heterogeneity arises from stochasticity in molecular events or from noise. For instance, occurrence of noise have been found to be exploited by cells to survive a variety of environmental changes [103] or to increase sensitivity in signal transduction processes [47]. To model such stochastic systems, two main methods are used. The first comprises using *stochastic differential equations* (SDEs, derived from DEs by adding noise terms to the equations), the solutions for which can be numerically obtained either by computing many trajectories (Monte Carlo methods) or approximating their probability distribution and then calculating statistical measures (such as mean and variance). Notably, with this method noise is imposed on the system and represented by mathematical terms chosen *a priori*, instead of arising from

the underlying physical interactions.

The second method is a very successful and exact one introduced nearly 30 years ago, and recently enhanced to cope with different reaction timescales or space constraints. With this approach, molecules are modelled individually and reaction events are calculated by their probability, basing on the Chemical Master Equation [41]. For simulation, usually the Gillespie stochastic simulation algorithm (SSA) [40] is used. It does not try to numerically solve the Master Equation for a given system, but is a systematic, computer-oriented procedure in which Monte Carlo techniques are employed to numerically simulate the discrete Markov process that the Master Equation describes analytically.

Of note, the price to pay for having a more physically realistic model is the considerable increase in computational time and the need for specialized algorithms [102].

Chapter 3

A novel synthetic network in yeast: IRMA

Here we will introduce IRMA, a synthetic network we built in the yeast *Saccharomyces cerevisiae* to benchmark modelling approaches. In this Chapter, we will present the design, topology and construction of the network. As described in the following Chapter, it was necessary to go through iterative refinement steps both in the model and in the experimental data-set. The mathematical modelling was fundamental to design *ad hoc* experiments to clarify the behaviour of the network. In Chapter 4 we will then present in details the approaches we used to construct a dynamical model of IRMA based on Ordinary Differential Equations. In Chapter 5, we will present an additional analysis we performed on the non-linear model with the aim of understanding if and how the network can be turned into an autonomous oscillator, or a bistable switch. Finally, in Chapter 6 we will analyse the re-

sponse of the network to external periodic input.

The work described in Chapters 3 has been partly described in [15]. All the *in vivo* experiments were carried out by Irene Cantone and Maria Aurelia Ricci, in Dr. Maria Pia Cosma's lab (TIGEM, Italy). Details about the experimental procedures are reported in Appendix D.

3.1 Introduction

The goal of our work was to provide the System biology community with an *in vivo* benchmark, which can be used as “ground truth” to test and compare modelling approaches and reverse-engineering inference strategies. At present, the usefulness and predictive ability of computation approaches in the field of Synthetic biology cannot be assessed and compared rigorously. To this aim we constructed, in the yeast *Saccharomyces cerevisiae*, a synthetic network of five genes regulating each other for In-vivo Reverse-engineering and Modelling Assessment (IRMA).

The network was designed to be negligibly affected by endogenous genes, and to respond to galactose, which triggers transcription of its genes. Our network (Figure 3.1), apparently simple, is in fact very articulated in its interconnections, which include regulator chains, single-input motifs, and multiple feedback loops, generated by the combination of transcriptional activators and repressors.

In order to gain information about the network dynamics, we analysed the transcriptional response of network genes after two different perturbation

strategies: performing a single perturbation and measuring mRNA changes at different time points, or performing multiple perturbations and collecting mRNA measurements at steady state. Data are presented in Chapter 4.

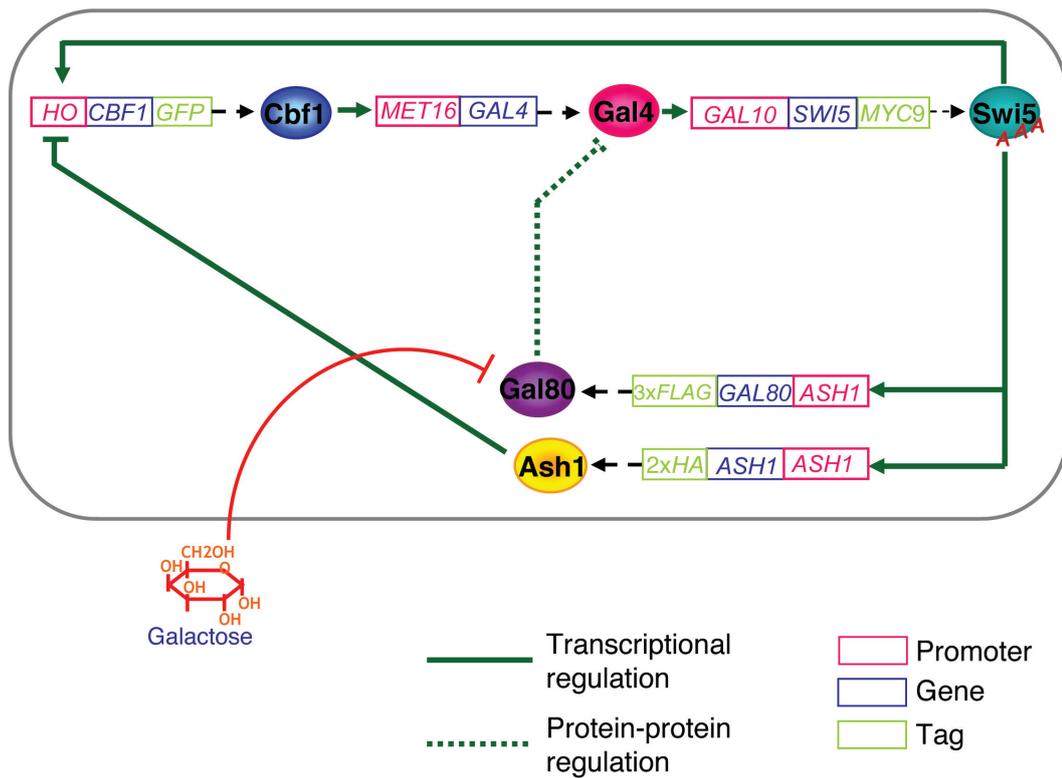


Figure 3.1: **Construction of IRMA, a synthetic network in yeast.** Schematic diagram of the synthetic gene network is represented. New transcriptional units (rectangles) were built by assembling promoters (red) with non-self coding sequences (blue). Genes were tagged at the 3' end with the specified sequences (green). Each cassette encodes for a protein (represented as a circle) regulating the transcription of another gene in the network (solid green lines). The resulting network, IRMA, is fully active when cells are grown in presence of galactose, while it is inhibited by the Gal80-Gal4 interaction in presence of glucose.

3.2 Construction of the synthetic network

3.2.1 Choice of model organism

We chose as model organism yeast *Saccharomyces cerevisiae* because it is the simplest eukaryote and it shares both transcriptional machinery structure and gene transcription mechanisms with higher eukaryotes. Considering basic biological concepts, a yeast cell is more similar to a human cell than a bacterium one. The DNA is wrapped around proteins called histones to form bead-like structures called nucleosomes, and the chromosomes are sequestered in a cellular compartment called the nucleus. For these reasons, yeast is classified as a eukaryote, as are humans, flies, worms and plants. Most of what we know about eukaryotic gene regulation comes from studies of the yeast *Saccharomyces cerevisiae*. Expression of a typical eukaryotic gene is more complex than the one of a bacterial gene, because there can be different layers of control which involve the presence of nucleosomes and nuclei. Furthermore, among eukaryotes, the yeast has got other convenient features, which led us to choose it as model. This organism grows rapidly, about 20-fold faster than mammalian cells and is only 3-fold slower than *Escherichia coli*. It is unicellular and can be easily cultured and manipulated. Mutants can be selected or recognised by simple assays, and sequences in and around genes can be altered at will. The genome is completely sequenced and comprises about 6000 genes, only about 2000 more than *E. coli*.

3.2.2 Choice of the network genes

Particular care was taken in the choice of genes in order to isolate the network from cellular environment. We searched in literature and in the SGD (Saccharomyces cerevisiae Genome Database; www.yeastgenome.org) for genes, which show some essential features. In particular:

- we chose non-essential and non-redundant TF-genes, which do not present synthetic lethality and, therefore, can be knocked out without affecting yeast viability;
- we selected well-characterised promoter/TF-encoding-gene pairs, belonging to distinct and non-redundant pathways, to further minimize external feedbacks on the network due to pathway crosstalk;
- we chose promoters for which a single transcription factor (TF) is sufficient and essential to activate transcription. Thus, by removing the endogenous TF, we maximally reduced influences from the cellular environment on each promoter.

Specifically, we selected as activators and repressors encoding genes: *SWI5*, *ASH1*, *CBF1*, *GAL4* and *GAL80*; as promoter genes: *HO*, *ASH1*, *MET16* and *GAL10* (Figure 3.1).

3.2.3 Choice of network motifs and resulting topology

In order to obtain a good benchmark, we aimed at constructing a synthetic network that captures the behaviour of larger eukaryotic gene networks on

smaller scale, and which includes also a variety of regulatory interactions.

In particular, the topology includes regulatory motifs (see Chapter 2) which are peculiar of yeast.

The network, depicted in Figure 3.1, is organized in such a way that each gene controls transcription of at least another gene in the network. We decided to use the Regulator Chain (Cbf1-Gal4-Swi5 regulators) and the Single Input (Swi5 which activates three promoters) motifs in order to have a sequence of transcriptional events, which can be separately analysed in time. We added to them both a positive (Swi5 activates *HO* transcription thus closing the circuit) and a negative transcriptional feedback loop (Ash1 represses *HO* transcription), thus obtaining a Multi Component Loop, with the aim of enriching the dynamic behaviour of the network.

Finally, in order to provide the circuit of a “switch”, we also used a negative feedback loop composed of a protein-protein interaction (Gal80-Gal4) that can “turn-off” the system in response to an external stimulus (depletion of galactose from the culturing medium). Galactose activates the *GAL10* promoter, cloned upstream of *SWI5* in the network, and it is able to activate transcription of all the five network genes. In presence of a non inducing medium as glucose, Gal4 is inactive because of the binding of the Gal80 repressor to its activation domain through the formation of a complex, preventing interaction of the transcription machinery [108]. In the presence of galactose, Gal4 activator binds to the multiple *UAS_{GAL}* elements in the promoter and leads to activation of transcription. Relief of inhibition by Gal80 is dependent on a functional Gal3 protein. Biochemical studies have shown

that Gal3 interacts with Gal80, and it is this interaction that is sensitive to the presence of galactose [108]. A diagram of the mechanism is shown in Figure 3.2.

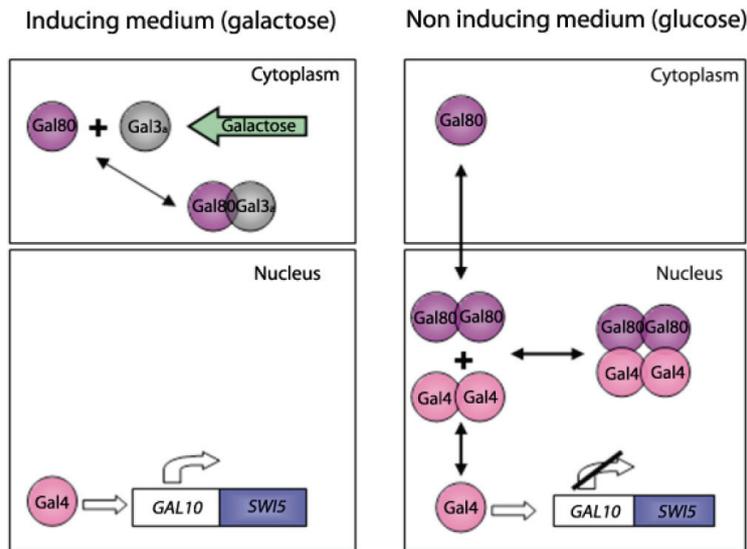


Figure 3.2: **Schematic representation of the galactose pathway regulating the *GAL10* promoter.** When the galactose medium is present, the activated Gal3 alters the free concentration of Gal80 through sequestration in the cytoplasm, thus relieving its inhibition on Gal4. In presence of glucose, the dimerized form of Gal80 directly to the Gal4 dimer.

The system can be induced also by another external signal, methionine, which regulates the activity of the *MET16* promoter. Methionine modulates the expression of all the *MET* genes by affecting the formation of the Cbf1-Met4-Met28 transcriptional complex [62]. High levels of methionine increase the ubiquitination and the subsequent degradation of the activator Met4, indeed inhibiting the transcription [17]. Thus, *MET16* expression is completely turned off in the presence of high methionine levels and, even at intermediate methionine levels, its transcription appears to be strongly decreased.

3.3 Testing *in vivo* the network response to the inducers.

3.3.1 Switching the network on by injecting galactose

At first, we tested transcription of network genes upon culturing cells in presence of galactose or glucose. In order to easily monitor the possibility of switching on the network by culturing the cells in galactose, a fluorescent tag (GFP, Green Fluorescent Protein) was cloned at the 3' end of the *CBF1* ORF (Figure 3.1). Living yeast cells grown with different carbon sources (galactose or glucose) were analysed by fluorescent microscopy. As shown in Figure 3.3, positive green cells were visualized only when IRMA was cultured in galactose-containing medium.

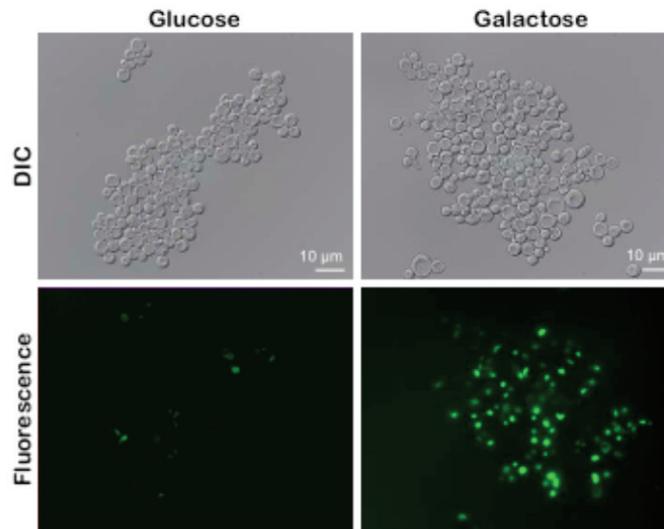


Figure 3.3: **Galactose triggers activation of IRMA synthetic network.** Live imaging of IRMA cells grown in glucose and galactose containing medium. Scale bar, 10 μ m; 63X magnification.

3.3.2 Inducibility of the network using methionine

In Figure 3.4, we show *in vivo* data (from both semi-quantitative and quantitative real-time RT-PCR) representing the expression levels of the *MET* genes, including *MET16*, when yeast cells are grown in the presence of low ($10\mu M$) or high ($1000\mu M$) methionine concentration. The levels are compared with the standard yeast growing condition complete medium (YPD), which contains an intermediate concentration of methionine ($140\mu M$) and thus show an intermediate level of *MET* genes expression. *MET16* expression is tightly regulated by methionine concentrations: it is completely turned off in the presence of high methionine levels and, even at intermediate methionine levels (the control condition), its transcription appears to be strongly decreased.

In Figure 3.5, we show the transcription levels of the genes of IRMA at steady state upon culturing cells in the presence of different concentrations of methionine, both in glucose and in galactose containing medium. Even in the presence of glucose (network off in the control standard growing condition YEP, methionine= $140\mu M$), network genes are activated in low methionine containing medium, and reach the same expression levels that they have in the cells grown in galactose (network on in YEP). Thus, the increased *GAL4* expression, due to *MET16* activation after the removal of methionine, turns on all the network genes, while addition of methionine inhibits them, independently from galactose.

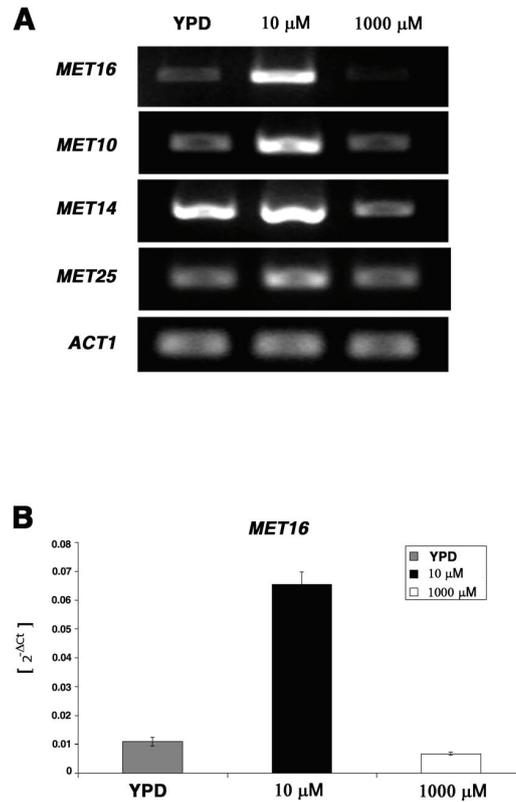


Figure 3.4: **Expression of *MET* genes in wild type yeast cells.** *MET* genes regulated by Cbf1 are transcriptionally activated in the presence of low levels of methionine (10 μ m) while they are repressed at high methionine concentrations (1000 μ m). Semi-quantitative (A) and quantitative (B) RT-PCR (normalization against *ACT1* gene) of *MET* genes were performed on total RNA extracted from yeast cells grown in the standard complete medium YPD (140 μ m of methionine) and at two different methionine concentrations.

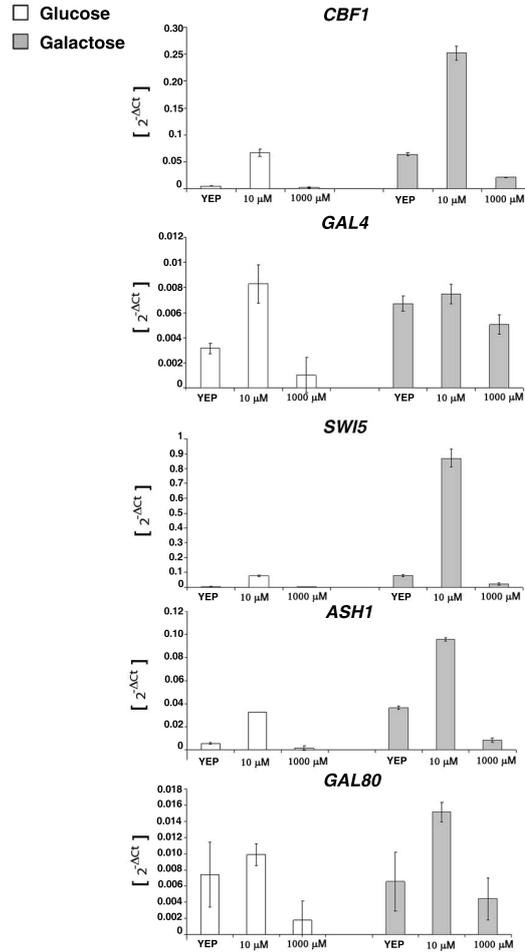


Figure 3.5: **Methionine modulates IRMA genes expression.** Expression levels of IRMA genes at different methionine concentrations in glucose (white bars) or in galactose/raffinose (grey bars). The control is the standard complete medium, YEP, which contains 140 μM of methionine. Data represent the $2^{-\Delta Ct}$ (mean of two experiments \pm Standard Error).

3.4 Discussion

In this Chapter, we presented the synthetic network we built in yeast *Saccharomyces cerevisiae*, describing its biological features. The design principles, aimed at constructing a good benchmarking tool, were illustrated. In the next Chapter, we will detail the mathematical modelling of the network through all the iterative steps and, consequently, the model-guided experimental design and *in vivo* results.

Chapter 4

Derivation, identification and validation of the mathematical model of IRMA network.

In this Chapter we will detail the mathematical modelling of the synthetic biological pathway network described in Chapter 3. We will illustrate all the steps required: model derivation, experimental design, parameter identification and model validation (see Chapter 2).

In order to build a model able to correctly predict the dynamical changes in the mRNA concentrations of the five network genes following both internal and external perturbations (i.e. gene over-expression, galactose addition, etc.), we adopted a gray-box approach [83], described in Chapter 2, following all the steps mentioned therein. For step (i) (derivation of the model equations), we used a differential equations (DEs) based approach. The task was

challenging since, to our knowledge, up to now quantitative DEs mathematical models have been developed for synthetic networks composed of a smaller number of genes than IRMA (e. g. [39], [32], [107], [60], [101]). Regarding the identifiability issue implicated in the step (ii) (parameters identification), we adopted the novel approach proposed by Raue and colleagues (see [90]), able to deal with non-linear models with an high number of parameters. This approach exploits the profile likelihood and is able to detect both structural and practical non-identifiable parameters. For the parameters identification (step (ii)), in order to cope with the high number of unknown quantities, the noise of experimental data and the presence of non-linear aspects in the optimisation procedure, we used an *ad hoc* designed Hybrid Genetic Algorithm (see Appendix B for further details). Finally, for the model validation (step (iii)), we tested the predictions of the model against data not used for the parameters identification. Of note, the identified parameters of the mathematical models are reported in Appendix A, the details about the algorithm used for the identification in Appendix B, the MATLAB files used for simulations in Appendix C and the experimental procedures of the *in vivo* experiments in Appendix D.

The results presented in this Chapter have been partly described in [73].

4.1 Derivation of model equations: step (i).

Model A

For each species in the network represented in Figure 3.1, i.e. each mRNA (italic capital letters) and correspondent protein concentration (roman small letters), we wrote one equation which expresses its change in time as the result of production and degradation:

$$\frac{d[CBF1]}{dt} = \alpha_1 + v_1 H^{+-}([Swi5], [Ash1]; k_1, k_2, h_1, h_2) - d_1[CBF1] \quad (4.1)$$

$$\frac{d[Cbf1]}{dt} = \beta_1[CBF1] - d_2[Cbf1], \quad (4.2)$$

$$\frac{d[GAL4]}{dt} = \alpha_2 + v_2 H^+([Cbf1]; k_3, h_3) - d_3[GAL4], \quad (4.3)$$

$$\frac{d[Gal4]}{dt} = \beta_2[GAL4] - d_4[Gal4], \quad (4.4)$$

$$\frac{d[SWI5]}{dt} = \alpha_3 + v_3 H^+([Gal4^{free}]; k_4, h_4) - d_5[SWI5], \quad (4.5)$$

$$\frac{d[Swi5]}{dt} = \beta_3[SWI5] - d_6[Swi5], \quad (4.6)$$

$$\frac{d[GAL80]}{dt} = \alpha_4 + v_4 H^+([Swi5]; k_5, h_5) - d_7[GAL80], \quad (4.7)$$

$$\frac{d[Gal80]}{dt} = \beta_4[GAL80] - d_8[Gal80], \quad (4.8)$$

$$\frac{d[ASH1]}{dt} = \alpha_5 + v_5 H^+([Swi5]; k_6, h_6) - d_9[ASH1], \quad (4.9)$$

$$\frac{d[Ash1]}{dt} = \beta_5[ASH1] - d_{10}[Ash1]. \quad (4.10)$$

The first two terms, on the right-hand side of the mRNA equations, represent the production, where α are the basal transcription rates; v are the maximal transcription rates modulated by the Hill functions, $H^+(y; k, h) = \frac{y^h}{y^h + k^h}$,

$H^-(z; k, h) = \frac{k^h}{y^h + k^h}$ and $H^{+-} = H^+(y; k, h)(\cdot, +)H^-(z; k_1, h_1)$, modelling transcriptional activation, repression or a combination of the two, respectively; $(\cdot, +)$ indicates that we can either sum or multiply the Hill functions in the case of multiple regulation; y and z represent transcription factor levels, h are the Hill coefficients (pure numbers that refer to the cooperativity of the activation binding reaction) and k are the Michaelis-Menten constants, equal to the amount of transcription factor needed to reach half maximal activation (or repression). For protein equations, the production rates are β , i.e. the maximal translation rates. Degradations of mRNAs and proteins are represented by d , i.e. the degradation constants. Gal4^{free} in equation (4.5) depends on the interactions of the galactose pathway with the network genes. In the model, the concentrations and the Michaelis-Menten parameters k are reported in arbitrary units [*a.u.*], the basal activities α in [*a.u. min*⁻¹], the maximal transcription rates v in [*a.u. min*⁻¹], the translation rates β in [*min*⁻¹], the degradation constants d in [*min*⁻¹].

When writing the above model, we made the following assumptions: [**A1**] the *transcriptional activity* of each promoter is leaky (α); [**A2**] the degradation kinetics of both mRNAs and proteins are first-order; [**A3**] the protein production terms are proportional to the corresponding mRNA concentrations; [**A4**] the transcriptional *activation-repression* of each promoter by a transcription factor can be modelled as a Hill function [56] and the *HO* promoter driving the expression of *CBF1* can be modelled either by adding the H^+ and H^- functions (i.e. the promoter is activated by *SWI5 OR* repressed by *ASH1*), or by multiplying them (i.e. the promoter is activated by *SWI5*

AND repressed by *ASH1*) (see [2] and [70]). We chose between these two forms only during step (iii) of the modelling process, as described later.

In order to define the Gal4^{free} term in eq. (4.5), we needed to describe the effect of the galactose pathway on the network dynamics. The biological mechanism is shown in Figure 4.1 (A). The concentration of Gal4^{free} is the amount of Gal4 protein that is not involved in the formation of the protein-protein complex with Gal80 and hence activates the *GAL10* promoter driving *SWI5* expression. In the literature, very detailed models of the galactose pathway have been presented ([10], [111]). We decided to simplify such paradigms and assumed ([A5]) that Gal80 directly binds to galactose ([GAL], the input of our model) in galactose growing condition, while Gal4 and Gal80 form the complex Gal4Gal80, when the yeast is grown in glucose (Figure 4.1 (B)). Under this assumption, the simplified physical mechanism can be described by the mass balance laws:

$$[\text{Gal4}] = [\text{Gal4}^{free}] + [\text{Gal4Gal80}], \quad (4.11)$$

and

$$[\text{Gal80}] = [\text{Gal80}^{free}] + [\text{Gal4Gal80}] + [\text{GALGal80}], \quad (4.12)$$

where $[\text{Gal4Gal80}]$ and $[\text{GALGal80}]$ indicate the concentrations of the complexes. The rates of these complexes can be modelled assuming reversible

reactions for them ($A + B \rightleftharpoons AB$), i.e.

$$\frac{d[\text{Gal4Gal80}]}{dt} = K_1[\text{Gal4}^{free}][\text{Gal80}^{free}] - K_2[\text{Gal4Gal80}], \quad (4.13)$$

$$\frac{d[\text{GALGal80}]}{dt} = K_3[\text{GAL}][\text{Gal80}^{free}] - K_4[\text{GALGal80}], \quad (4.14)$$

with K being rate constants (K_1 is measured in [$a.u.^{-1} min^{-1}$], K_2 and K_4 in [min^{-1}], K_3 in [$nM^{-1} min^{-1}$] if the concentration of galactose, [GAL], is measured in [nM]).

The full model is described by eqs. (4.1)-(4.10) together with eqs (4.13), (4.14) and consists of 12 equations and 41 parameters (Model A).

4.2 Simplification of the complete model: Model B

If we assume that the time scale for the protein synthesis rate (including translocation and post-translational modifications) is much smaller than the time scale for the mRNA synthesis rate [49], the protein concentrations are monotonically increasing functions of their corresponding mRNA concentrations at any time. Thus, by considering mRNA transcription and translation as a single step of synthesis for the five genes of the network ([A6]), equations (4.2), (4.4), (4.6), (4.8) and (4.10) can be removed together with the associated 10 unknown parameters, leading to a *simplified non-linear model* (Model B) of IRMA (degradation constants renumbered and Hill functions in explicit form):

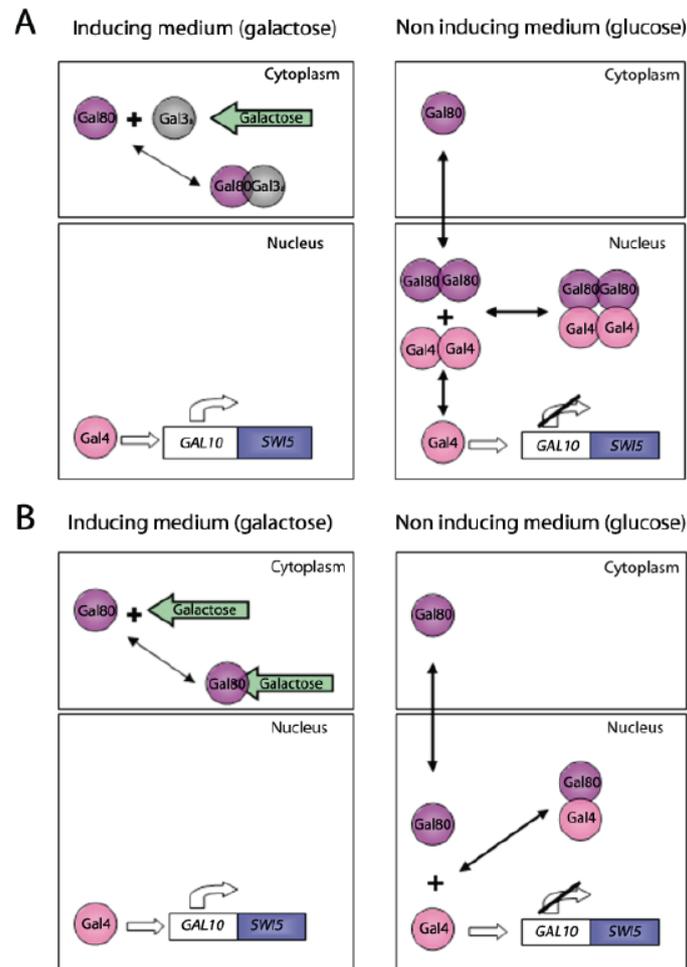


Figure 4.1: **Interactions between the galactose pathway and IRMA genes.** (A) Schematic representation of galactose pathway. When the galactose medium is present, the activated Gal3 alters the free concentration of Gal80 through sequestration in the cytoplasm, thus relieving its inhibition on Gal4. In presence of glucose, the dimerized form of Gal80 directly to the Gal4 dimer. (B) Simplified representation of the galactose induced switch described by models A and B. Here it's assumed that Gal80 directly binds to galactose when the network is on while Gal4 and Gal80 form the complex Gal4Gal80 without any prior dimerization when the network is on.

$$\frac{d[CBF1]}{dt} = \alpha_1 + v_1 \left(\frac{[SWI5]^{h_1}}{k_1^{h_1} + [SWI5]^{h_1}} \right) (\cdot, +) \left(\frac{k_2^{h_2}}{k_2^{h_2} + [ASH1]^{h_2}} \right) - d_1[CBF1] \quad (4.15)$$

$$\frac{d[[GAL4]]}{dt} = \alpha_2 + v_2 \left(\frac{[CBF1]^{h_3}}{k_3^{h_3} + [CBF1]^{h_3}} \right) - d_2[GAL4], \quad (4.16)$$

$$\frac{d[SWI5]}{dt} = \alpha_3 + v_3 \left(\frac{([GAL4] - [Gal4Gal80])^{h_4}}{k_4^{h_4} + ([GAL4] - [Gal4Gal80])^{h_4}} \right) - d_3[SWI5], \quad (4.17)$$

$$\frac{d[GAL80]}{dt} = \alpha_4 + v_4 \left(\frac{[SWI5]^{h_5}}{k_5^{h_5} + [SWI5]^{h_5}} \right) - d_4[GAL80], \quad (4.18)$$

$$\frac{d[ASH1]}{dt} = \alpha_5 + v_5 \left(\frac{[SWI5]^{h_6}}{k_6^{h_6} + [SWI5]^{h_6}} \right) - d_5[ASH1], \quad (4.19)$$

$$\begin{aligned} \frac{d[Gal4Gal80]}{dt} &= K_1([GAL4] - [Gal4Gal80])([GAL80] - [Gal4Gal80] - [GALGal80]) + \\ &\quad - K_2[Gal4Gal80] \end{aligned} \quad (4.20)$$

$$\frac{d[GALGal80]}{dt} = K_3[GAL]([GAL80] - [Gal4Gal80] - [GALGal80]) - K_4[GALGal80], \quad (4.21)$$

where $(\cdot, +)$ in (4.15) indicates that, according to assumption [A4], the multiple regulation of *CBF1* can be modelled either as an AND or an OR logic gate. Note the complexes equations (4.20), (4.21) were derived from the rate equations (4.13), (4.14) by substituting the expressions of $[Gal4^{free}]$ and $[Gal80^{free}]$ derived from the mass balance laws (4.11) and (4.12) under the assumption [A6]. In what follows, we will explore both possibilities, showing how comparison of the model predictions with the experimental data motivated the final choice. Equations (4.20) and (4.21) were obtained by substituting eqs. (4.11) and (4.12) in eqs. (4.13) and (4.14), under the assumption of proportionality between the protein levels of Gal4 and Gal80

and the corresponding mRNAs. Model B consists of 7 differential equations ((4.15)-(4.21)) and contains 31 unknown parameters.

4.2.1 Identification of model parameters: step (ii)

In order to reduce the number of unknown parameters in Model B, we assumed that: [A7] all the promoters have null basal activity and unitary transcription rate, so that $\alpha = 0$ and $v = 1$; [A8] for the cooperativity coefficients h in the Hill functions we can consider only two options: set them all to 1 (*monomers approach*), or set all to 1 with the exception of h_3 and h_4 , which are equal to 2 (*dimers approach*) in order to model the higher cooperativity of Cbf1 and Gal4 respectively on the *MET16* promoter and the *GAL10* promoter [50] [42]. Parameters K_1 , K_2 , K_3 and K_4 in equations (4.20), (4.21) were fixed *a priori* from literature [3]. Their values are reported in Table A.1. The remaining 11 parameters were unknown and needed to be estimated from experimental data. To this end, we collected data of mRNAs expression levels during a time course experiment, by shifting cells from glucose to galactose “switch-on” experiment) as described in Appendix D. There are four versions of Model B due to assumptions [A4] (AND/OR regulation of the *HO* promoter driving *CBF1* expression) and [A8] (dimers versus monomers). We labelled the four different versions of Model B as B1 (AND/-Monomers), B2 (OR/monomers), B3 (AND/dimers) and B4 (OR/dimers). Identifiability analysis showed that all the 11 unknown parameters of Models B1, B2, B3 and B4 are structurally, but not practically, identifiable in the sense of Raue [90]. Thus, the non-identifiability does not arise from incom-

plete observation of the internal model states or redundant parametrisation, but from the noisy nature and/or from the insufficient amount of experimental data. This makes the qualitative identification procedure we used the only viable option. We proceeded with the identification in order to evaluate the descriptive performance of the models and to start discriminating between different modelling possibilities (see Appendix B for details about the identification procedure). This was done by comparing *in silico* and *in vivo* data for each of the four B models, using direct inspection and comparison of the corresponding values of the cost function J (see Appendix B).

The identified parameters are listed in Table A.1. Results for Models B1 and B2 are shown in Figures 4.2 (A) and 4.3 (A), respectively. Model B1 has a lower cost function ($J=4.37$) than the value obtained with Model B2 ($J=7.951$). Hence, modelling the multiple regulation of *CBF1* as a product (AND) seems to capture more accurately the dynamics of the *HO* promoter. Results for the Models B3 and B4 are shown in 4.3 (B) and (C), respectively. Model B3 (AND/dimers) has a cost function $J=2.83$, much lower than the other three models, and, thus, it was selected for the next step.

4.2.2 Validation of model predictive performance: step (iii)

In order to assess the predictive ability of Model B3, i.e. if the model is able to predict the behaviour of the network to new perturbations, we measured the gene expression response of the five network genes following exogenous over-expression of each of the five genes under the control of a strong constitutive

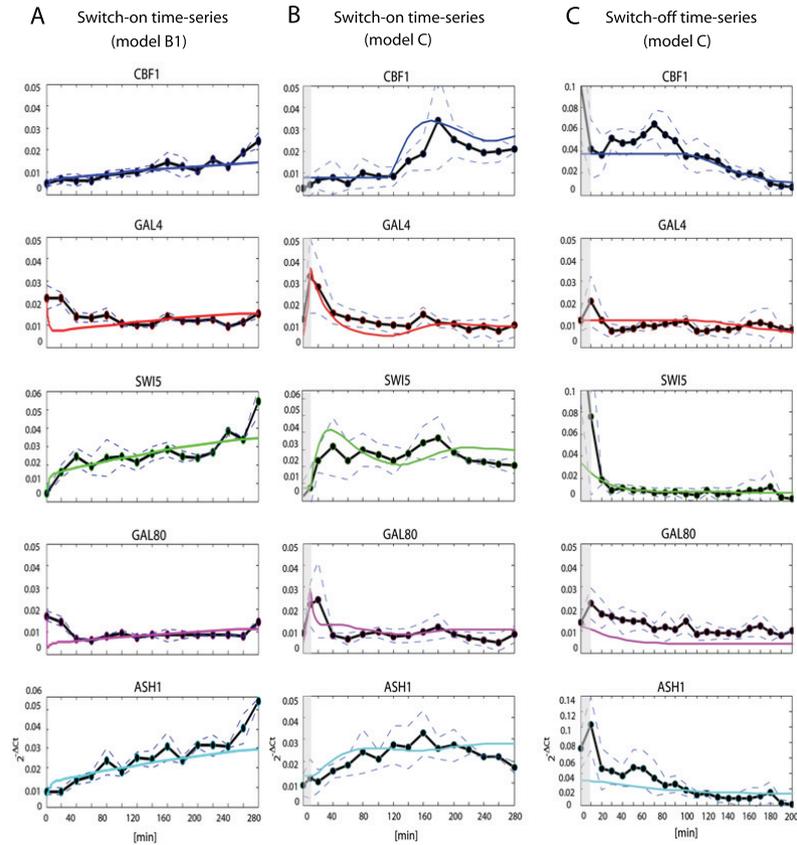


Figure 4.2: **Identification and validation results on time-series data. Models B1 and C.** Circles represent average expression data for each of the IRMA genes at different time points. Dashed lines represent standard errors. Continuous colored lines represent *in silico* data. (A) Identification results of Model B1 on the preliminary 5 hours “switch-on” time-series (average of 4 time-series). (B) Identification results of Model C on the new 5 hours “switch-on” data-set (average of 5 time-series). (C) Validation of Model C on the 3 hours “switch-off” data-set (average of 4 time-series).

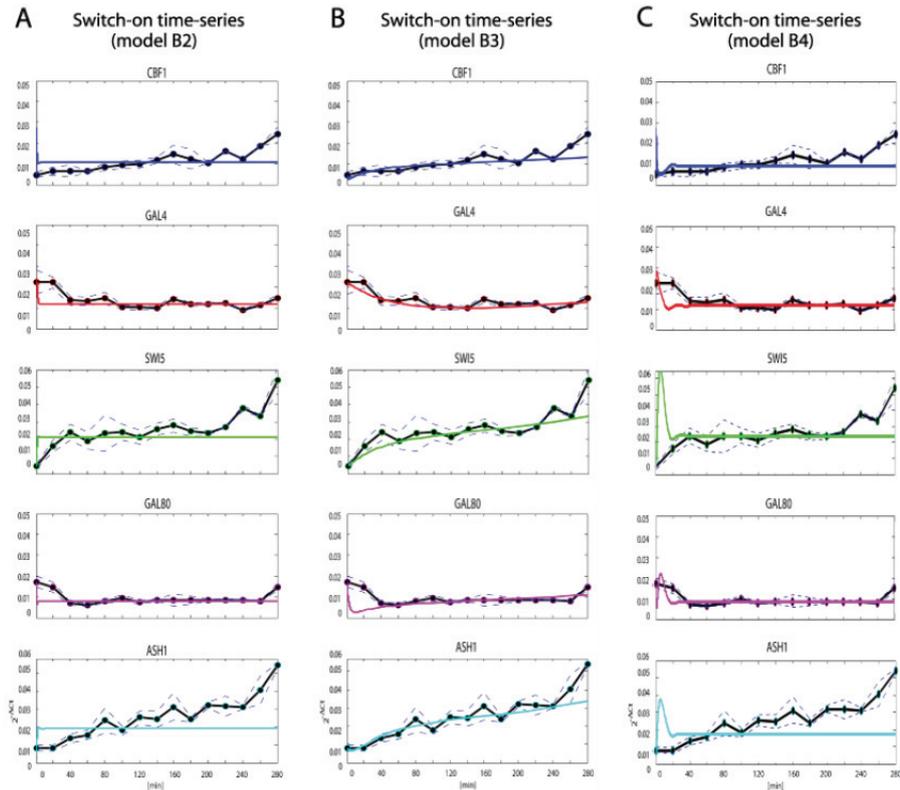


Figure 4.3: **Identification results on time-series data. Models B2, B3 and B4.** Circles represent average expression data for each of the IRMA genes at different time points. Dashed lines represent standard errors. Continuous colored lines represent *in silico* data. (A) Identification results of the model B2 on the preliminary “switch-on” time-series. (B) Identification results of the model B3 on the preliminary “switch-on” time-series. (C) Identification results of the model B4 on the preliminary “switch-on” data-set.

promoter, as described in Appendix D. Such over-expression experiments were performed both in glucose and in galactose. We will refer to these two experimental data-sets as the “Galactose steady-state” and “Glucose steady-state” (Figure 4.4 (A), (C)).

We performed *in silico* the over-expression experiments: as initial conditions, we used the steady states predicted by the model in unperturbed conditions (either in glucose or in galactose), and in addition we applied a constant input, corresponding to the gene overexpression, to each of the five equations, in order to match the experimental data of the perturbed gene. We collected the predicted steady state of the other genes, and compared it with data.

It resulted that Model B3, despite its good descriptive performance, has a very poor predictive power (simulations not shown).

Therefore, we tested the predictive performance also for Model B1 (the second best as regards descriptive performance). Results are shown in Figure 4.4 (B), (D). Model B1 is able to partly describe and predict the network behaviour. There are still some major pitfalls: (a) two quantities (the concentrations of the two complexes) are present in the model, but cannot be measured experimentally. They were introduced by assuming a simplified mechanism for the interactions between the medium, Gal4 and Gal80 (assumption [A5] and Figure 4.1), but they are not physically consistent, and thus not measurable. (b) The “switch-on” data-set shows almost monotonic dynamics for the genes of IRMA, regardless of its complex topology. This data-set is an average of four independent experiments, three lasting 3 hours, and just one lasting 5 hours. Moreover, in such data, the early dynamic be-

haviour of the genes *GAL4* and *GAL80* is highly unexpected. We should observe an increase of all the mRNA concentrations following addition of galactose (“switch-on”), whereas *GAL4* and *GAL80* show a decrease during the initial 40 minutes, which Model B is unable to reproduce.

This modelling stage indicates that Model B has to be refined, and that new experiments are needed in order both to obtain a better characterisation of the dynamics of the synthetic network and to try to cope with the practical non-identifiability of the parameters.

4.3 Additional experimental investigation

We performed one additional 5 hours “switch-on” time-series (see Appendix D), this time including as the first point of the time-series the expression level of the network genes after growing cells overnight in glucose, just before shifting them from glucose to galactose [15]. The second point, taken after 10 minutes, is measured just after the shift has occurred and is equivalent to the first point of the previous time-series. The addition of this point to the data is fundamental to clarify the inconsistency in the early dynamics of *GAL4* and *GAL80*. The new averaged data-set (Figure 4.2 (B)) shows that the standard washing steps, needed to shift cells from glucose medium to the fresh new galactose-containing medium, induce a transient increase in mRNA levels of *GAL4* and *GAL80* (Figure 4.2 (B), grey bars). This effect is not dependent on galactose addition, but uniquely on the washing steps [15], and it is probably due to the transient deprivation of carbon source

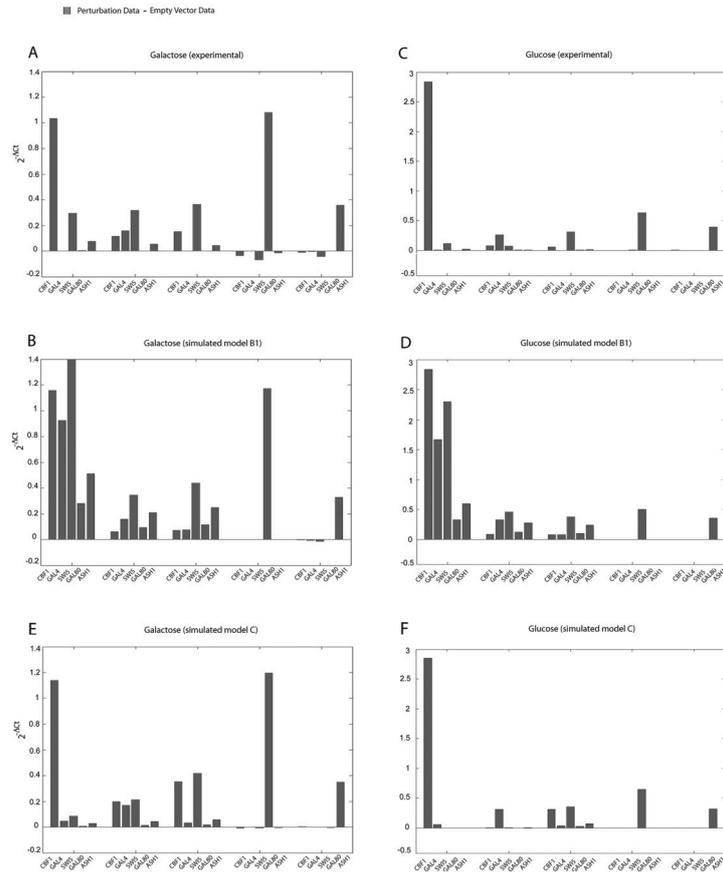


Figure 4.4: **Experimental and simulated over-expression experiments. Models B1 and C.** (A), (C) Difference between *in vivo* expression levels of IRMA genes after over-expression of each gene from the constitutive *GPD* promoter and levels after transformation of the empty vector. IRMA cells were transformed with each of the constructs containing one of the five genes or with the empty vector. At least three different colonies were grown in glucose (C) and in galactose-raffinose (A) up to the steady-state levels of gene expression. Quantitative PCR data are represented as $2^{-\Delta Ct}$ (average data from different colonies). (B),(D) *In silico* data obtained by simulating the over-expression of each gene with Model B1. (E), (F) *In silico* data obtained by simulating the over-expression of each gene with Model C.

during washing, which attenuates the degradation levels of *GAL4* and *GAL80* mRNAs [53].

Also, in the new averaged data-set, the activation of *CBF1* appears to be delayed with respect to the other Swi5 targets, respectively *GAL80* and *ASH1*. Such delay, not evident from the preliminary data-set, is physically due to the sequential recruitment of chromatin modifying complexes to the *HO* promoter, which follows binding of Swi5 ([11], [20]).

We performed four additional experiments, shifting cells from galactose to glucose, thus switching off gene expression in the network, as described in Appendix D. The averaged time-series data-set (Figure 4.2 (C)) was used for a further validation of the model predictive performance. We will refer to this data-set as the “switch-off” data-set.

4.4 Model refinement: Model C (step (i))

At this stage, we had to properly refine the model both to be able to capture the new features highlighted by the new data-set and to remove unsuitable model complexity. First of all, we made the following extra modelling assumptions: [A9] a fix time delay, τ , equal to 100 minutes, is added in the activation of the *HO* promoter by Swi5; [A10] a transient decrease in the mRNA degradation of *GAL4* and *GAL80* of value $\Delta\beta_1$ and $\Delta\beta_2$ ($[min^{-1}]$) is added for an interval of 10 minutes to describe the effect of the washing steps (Δ represents the transient duration of the washing effect).

Secondly, in order to remove from the model the unmeasured complexes con-

centrations describing the effects of galactose on the network, we considered two possible approaches: (1) to take the quasi steady-state approximation of the protein complexes dynamics (i.e. by setting the left-hand sides of (4.13) and (4.14) to 0); (2) to consider a new phenomenological non-linear function describing the effect of galactose. In the first case, steady-state approximation leads to the presence of an algebraic constraint thus turning the problem into a differential algebraic model with delays (DDAEs). This kind of problems are particularly cumbersome to solve and analyse from a mathematical viewpoint (see [61] for further details). To avoid this, we proceeded by finding a simple but effective phenomenological non-linear function to model the effect of the galactose pathway on the dynamics of *SWI5*, which is regulated by the *GAL10* promoter.

We assumed [A11] that the protein-protein interaction between Gal80 and Gal4 can be modelled as a direct inhibition of *GAL80* on the promoter of *SWI5*, and that the strength of such inhibition depends on the medium (strong inhibition in glucose, weak inhibition in galactose). Actually we assumed that the *GAL10* promoter is activated by *GAL4* and non-competitively inhibited by *GAL80* [19].

The resulting phenomenological DDEs model (Model C), derived from Model B1, is:

$$\frac{d[CBF1]}{dt} = \alpha_1 + v_1 \left(\frac{[SWI5(t - \tau)]^{h_1}}{k_1^{h_1} + [SWI5(t - \tau)]^{h_1}} \right) \cdot \left(\frac{k_2^{h_2}}{k_2^{h_2} + [ASH1]^{h_2}} \right) - d_1[CBF1], \quad (4.22)$$

$$\frac{d[GAL4]}{dt} = \alpha_2 + v_2 \left(\frac{[CBF1]^{h_3}}{k_3^{h_3} + [CBF1]^{h_3}} \right) - (d_2 - \Delta\beta_1)[GAL4], \quad (4.23)$$

$$\frac{d[SWI5]}{dt} = \alpha_3 + v_3 \left(\frac{[GAL4]^{h_4}}{(k_4^{h_4} + ([GAL4]^{h_4})(1 + \frac{[GAL80]^{h_7}}{\hat{\gamma}^{h_7}}))} \right) - d_3[SWI5], \quad (4.24)$$

$$\frac{d[GAL80]}{dt} = \alpha_4 + v_4 \left(\frac{[SWI5]^{h_5}}{k_5^{h_5} + [SWI5]^{h_5}} \right) - (d_4 - \Delta\beta_2)[GAL80], \quad (4.25)$$

$$\frac{d[ASH1]}{dt} = \alpha_5 + v_5 \left(\frac{[SWI5]^{h_6}}{k_6^{h_6} + [SWI5]^{h_6}} \right) - d_5[ASH1], \quad (4.26)$$

which consists of only 5 equations without any additional constraint.

The constant $\hat{\gamma}$ in (4.24) is the Michaelis-Menten coefficient of the phenomenological description of the inhibition of *GAL80*, which is assumed to be dependent on the medium (we use the symbol $\hat{\gamma}$ to indicate medium-dependent quantities). This phenomenological DDEs model consists of 5 differential equations ((4.22)-(4.26)) and 31 unknown parameters.

4.4.1 Identification of the model parameters and validation of its predictive performance: steps (ii) and (iii)

We set all of the Hill coefficients to 1 (monomers). For the identification of the remaining parameters, we used again the “switch-on” data-set, but this time using as initial values the simulated steady-state mRNA levels in glucose. The

identifiability analysis showed that all the unknown parameters of Model C are again structurally identifiable, but not practically. Identification results are shown in Figure 2 (B) and the inferred parameters in Table A.1. The model captures the delay in *CBF1* activation and the small variations of *GAL4* and *GAL80*.

In order to validate the model predictive performance, we used again the “Glucose steady-state” and “Galactose steady-state” over-expression experiments, and compared them with their *in silico* counterparts by simulating the over-expression of each of the five genes using Model C (Figure 4.4 (E), (F)), as described in Section 4.2.2.

We further validated the predictive performance of the Model C against the “switch-off” time-series by simulating *in silico* the “switch-off” experiment (i.e. setting the medium-dependent parameters to their values in glucose and starting the simulation from the steady-state equilibrium in galactose) (Figure 4.2 (C)).

Model C has good descriptive and predictive performance. At this stage, it represents the best compromise between model complexity and performance given the experimental data-set. The model is indeed able to qualitatively predict network behaviour to new perturbations, thus achieving the aim we set for the modelling task. However, the 24 identified parameter values are likely to be different from their physical values. For example, model parameters (Table A.1) indicate that the inhibition of Ash1 on *CBF1* is so weak that can be neglected, even if in the literature it has been reported otherwise [20].

4.4.2 Experimental identification of the Hill function parameters

At this point, we needed to clarify the biological properties of the *HO* promoter by taking direct measurements of the promoter' parameters. We thus performed promoter strength experiments by measuring the transcriptional response of the promoters of *GAL10*, *MET16*, *ASH1* and *HO*, the latter when regulated by both Swi5 and Ash1. For details refer to Appendix D. Actually, we could have performed these experiments from the beginning, since the Hill functions were almost unchanged during the model refinement, with the exception of the *GAL10* and *HO* promoters modelling. However, since each experiment is costly and time consuming, we tried at each step to only perform those experiments that the mathematical modelling deemed indispensable. The need of performing promoter strength experiments arose after the identification of Model C since we did not trust the identified Hill functions parameters.

The model is now significantly improved, and the number of parameters that are practically not identifiable from the “switch-on” data-set can be significantly reduced.

For all of the promoters, we fitted the Hill function used in Model C. For each promoter, we fitted to data the equation at steady state of the gene whose expression is driven by the promoter itself. For example, in the case of *HO* promoter, the function fitted was the right-hand side of equation (4.22),

thus obtaining:

$$[CBF1] = \frac{\alpha_1}{d_1} + \frac{v_1}{d_1} \left(\frac{[SWI5]^{h_1}}{(k_1^{h_1} + [SWI5]^{h_1}) \cdot \left(1 + \frac{[ASH1]^{h_2}}{k_2^{h_2}}\right)} \right). \quad (4.27)$$

For the fitting, the hybrid genetic algorithm was used (see Appendix B). In order to identify the phenomenological law for the *GAL10* promoter in eq. (4.24), we fitted all the possible forms of the inhibition law (non-competitive, uncompetitive and competitive [19]). Uncompetitive inhibition was found to give the best fitting (data not shown). Finally, it became apparent from the new experimental data and the results of the fitting, that galactose not only weakens the inhibition of Gal80 on the *GAL10* promoter (assumption [A11] in Model C), but also allows a faster activation of the *GAL10* promoter. Moreover, in galactose such activation is possible for values of *GAL4* lower than in glucose.

The kinetic parameters that were physically estimated are given in Table A.1, while the data and the relative fitting in Figures 4.5 and 4.6.

4.5 Further model refinement: Model D (back to step (i))

To model the effect of galactose and, in particular, the behaviour of the *GAL10* promoter, Model C needed to be further refined. In particular, since galactose was found to affect all of the parameters describing the *GAL10* promoter activity, we considered two additional parameters in the model to

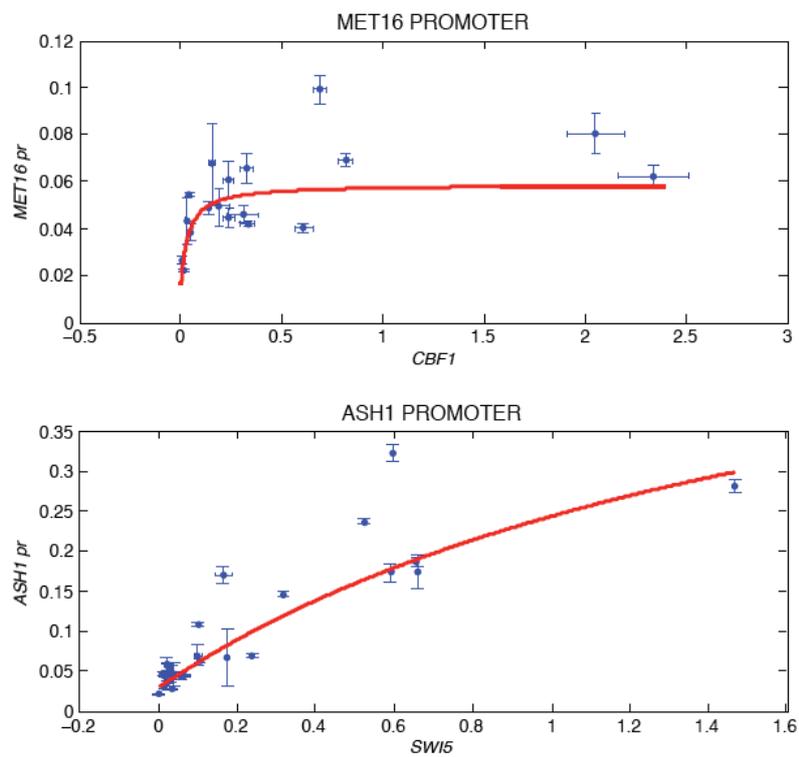


Figure 4.5: **Fitting of experimental data of promoters strength to Hill function.** Data are shown as expression values ($2^{-\Delta Ct}$). x-axis: expression of the transcription factor; y-axis: expression of the target gene.

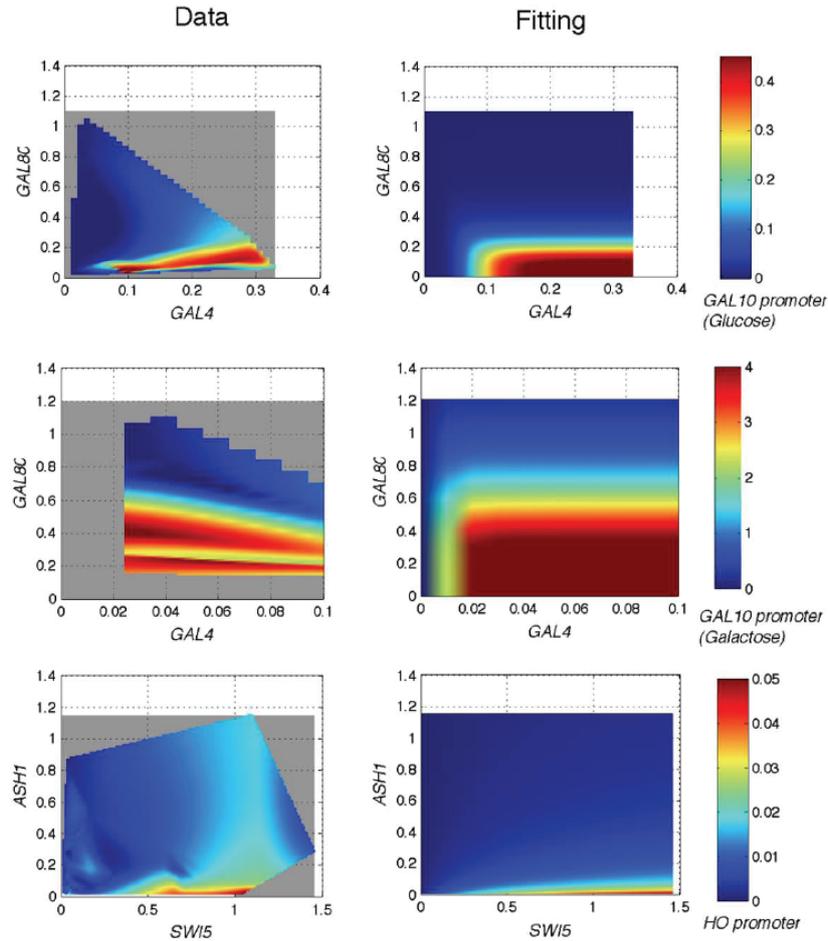


Figure 4.6: **Fitting of experimental data of promoters strength to Hill function.** Data are shown as expression values ($2^{-\Delta Ct}$). By the function *griddata* of MATLAB a surface was fitted with cubic interpolation to the promoter strength data; x-axis and y-axis: expression of the transcription factor; z-axis: expression of the target gene. On the left panel data are shown (grey area represent regions in which data are not present); on the right panel fitting results are shown.

be explicitly dependent on the medium.

Thus, we derived a new model (Model D) consisting of the equations (4.22), (4.23), (4.25), (4.26) of Model C and of the following equation for *SWI5*:

$$\frac{d[SWI5]}{dt} = \alpha_3 + \hat{v}_3 \left(\frac{[GAL4]^{h_4}}{(\hat{k}_4^{h_4} + ([GAL4]^{h_4})(1 + \frac{[GAL80]^{h_7}}{\hat{\gamma}^{h_7}}))} \right) - d_3[SWI5], \quad (4.28)$$

where the symbol $\hat{}$ indicates parameters dependent on the medium. From the analysis of data, we found that the value assumed by \hat{v}_3 in galactose is 9 times bigger than the one in glucose. Analogously, the value of \hat{k}_4 is 9 times bigger in glucose than in galactose (see Appendix A).

4.5.1 Identification of parameters and validation of model D: step (ii) and (iii)

The refined DDEs model (eqs. (4.22), (4.23), (4.25), (4.26), (4.28)) contains 33 unknown parameters. From the promoter data-set, we estimated 16 parameters, including the medium-dependent ones (Appendix A). From such data, we could not fit degradation constants, nor the washing effect parameters ($\Delta\beta_1$ and $\Delta\beta_2$). Thus, the remaining 17 parameters were evaluated from the “switch-on” experiment (Table A.1). In simulations, the initial values of mRNA concentrations were set to the steady state values predicted by the model in glucose. The *in silico* “switch-on” time-series is shown in Figure 4.7 (A). Also in this case, we tested the predictive ability of the model performing *in silico* the previously described “Glucose steady-state” and “Galactose

steady-state” over-expression experiments and the “switch-off” time-series (Figure 4.8 (B), (D) and Figure 4.7 (B)). By comparing data and simulations, it appears that Model D is quite similar to Model C, the only difference being that, this time, some of its physical parameter values have been directly measured. Now, Model D parameters confirm that the Ash1 inhibition of the *HO* promoter is indeed strong, as reported in the literature [20].

There are still discrepancies between the *in vivo* and *in silico* initial values of *CBF1*, *SWI5* and *ASH1* in the “switch-off” data-set, and in the predicted steady state of mRNA levels in galactose. We attribute them to the unmodelled effect of protein dynamics, which have been removed from the original model due to the lack of experimental measurements. In particular, we noticed that the Gal4 protein is stable [81], and therefore even a small, or transient, increase in its mRNA level is able to induce the *GAL10* promoter, regulating Swi5 in our network. Since we do not explicitly model protein dynamics, a small increase in *GAL4* mRNA cannot fully activate the *GAL10* promoter in the model and does not cause the increase in *SWI5* mRNA seen *in vivo*. In order to verify this hypothesis, we modified Model D by additionally modelling the protein level of Gal4. Thus, in the model we added the following equation for Gal4 protein (which is assumed to be linearly dependent on *GAL4* mRNA):

$$\frac{d[\text{Gal4}]}{dt} = v_{tr}[\text{GAL4}] - d_{pr}[\text{Gal4}]. \quad (4.29)$$

As a consequence, a new variable in the activation law of Swi5 has been

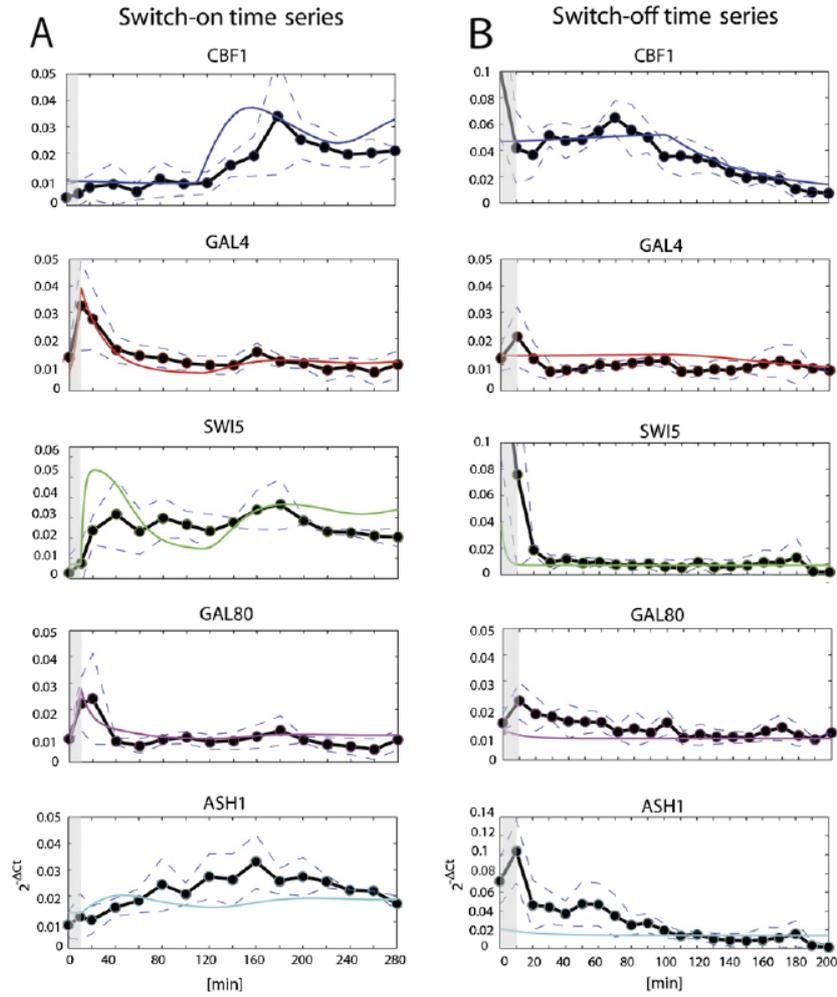


Figure 4.7: **Identification and validation results on time-series data. Model D.** Identification results on time-series data. Circles represent average expression data for each of the IRMA genes at different time points. Dashed lines represent standard errors. Continuous colored lines represent in silico data. (A) Identification results of the model D on the “switch-on” time-series (average of 5 time-series). (B) Validation results of the model D on the “switch-off” time-series (average of 4 time-series).

inserted:

$$\frac{d[SWI5]}{dt} = \alpha_3 + \hat{v}_3 \left(\frac{[\text{Gal4}]^{h_4}}{(\hat{k}_4^{h_4} + ([\text{Gal4}]^{h_4})(1 + \frac{[\text{GAL80}]^{h_7}}{\hat{\gamma}^{h_7}}))} \right) - d_3[SWI5]. \quad (4.30)$$

We fitted the parameters in equation (4.29) from the “switch-on” data-set (Table A.1). In particular, the estimated degradation rate of Gal4 protein is lower than all the other degradation rates, in accordance with the experimental results in [81]. Consequently, we slightly modified two parameters of the *GAL10* promoter (Table A.1). Note that such parameters were previously estimated from the promoter data-set, but in such experiments we measured the levels of the *GAL10* promoter depending on the mRNA and not on the protein level of Gal4. The *in silico* “switch-on” and “switch-off” time-series look almost identical to the simulations of Model D (data not shown), but the quality of the predictions of the “Glucose steady-state” and “Galactose steady-state” over-expressions is significantly improved (see Figure 4.8 (E), (F)). In particular, the increase in *SWI5* expression, due to the accumulation of Gal4 protein, is captured (e.g. Figure 4.8 (E), over-expression of *CBF1* and *GAL4*).

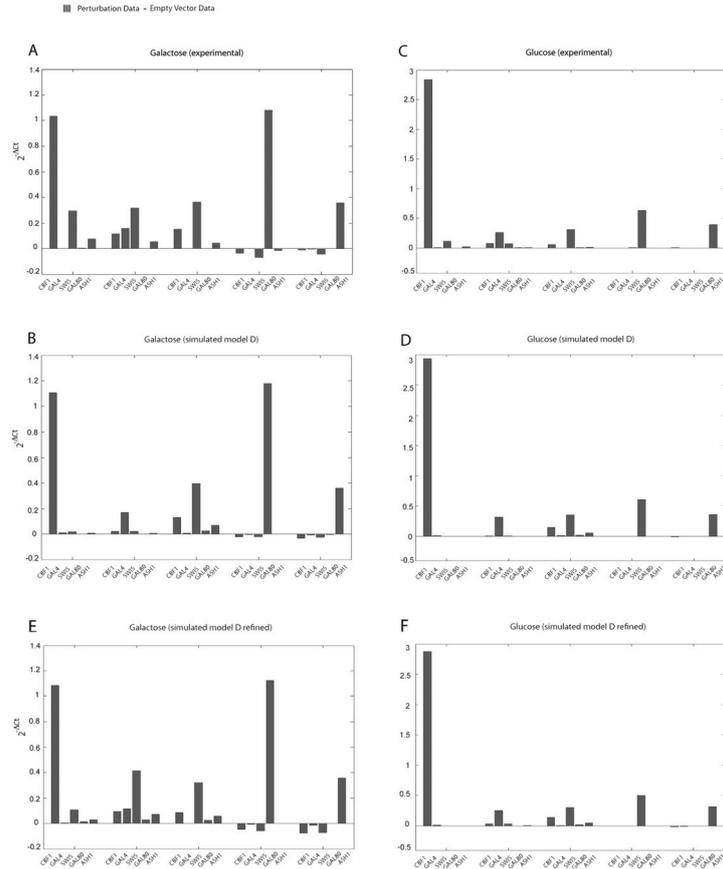


Figure 4.8: Experimental and simulated over-expression experiments. Models D and D refined.

(A), (C) Difference between *in vivo* expression levels of IRMA genes after over-expression of each gene from the constitutive *GPD* promoter and levels after transformation of the empty vector. IRMA cells were transformed with each of the constructs containing one of the five genes or with the empty vector. At least three different colonies were grown in glucose (C) and in galactose-raffinose (A) up to the steady-state levels of gene expression. Quantitative PCR data are represented as $2^{-\Delta Ct}$ (average data from different colonies). (B), (D) *In silico* data obtained by simulating the over-expression of each gene with Model D. (E), (F) *In silico* data obtained by simulating the over-expression of each gene with Model D refined.

4.6 Discussion

In this Chapter, we described in detail the steps required to build a mathematical model of a synthetic biological pathway. The whole modelling procedure is schematically described in Figure 4.9. This framework can be applied equally well to naturally occurring networks in the cell, thus transforming the drawing of a biological pathway into a computational model. Such a model can then be easily probed *in silico* and its predictions checked against experimental data in order to validate the correctness of biological hypotheses. When inconsistencies between modelling and experiments arise, this is a clue that something important is missing in our drawing of the biological pathway. We can identify this missing link by appropriately modifying the computational model using our biological knowledge, until a better agreement between simulated and experimental data is achieved.

In our example, modelling pointed to an inconsistency between the *in silico* and *in vivo* behaviour of *GAL4* and *GAL80* during the glucose-to-galactose shift (“switch-on”); their decrease in concentration could not be captured by the model, which was simply based on the drawing in Figure 3.1, i.e. on the known biological function of the promoters and proteins in the network. This hinted to the possibility of an unmodelled effect and prompted further experimental investigation of what this could be. We discovered that cell manipulation during the washing steps (needed to perform the medium shift) induced a transient increase in *GAL4* and *GAL80*.

Modelling can also suggest that additional experimental investigation is needed. In particular, we had to face the issue of practical identifiability for the model

parameters. Biological systems, as well as economical ones, often suffer from this problem due to the intrinsic experimental noise [65]. However, when it is possible, extra-experiments can be performed in order to reduce the number of practical non-identifiable parameters. In our case, we enlarged the available data-set by performing the promoter strength experiments.

During the modelling process, the modeller needs to simplify some aspects of the model and to increase the level of details of others, always taking into account the amount and quality of experimental data. For example, we showed that adding an equation for Gal4 protein improves the predictive power of the model. The quality of the fitting and the predictions could be further improved by modelling the proteins levels of all the genes in the network. However, in the actual version of the network, it is not possible to measure protein levels with the exception of only one gene (Cbf1). Thus, the assumption of steady state for protein dynamics is required, not only in order to simplify the model, but mainly to do not introduce the problem of over-fitting and non-uniqueness of parameters for proteins. In order to decide what can be simplified, and what needs to be modelled in more details, it is necessary to go through iterative refinement steps both in the model and in the experimental data-set.

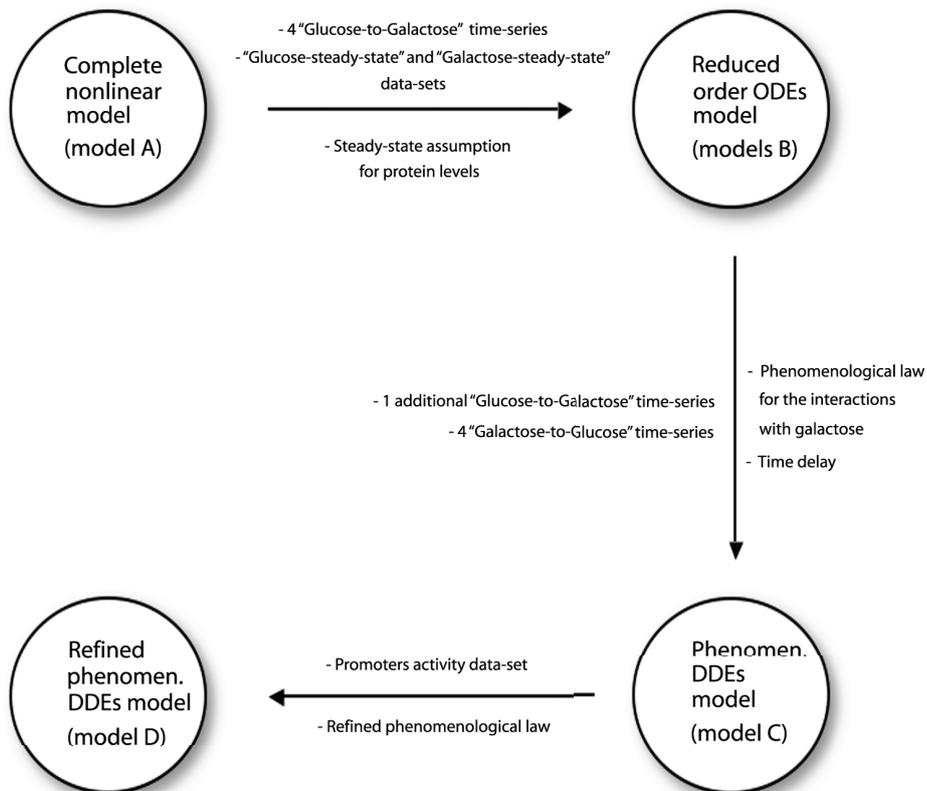


Figure 4.9: **Scheme of the whole modelling and experimental procedure.**

Schematic representation of the steps performed in the refinement of the mathematical model of IRMA and in the set-up of the experiments.

Chapter 5

Turning IRMA into an autonomous oscillator or a bistable switch: non-linear analysis and continuation results

In this Chapter we will show how to use novel tools from numerical bifurcation theory, together with recent results on the link between the dynamics and topology of networks, in order to redesign a synthetic circuit. The need to modify a synthetic network after its biological implementation is common practice in Synthetic biology. When a novel network is built, e.g. a synthetic oscillator, the design at the very beginning is often difficult and can

lead to misleading results mainly due to the lack of quantitative characterisation of network components [31]. In our model-supported approach, the analysis of the previously identified mathematical model allows to increase the predictability of the network dynamics and experimental re-engineering, decreasing the amount of *in vivo* experiments and *post hoc* tweaking to be performed [31, 71]. The model predictions are used to determine how to tune the system parameters, and hence their physical counterparts, in order to change the dynamic behaviour of the network. Of note, the use of bifurcation theory for classification and categorization of the dynamics of species in a reaction mechanism, initiated in [30], is now commonly adopted for the construction and fine-tuning of synthetic networks (see [33] for an overview). In particular, the aim is to understand if and how IRMA can be turned into a robust and tunable synthetic oscillator or a bistable switch. Oscillations have a crucial role in cell behaviour: the circadian clock and the cell cycle are common examples [82]. Currently, the interest of many researchers is focused on the properties of cellular oscillations that only depend on the topology of the reaction network, transcending the individual species involved [27, 46, 116].

In the case of IRMA, the goal is challenging, both in terms of the mathematical analysis and in terms of the *in vivo* implementation. Up to now, only small topologies have been analysed, and the synthetic oscillators experimentally built consist of a few genes (e.g. [32, 7, 38, 107, 101]). Moreover, to our knowledge, numerical continuation techniques for DDEs model have not been applied to the analysis of synthetic gene networks up to now. We

found that multi-step processing of gene products in the negative feedback loop and strong cooperativity in gene regulation are the ingredients to elicit robust oscillations.

In addition, we discovered that by reducing the topology of the network to a single positive feedback loop, IRMA can be turned into a bistable system (a “toggle switch”, that toggles between two discrete, alternative stable steady states). Hysteretic examples have been observed in several natural examples, including the control of lactose utilization in *E. coli*, and ensuring unidirectional cell-cycle progression in eukaryotes [87]. Synthetic switches have been built both in bacterial [39] and mammalian [60] cells for a variety of applications (e.g. gene therapy, construction of bio-sensors and research tools).

The Matlab code of all the models presented in this Chapter is given in Appendix C.

The results presented in this Chapter have been partly published in [72].

5.1 Turning IRMA into an oscillator

With the aim of tuning the dynamics of IRMA and turning it into an autonomous biochemical oscillator, we shall seek to achieve the desired dynamic behaviour by appropriately varying the model parameters. In so doing, it is obviously fundamental both to remain inside the physically feasible range and to minimize the number of changes to the existing network topology and nominal parameter values, in order to speed up the experimental implementation.

In our specific case, the number of physical parameters is quite high (33), thus an exhaustive exploration of the parameter space would be excessively complicated and time consuming. On the other hand, from the analytical view point it is cumbersome to get any results about the structural stability of equilibria under parameters variations since the system is time-delayed and highly non-linear, due to the large value that the Hill coefficients can assume. For the case of our multi-parametric delayed gene network, it is then crucial to restrict the number of parameters to be changed to induce sustained oscillations. For the selection of the parameter subset to be used to carry out the bifurcation analysis, we used as guidelines the links between the topology and the occurrence of autonomous oscillations presented in the recent literature [27, 46, 116, 84, 36, 109]. Exploiting the interplay between parameter variations and network geometry, we decided to vary those parameters which can affect the topology (adding-removing links).

In the analytical studies of simple two-components networks modelled by differential equations [27, 46, 116], it was proposed that the presence of a negative feedback loop and high Hill coefficients in the kinetic functions are the key ingredients for the occurrence of oscillatory behaviour. In [84], the authors consider larger systems with three genes, postulating four general requirements for biochemical oscillations: negative feedback, time delay, sufficient non-linearity of the reaction kinetics and proper balance of the timescales of the reactions. In particular, a negative feedback loop with at least three components can generate oscillations, even without an explicit time delay. It has been further demonstrated that the inclusion of a positive auto-feedback

loop can help in obtaining an oscillatory dynamic behaviour [36]. Extending such an idea, in [109] the authors consider topologies in which, in addition to a negative feedback loop, also a positive one is present, showing that it is generally difficult to adjust a negative feedback oscillator frequency without compromising its amplitude, whereas with positive-plus-negative feedback one can achieve a widely tunable frequency and near-constant amplitude. Thus, positive-plus-negative oscillators appear to be more robust and easier to evolve, rationalizing why they are found in contexts like heartbeats and cell cycles [109].

For the analysis of the IRMA network, we decided to consider only the galactose growing condition, since in such a condition the network is “switched on” and the genes are significantly expressed. Note that, in such condition the protein-protein interaction between Gal4 and Gal80 is not occurring (see Section 3.2). Thus, the topology of IRMA consists of two loops composed only of the transcriptional interactions active in galactose: one delayed positive feedback loop (DFBL) among the genes *CBF1*, *GAL4*, *SWI5* with a delayed reaction due to the presence of the *HO* promoter (see Section 4.3), and one negative feedback loop (NFBL) among the genes *CBF1*, *GAL4*, *SWI5*, *ASH1* (Figure 5.1 (A)). The presence of intermediate states in such negative loop suggests that the network has the potentiality of being turned into an autonomous oscillator, if a proper tuning of the parameters is performed.

In what follows, we will analyse 3 possible re-engineering scenarios in order both to compare the oscillator tunability and robustness due to different network topologies and to explore different experimental strategies for their

implementation.

5.1.1 Scenario 1: stable oscillations keeping the activation of Swi5 on *CBF1* (DDEs model). Simulation and continuation results.

The mathematical model is the one presented in 4.5, but, because of the above considerations, the medium-dependent parameters in the equation of *SWI5* are fixed to their values in galactose. Moreover, since the cells will not be switched from the glucose to the galactose growing condition, the starvation effect induced by the medium shift can be removed.

Letting $[CBF1] = x_1$; $[GAL4] = x_2$; $[SWI5] = x_3$; $[GAL80] = x_4$; $[ASH1] = x_5$, the model thus becomes:

$$\frac{dx_1}{dt} = \alpha_1 + v_1 \left(\frac{x_3^{h_1}(t - \tau)}{(k_1^{h_1} + x_3^{h_1}(t - \tau)) \cdot \left(1 + \frac{x_5^{h_2}}{k_2^{h_2}}\right)} \right) - d_1 x_1, \quad (5.1)$$

$$\frac{dx_2}{dt} = \alpha_2 + v_2 \left(\frac{x_1^{h_3}}{k_3^{h_3} + x_1^{h_3}} \right) - d_2 x_2, \quad (5.2)$$

$$\frac{dx_3}{dt} = \alpha_3 + v_3 \left(\frac{x_2^{h_4}}{k_4^{h_4} + x_2^{h_4} \left(1 + \frac{x_4^{h_7}}{\gamma^{h_7}}\right)} \right) - d_3 x_3, \quad (5.3)$$

$$\frac{dx_4}{dt} = \alpha_4 + v_4 \left(\frac{x_3^{h_5}}{k_5^{h_5} + x_3^{h_5}} \right) - d_4 x_4, \quad (5.4)$$

$$\frac{dx_5}{dt} = \alpha_5 + v_5 \left(\frac{x_3^{h_6}}{k_6^{h_6} + x_3^{h_6}} \right) - d_5 x_5, \quad (5.5)$$

By looking at the values of the kinetic parameters estimated from *in vivo* data (Table A.2, Nominal Value column), it emerges that all the interactions

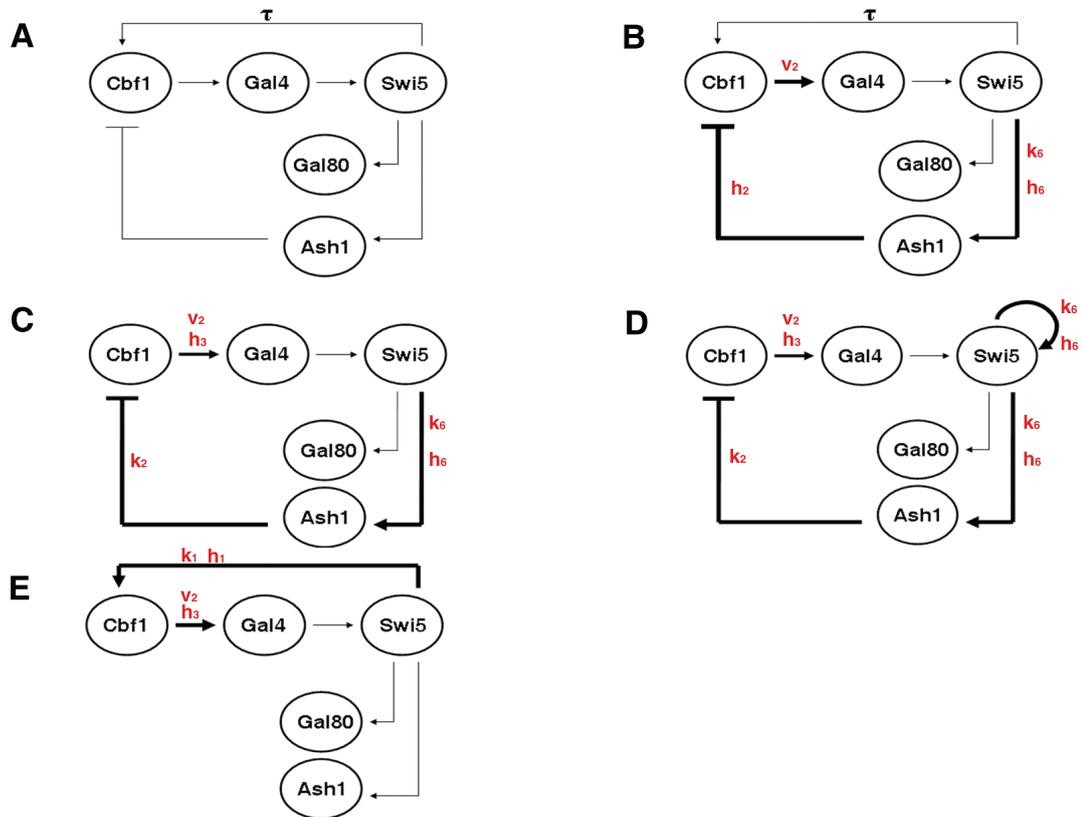


Figure 5.1: Re-engineering the topology in order to turn IRMA into an oscillator or a switch.

Comparison between the topology of the actual version of the network (A) and the re-engineered topologies (B)-(E). A thicker line corresponds to an increase of the strength (in terms of Michaelis-Menten coefficient and/or Hill coefficient and/or maximal transcriptional velocity) of the corresponding interaction. The parameters in red are the ones varied from the nominal value. (A) Topology of IRMA in galactos. (B) Re-engineering of IRMA, Scenario 1. Tuning the parameters v_2 , k_6 , h_2 and h_6 we increase the strength of the following interactions: Cbf1 on Gal4, Swi5 on Ash1 and Ash1 on Cbf1. Both the original positive and the negative feedback lops are present. (C) Re-engineering of IRMA, Scenario 2. Tuning the parameters v_2 , k_1 , k_2 , k_6 , h_3 and h_6 we increase the strength of the following interactions: Cbf1 on Gal4, Swi5 on Ash1 and Ash1 on Cbf1. The original positive feedback loop is removed. (D) Re-engineering of IRMA, Scenario 3. The topology is identical to the one in Scenario 2 with the addition of a positive auto-feedback loop on Swi5. The tuned parameters are: v_2 , k_1 , k_2 , k_6 , h_3 and h_6 . (E) Re-engineering of IRMA, Scenario 4. Tuning the parameters v_2 , k_1 , k_2 , h_1 and h_3 we increase the strength of the following interactions: Cbf1 on Gal4, Swi5 on Cbf1. The negative feedback loop is removed.

in the NFBL loop are balanced in terms of strength and timescales, except for the maximal velocity of transcription of the *MET16* promoter v_2 (which drives the expression of *GAL4*) and the Michaelis-Menten coefficient k_6 , which describes the strength of the activation of *Swi5* on *ASH1* gene. In particular, the parameter v_2 is two order of magnitude lower than all other maximal transcriptional rates while the Michaelis-Menten k_6 coefficient is one order of magnitude higher. Thus, in order to balance the strength of the regulations involved in the negative feedback loop, we started by decreasing the value of k_6 and increasing the value of v_2 , as schematically shown in Figure 5.1 (B). Then, we evaluated the effect of the non-linearity of the reaction kinetics generated by the Hill functions on the network behaviour. Since the stiffness of such sigmoidal function is determined by the Hill coefficients, which describe the cooperativity of gene regulation, we performed our numerical investigations by increasing the Hill coefficients h_2 and h_6 (Figure 5.1 (B)). With the parameters choice reported in Table A.2 (Scenario 1 A column), the dynamic behaviour of the network appears like in Figure 5.2. Here, oscillations have period equal to 120 minutes, thus close to the yeast cell cycle period in galactose; the amplitude is physically feasible and observable for all the mRNAs, but *CBF1*.

Once oscillations are obtained, a fundamental step in the theoretical analysis is the investigation of the robustness and the tunability of the oscillator. To this aim, we used numerical continuation techniques [96].

The transition from a stable steady state solution to a periodic state happens through a supercritical Hopf bifurcation, which occurs when the real part of

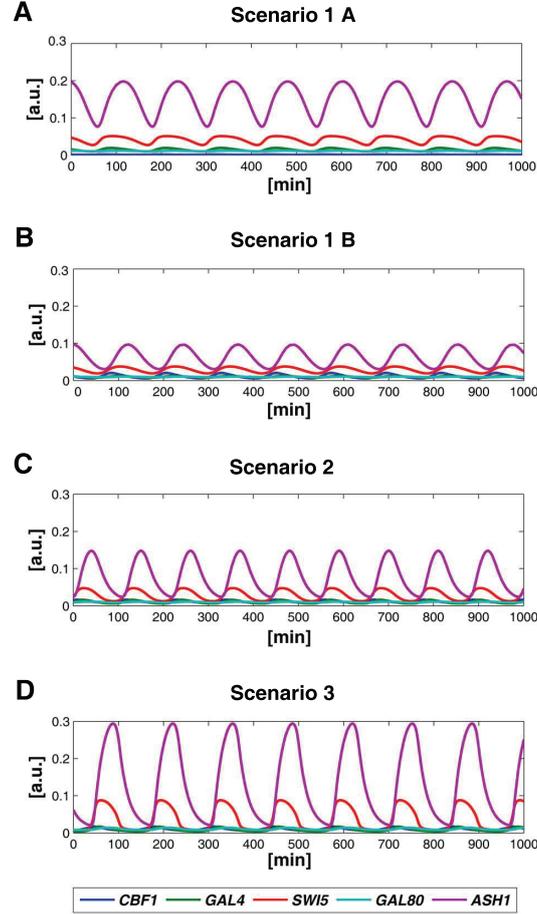


Figure 5.2: **Turning IRMA into an oscillator: time simulations.**

(A) Scenario 1 A, simulations of the DDEs model; parameters v_2 , k_6 , h_2 and h_6 were varied from their nominal values (Table A.2, Scenario 1 A column). Period of the oscillations=120 minutes. (B) Scenario 1 B, simulations of the DDEs model; parameters k_6 , h_2 and h_6 were varied from their nominal values like in Scenario 1 A (Table A.2, Scenario 1 B column), while v_2 was tuned according to the continuation results in order to increase the values of *CBF1*. Period of the oscillations=120 minutes. (C) Scenario 2, simulations of the ODEs model; parameters v_2 , k_1 , k_2 , k_6 , h_3 and h_6 were varied from their nominal values (Table A.2, Scenario 2 column). Period of the oscillations=110 minutes. (D) Scenario 3, simulations of the ODEs model; parameters v_2 , k_1 , k_2 , k_6 , h_3 and h_6 were varied from their nominal values (Table A.2, Scenario 3 column). Period of the oscillations=133 minutes.

a complex conjugate pair of eigenvalues of the Jacobian matrix crosses zero, while the real parts of all other eigenvalues remains negative. The software used to perform numerical continuation is DDE-BIFTOOL [35], the first general-purpose package for bifurcation analysis of DDEs. The characteristic matrix appearing in the stability theory for DDEs has an infinite number of eigenvalues because of the infinite-dimensional nature of DDEs. To determine the local stability of an equilibrium, in DDE-BIFTOOL [35] a linear multi-step method is applied to the variational equation and the approximations to the rightmost (stability determining) characteristic roots are computed. In case of periodic solution of period T , a discrete approximation on a mesh in $[0, T]$ and its period are computed as solutions of the corresponding periodic boundary value problem by using a piecewise polynomial collocation. The local asymptotic stability of a periodic solution is determined by the spectrum of the linear so-called monodromy operator [35]. Further details on the employed methods and the underlying theory can be found in [69, 34]. The limit cycle can be continued on each of the 4 parameters we are varying (k_6, v_2, h_2, h_6). Moreover, once the Hopf bifurcation is localized, it is possible to continue it on all the pairs obtainable by combining such 4 parameters. From continuation results represented in Figure 5.3, it emerges that keeping the Michelis-Menten parameter k_6 low (i.e. keeping the activation of Swi5 on *ASH1* strong enough) is fundamental to guarantee persistent oscillations. The range of k_6 that allows the desired dynamics is further enlarged when the k_1 coefficient increases (see Figure 5.3 (A)): it means that, if the strength of the positive loop decreases, oscillations are guaranteed only if the strength

of the negative loop decreases as well. Figure 5.3 (B) shows that k_6 must be kept small if the maximal transcriptional velocity of the *MET16* promoter increases, remarking that the reaction in the loop must be balanced in terms of strength. In Figure 5.3 (C) and (D) we continue the Hopf bifurcation to analyse the relationship between the Hill coefficients h_2 and h_6 and the Michealis-Menten parameter k_6 , showing that if the activation of Swi5 on *ASH1* is strong enough, the cooperativity coefficient can be decreased without losing persistent oscillations.

Furthermore, continuation allowed us to investigate the tunability of the oscillator in terms of amplitude and period (Figure 5.3 (E) and (F)). We found that the amplitude and the period of the oscillations are tunable individually, thus confirming what stated in [109] for topologies that include both a negative and a positive feedback loop. The parameter that was found to affect the period of the oscillations the most is h_2 : increasing it can enlarge the period up to 18 minutes (Figure 5.3 (F)), but the amplitude of the oscillations stays almost constant (results not shown). Regarding the amplitude, we found that it can be tuned by varying the parameter v_2 inside the range that ensures oscillations (Figure 5.3 (E)). Thus, using continuation, we found how to increase the amplitude of *CBF1* oscillations. By simulating the dynamics of the network using the parameters of Scenario 1 B (all parameters identical to Scenario 1 A, but v_2 set equal to the value for which the amplitude of x_1 has its maximum in Figure 5.3 (E)), we get observable oscillations for all the genes (Figure 5.2 (B)).

Finally, it is useful to test for the robustness of the oscillator under initial

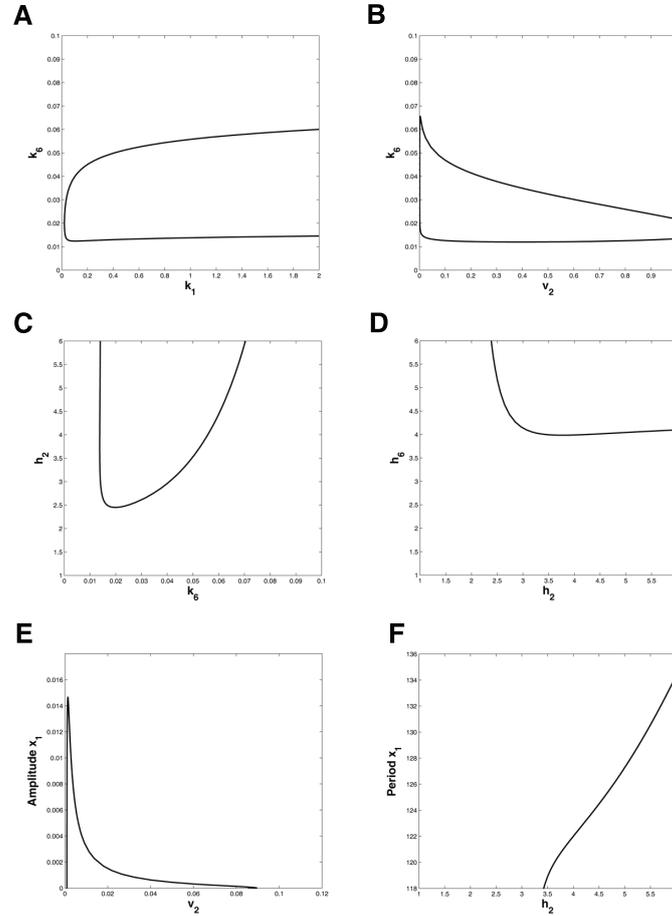


Figure 5.3: **Continuation results for Scenario 1.**

Continuation results for Scenario 1A using DDE-BIFTOOL software. (A) Two parameters continuation of the Hopf bifurcation on parameters k_1 (Michealis-Menten coefficient of the *HO* promoter) and k_6 (Michealis-Menten coefficient of the *ASH1* promoter). (B) Two parameters continuation of the Hopf bifurcation on parameters v_2 (maximal transcriptional rate of the *MET16* promoter) and k_6 (Michealis-Menten coefficient of the *ASH1* promoter). (C) Two parameters continuation of the Hopf bifurcation on parameters k_6 (Michealis-Menten coefficient of the *ASH1* promoter) and h_2 (Hill coefficient of the *HO* promoter). (D) Two parameters continuation of the Hopf bifurcation on parameters h_2 (Hill coefficient of the *HO* promoter) and h_6 (Hill coefficient of the *ASH1* promoter). (E) Tunability of the oscillations in terms of amplitude. Amplitude of x_1 (level of the *CBF1* gene) continuing the periodic solution on v_2 (maximal transcriptional rate of the *MET16* promoter). (F) Tunability of the oscillations in terms of period. Period of x_1 (*CBF1* gene) continuing the periodic solution on h_2 (Hill coefficient of the *HO* promoter).

conditions variations. To this aim, we performed a significant number of time simulations (5000) fixing the parameters to the values in Table A.2 by changing randomly the initial conditions for all the five genes, keeping all of them into a physical reasonable range ($[0 \ 1]$ [a.u.]). The simulations show robustness with all trajectories converging to limit cycles of period 1.

Experimental implementation of Scenario 1 *in vivo*.

At this point, it is crucial to address the feasibility of re-engineering IRMA *in vivo* according with our theoretical results.

In order to increase the maximal transcription velocity v_2 of the *MET16* promoter, the idea is to decrease the level of methionine in the yeast. As we reported in Chapter 3, methionine modulates the expression of the *MET* genes by affecting the formation of the Cbf1-Met4-Met28 transcriptional complex [62]. The activation of Cbf1 on Gal4 is the weakest in the actual version of the network, being the *MET16* promoter weak for the methionine concentrations used in our medium ($140 \mu m$) [15].

From the experimental results presented in Section 3.3.2, we can conclude that increasing the the maximal transcriptional rate v_2 , that determines the steady state of the *MET16* promoter and allows to tune the amplitude of the oscillations, can be achieved by simply decreasing the level of methionine in the medium.

Regarding the changes to the *ASH1* promoter, we need to vary h_6 and k_6 . This can be done by replacing this promoter with a stronger one. A possible candidate is the *EGT2* promoter [95]. Since this gene is activated even by

low levels of Swi5, as well as, by the mutant version of Swi5 (Swi5-AAA) that is present in IRMA [15], it should ensure a low Michaelis-Menten parameter k_6 , required for obtaining the oscillatory behaviour. Moreover, six putative binding sites have been identified [95], thus ensuring a high Hill coefficient h_6 .

The last parameter to be tuned is the Hill coefficient h_2 . Actually, this would be difficult since, in the analysed scenario, all the kinetic parameters of the *HO* promoter are kept equal to their nominal values, but h_2 , that describes the cooperativity of the inhibition of Ash1 on such promoter. Increasing such cooperativity could be implemented *in vivo* by increasing the number of binding sites for Ash1 on the *HO* promoter, although it has not been previously demonstrated that experimental re-engineering would affect only the Hill coefficient and not other parameters, e.g. the Michaelis-Menten constant of the promoter. Furthermore, such promoter is also activated by Swi5 and the regulatory mechanisms are quite complex [15]. We can conclude that the re-engineering of the *HO* promoter could be troublesome.

5.1.2 Scenario 2: stable oscillations by removing the activation of Swi5 on *CBF1* (ODEs model). Simulation and continuation results.

The positive loop in Scenario 1 seems difficult to implement *in vivo*. Therefore, we considered a second scenario, in Figure 5.1 (C), in which the delayed activation of Swi5 on *Cbf1* is removed and the topology of IRMA is reduced to

a negative feedback loop through the genes *CBF1*, *GAL4*, *SWI5* and *ASH1*. In the model, this corresponds to fixing the Michealis-Menten coefficient k_1 to zero or equivalently rewrite equation (5.1) as:

$$\frac{dx_1}{dt} = \alpha_1 + v_1 \left(\frac{k_2^{h_2}}{k_2^{h_2} + x_3^{h_2}} \right) - d_1 x_1. \quad (5.6)$$

Again, we tuned both the strength of the negative loop (by decreasing k_6 and increasing v_2) and the non-linearity of the reaction kinetics (by increasing the Hill coefficients h_3 and h_6). Moreover, we increased the strength of the inhibition of Ash1 on *CBF1* by reducing the value of the Michaelis-Menten coefficient k_2 . Using the parameters in Table A.2 (Scenario 2 column), simulations show the presence of sustained oscillations with period equal to 110 minutes (Figure 5.2 (C)). Note that the amplitude of the oscillations is predicted to be experimentally observable for all the genes, including *CBF1*.

Such a scenario can be analysed in terms of robustness to parameters variations and tunability by using the continuation tool DDE-BIFTOOL with no delayed variable. The most relevant continuation results, reported in Figure 5.4, lead to conclusions similar to the ones discussed for the first scenario. Namely, it is of utmost importance to keep the Michaelis-Menten parameters k_6 and k_2 low and the Hill coefficients h_3 and h_6 large enough. This confirms that, to have oscillatory behaviour, a proper balance of the reactions in the negative feedback loop is needed together with the presence of significant non-linearities.

Furthermore, through continuation we investigated the tunability of the oscil-

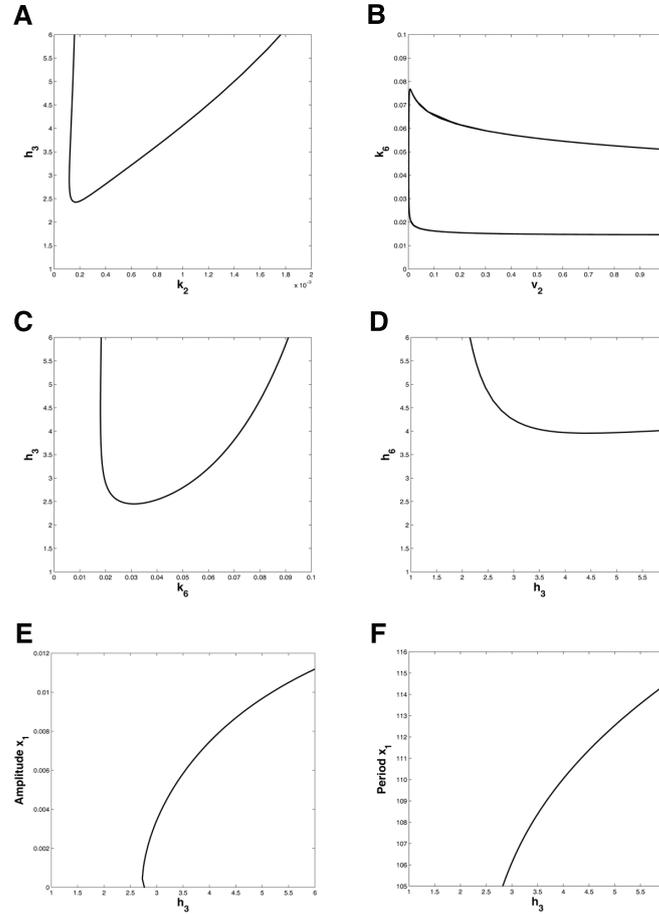


Figure 5.4: **Continuation results for Scenario 2.**

Continuation results for Scenario 2 using DDE-BIFTOOL software. (A) Two parameters continuation of the Hopf bifurcation on parameters k_2 (Michealis-Menten coefficient of the *HO* promoter) and h_3 (Hill coefficient of the *MET16* promoter). (B) Two parameters continuation of the Hopf bifurcation on parameters v_2 (maximal transcriptional rate of the *MET16* promoter) and k_6 (Michealis-Menten coefficient of the *ASH1* promoter). (C) Two parameters continuation of the Hopf bifurcation on parameters k_6 (Michealis-Menten coefficient of the *ASH1* promoter) and h_3 (Hill coefficient of the *MET16* promoter). (D) Two parameters continuation of the Hopf bifurcation on parameters h_3 (Hill coefficient of the *MET16* promoter) and h_6 (Hill coefficient of the *ASH1* promoter). (E) Tunability of the oscillations in terms of amplitude. Amplitude of x_1 (level of the *CBF1* gene) continuing the periodic solution on h_3 (Hill coefficient of the *MET16* promoter). (F) Tunability of the oscillations in terms of period. Period of x_1 (*CBF1* gene) continuing the periodic solution on h_3 (Hill coefficient of the *MET16* promoter)

lator, discovering that in Scenario 2, contrary to what found for Scenario 1, it is not possible to tune the amplitude independently of the period. The unique parameter that allows to tune the dynamics of oscillations is h_3 , that significantly affects both the period and the amplitude (Figure 5.4, (E) and (F)). Such results confirms what stated in [109] about the tunability of topologies composed only by a negative feedback loop.

Testing through simulations the network dynamics under varying initial conditions within the range $[0 \ 1]$ [a.u], we observed again that robustness is guaranteed. All the trajectories converge to limit cycles of period 1 (results not shown).

Experimental implementation of Scenario 2 *in vivo*.

The critical parameters which have to be tuned to implement scenario 2 *in vivo* are v_2 , k_6 , h_6 , h_3 , k_1 and k_2 . Concerning the first four, we could proceed like it has been described for Scenario 1: decrease the level of methionine in order to increase the strength of the activation of Cbf1 on Gal4 and replace the *ASH1* promoter with the *EGT2* promoter. Moreover, it is possible to tune also the h_3 parameter by changing the level of methionine in the yeast. In fact, the behaviour of the *MET16* promoter with low methionine concentrations should become switch like, thus leading to an increase of the stiffness of the sigmoidal Hill function modelled by the h_3 coefficient.

The tuning of parameters k_1 and k_2 requires two additional changes: first to replace the *HO* promoter with a promoter which is not activated by Swi5. Secondly, we need to replace *ASH1* gene with a gene whose expression is

driven by the *EGT2* promoter and that is able to inhibit strongly the new promoter. A good candidate inhibitor-promoter couple is given by *ROX1* repressor and *ANB1* promoter [59].

5.1.3 Scenario 3: stable oscillations by removing the activation of Swi5 on *CBF1* and by adding a positive auto-feedback loop on *SWI5* (ODEs model). Simulation and continuation results.

The topology proposed in Scenario 2 appears feasible for *in vivo* implementation and the oscillations appear robust to varying parameters and initial conditions. For the sake of completeness, we considered also the possibility of including in the network a positive feedback loop, in order to check if the robustness and the tunability of the oscillations increase, according to what shown in a number of works [68, 109, 105].

In Scenario 3, the topology of the network is the same as in Scenario 2 with the addition of an auto-activation reaction on *SWI5* (Figure 5.1 (D)). The parameters are the same of Scenario 2 (Table A.2), but in the ODEs the changes correspond to fixing the Michealis-Menten coefficient k_1 to zero, thus substituting (5.1) with (5.6), and adding an activation term in equation (5.3) that becomes:

$$\frac{dx_3}{dt} = \alpha_3 + v_3 \left(\frac{x_2^{h_4}}{k_4^{h_4} + x_2^{h_4} \left(1 + \frac{x_4^{h_7}}{\gamma^{h_7}}\right)} \right) + v_5 \left(\frac{x_3^{h_6}}{k_6^{h_6} + x_3^{h_6}} \right) - d_3 x_3. \quad (5.7)$$

Numerical simulations show sustained oscillations with period equal to 133 minutes (Figure 5.2 (D)). Note that the amplitude of the oscillations is physically feasible and observable for all the genes; in particular, it is significantly higher than in Scenario 2 for the genes *SWI5* and *ASH1*.

We can compare the robustness to parameter variations of Scenarios 2 and 3 by continuing the Hopf bifurcation on the same pairs of parameters considered previously. By comparing Figure 5.4 (A)-(D) and Figure 5.5 (A)-(D), it appears that the parameter's regions that ensure oscillatory behaviour are significantly enlarged. Moreover, unlike the single negative feedback topology, the topology of Scenario 3 allows to tune the amplitude of the oscillations independently from the period (Figure 5.5, (E) and (F)). The period of oscillations can be varied up to 30 minutes, while in Scenario 2 the maximum change was of 10 minutes. Such results confirm that the robustness and the tunability of the network can increase by adding a positive feedback loop.

Experimental implementation of Scenario 3 *in vivo*.

For the *in vivo* implementation, we need to apply the same changes of Scenario 2 and to add an extra-plasmid containing a *SWI5* responsive promoter upstream of the starting codon of *SWI5*. The previously described *EGT2* promoter is again a good candidate.

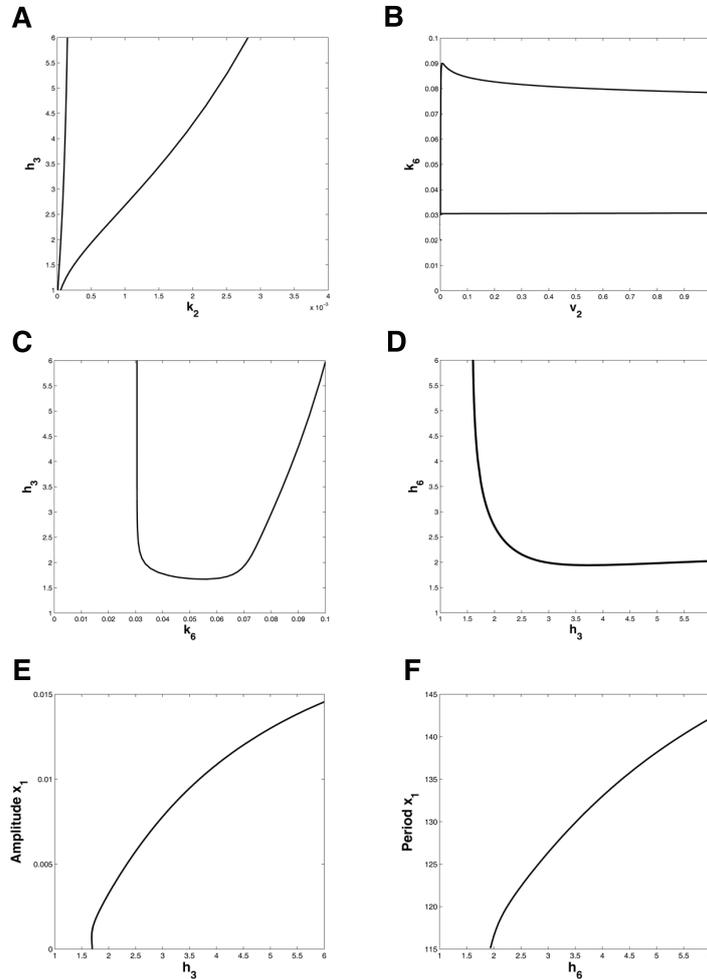


Figure 5.5: **Continuation results for Scenario 3.**

(A) Two parameters continuation of the Hopf bifurcation on parameters k_2 (Michealis-Menten coefficient of the *HO* promoter) and h_3 (Hill coefficient of the *MET16* promoter). (B) Two parameters continuation of the Hopf bifurcation on parameters v_2 (maximal transcriptional rate of the *MET16* promoter) and k_6 (Michealis-Menten coefficient of the *ASH1* promoter). (C) Two parameters continuation of the Hopf bifurcation on parameters k_6 (Michealis-Menten coefficient of the *ASH1* promoter) and h_3 (Hill coefficient of the *MET16* promoter). (D) Two parameters continuation of the Hopf bifurcation on parameters h_3 (Hill coefficient of the *MET16* promoter) and h_6 (Hill coefficient of the *ASH1* promoter). (E) Tunability of the oscillations in terms of amplitude. Amplitude of x_1 (level of the *CBF1* gene) continuing the periodic solution on h_3 (Hill coefficient of the *MET16* promoter). (F) Tunability of the oscillations in terms of period. Period of x_1 (*CBF1* gene) continuing the periodic solution on h_6 (Hill coefficient of the *ASH1* promoter).

5.2 Turning IRMA into a bistable switch.

Our investigation confirmed the flexibility of IRMA, thus we further explored the possibility of turning the network also into a bistable switch. A bistable system is one that toggles between two discrete, alternative stable steady states, in contrast to a monostable system. In biology, bistability has long been established in control of the cell cycle and other oscillations [14], and also recently reported in an artificial gene regulation network [39]. Bistability arises in signaling systems that contain a positive feedback loop or a mutually inhibitory, double negative feedback loop (which, in some regards, is equivalent to a positive feedback loop) [5]. Indeed, in [104] it is demonstrated that the existence of at least one positive feedback loop is a necessary condition for the existence of multiple steady states.

In our setting, the idea is to reduce the actual version of the topology to a 3 gene positive feedback loop between the genes *CBF1*, *GAL4* and *SWI5*, thus removing the inhibition on *CBF1* by Ash1. The corresponding mathematical model consists of equations (5.2)-(5.5) while equation (5.1) is replaced with:

$$\frac{dx_1}{dt} = \alpha_1 + v_1 \left(\frac{x_3^{h_1}}{(k_1^{h_1} + x_3^{h_1})} \right) - d_1 x_1. \quad (5.8)$$

The ODEs model can be analysed by continuing the steady state on the critical parameters. Figure 5.6 (A) and (B) show typical bistability continuation plots: continuing the steady state on k_1 and on h_1 two saddle-node bifurcations delimitate the bistability region in which 3 equilibria coexist, two stable and one unstable. In particular, we can notice that bistability is ensured for

k_1 inside the range [0.02 0.14] [a.u.] and h_1 in [2.3 40], thus the activation of Swi5 on Cbf1 must be strong enough. Figure 5.6 (C) shows the continuation of one saddle-node bifurcation point on two parameters: a codimension 2 bifurcation point (cusp) is detected, from which two branches delimiting the bistability region for the parameters v_2 and h_3 emanate. From such continuation, it emerges that bistability is guaranteed even if we do not vary v_2 and h_3 from their nominal values (Table A.2, Scenario 4 B column): continuing the steady state on k_1 , in Figure 5.6 D we observe again two saddle-node bifurcations delimitating the bistability region that, however, is now slightly smaller ([0.03 0.08] [a.u.]).

5.2.1 Experimental implementation of Scenario 4 *in vivo*.

For the *in vivo* implementation, a simple strategy is to replace the *HO* promoter by inserting the previously described *EGT2* promoter in front of the *CBF1* gene. Correspondingly, in the model the nominal values of k_1 and h_1 (Michaelis-Menten and Hill coefficient of the *HO* promoter in eq. (5.1)) are replaced respectively with k_6 and h_6 . In so doing, the strength and the non-linearity of the positive loop are increased.

Again, we can increase the strength of the activation of the *MET16* promoter by Cbf1 by tuning the parameters v_2 and h_3 as in the previously analysed scenarios by decreasing the methionine concentration in the medium. The overall re-engineering of the topology is schematically represented in Figure 5.1 (E); the parameters are reported in Table A.2, Scenario 4 A column.

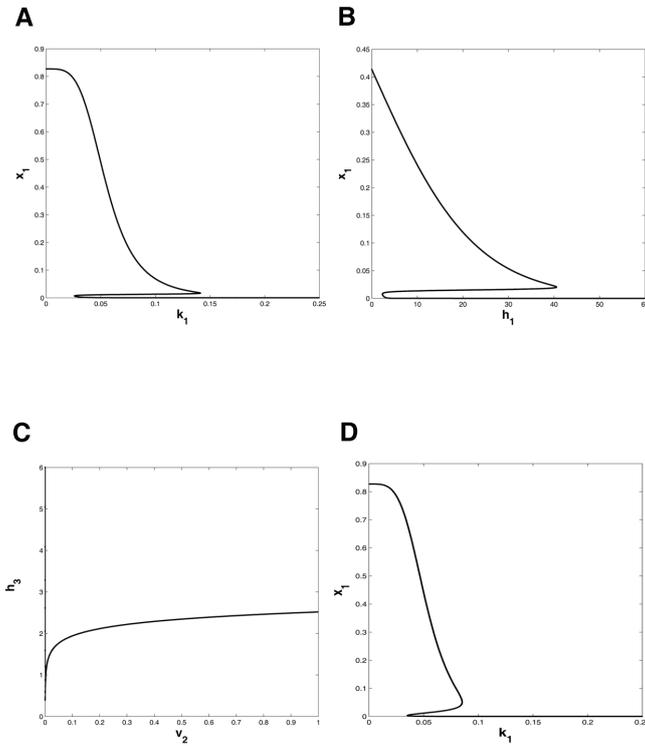


Figure 5.6: **Continuation results for Scenario 4.**

Continuation results for Scenario 4. (A) Scenario 4 A. One parameter continuation of the steady state on k_1 (Michealis-Menten coefficient of the *HO* promoter). Two saddle-node bifurcation points (at $(k_1, x_1)=(0.02, 0.007)$ and $(k_1, x_1)=(0.14, 0.01)$) delimitate the bistability region. (B) Scenario 4 A. One parameter continuation of the steady state on h_1 (Hill coefficient of the *HO* promoter). Two saddle-node bifurcations (at $h_1, x_1=(2., 0.008)$ and $h_1, x_1=(40, 0.019)$) delimitate the bistability region. (C) Scenario 4 A. Two parameters continuation of one saddle-node bifurcation point on v_2 (maximal transcriptional rate of the *MET16* promoter) and h_3 (Hill coefficient of the *MET16* promoter). The cusp bifurcation occurs at $(v_2, h_3)=(0.0005, 0.39)$. (D) Scenario 4 B. One parameter continuation of the steady state on k_1 (Michealis-Menten coefficient of the *HO* promoter). Two saddle-node bifurcation points (at $(k_1, x_1)=(0.03, 0.002)$ and $(k_1, x_1)=(0.08, 0.05)$) delimitate the bistability region.

5.3 Discussion

In this Chapter, using numerical and continuation techniques, we showed how IRMA can be turned into a robust and tunable oscillator, or a bistable genetic switch. The deterministic mathematical model, previously formulated and identified to allow data interpretation and experiment planning, is here analysed to guide the re-engineering of the network with predictable functions. Such *in vivo* re-engineering is actually work in progress in Dr. Diego di Bernardo Systems and Synthetic biology Lab in TIGEM.

IRMA showed great flexibility. Its topology can be re-engineered in a number of ways in order to achieve the desired dynamical behaviour. Of note, all the proposed changes are viable *in vivo*. The robustness to parameters changes and the tunability of the oscillator were assessed via continuations performed using the software DDE-BIFTOOL, the first package for bifurcation analysis of systems with delays that, up to now, has not been commonly used in the Synthetic biology community.

The major conclusion we can draw from our results is that, aiming at constructing a robust and tunable oscillator, the best option is to include in the topology both a delayed negative feedback loop and a fast positive one. This is the case explicitly analysed in Scenario 3 that results to be most robust and tunable as compared to Scenario 2, in which the topology of the network is reduced to a single negative feedback loop.

In the context of Synthetic biology, our model guided re-engineering framework can be applied to existing topologies with the aim of turning them into oscillators or switches. We analysed three topologies for the oscillator case

and one for the switch case. A crucial point was to minimize the number of experiments needed to modify the synthetic network. Surely, other possible ways to re-engineering IRMA can give rise to other oscillatory, switch-like and maybe more complex dynamical behaviours. Of note, once the best performing scenario has been chosen from our deterministic approach, it will be crucial to resort to stochastic simulations in order to estimate the impact of noise on the network dynamics [89]. Remarkably, resulting noise-induced bifurcations can lead to multi-stability or oscillatory dynamics in biochemical networks even when the deterministic description predicts a stable steady state for a certain parameter set [112], or for any parameter values [51].

Chapter 6

Response of the synthetic network in yeast to an external periodic input

In biology, much research effort has been spent on the analysis and investigation of synchronization of biological rhythms. One way to achieve synchronization is via entrainment to some external periodic input. In our everyday life, we experience many behavioural and physiological oscillations that are entrained with the external fluctuating environment. The most famous example is that of circadian rhythms in mammals, regulated by an endogenous biological clock entrained by external signals from the environment [44]. The environmental light-dark cycle, for example, acts as one of the most important pacemakers. Another important example of synchronization and coordination of biological clocks is the cell cycle through which cells periodically

duplicate their genome and divide [110]. Understanding the emergence and coordination of rhythmic phenomena regulating the activities of living organisms requires the investigation of the cooperative behaviour leading to synchronization. Of note, for non-linear systems, driving the system by an external periodic signal does not guarantee the system response also to be a periodic solution, as non-linear systems can exhibit harmonic generation or suppression and complex behaviour such as chaos or quasi-periodic solutions [63].

In this Chapter we will analyse the response of the oscillatory (re-engineered Scenario 2 presented in Section 5.1.2) and non-oscillatory version of the yeast synthetic network to an external periodic input. Such forcing can lead to entrainment, that means that the period of the forced oscillator is exactly the one of the external signal and that the phase of the oscillations is locked. We will analyse entrainment both via simulation and analytically, using recent contraction theory results [93]. Of note, in this Chapter we define *global entrainment* as the convergence of the forced system to some globally attracting limit cycle, whose period is identical to the period of the input, regardless of the properties of the input (period/amplitude of its oscillations).

6.1 Entrainment to periodic input; *in silico* experiments

6.1.1 Numerical results of the oscillatory version of the network (Scenario 2)

Here we apply a forcing oscillatory input to the re-engineered version of the network, we termed as Scenario 2 in Section 5.1.2. We recall that the topology of the network is the one presented in Figure 5.1 (C), and the corresponding mathematical model is:

$$\frac{dx_1}{dt} = \alpha_1 + v_1 \left(\frac{k_2^{h_2}}{k_2^{h_2} + x_5^{h_2}} \right) - d_1 x_1 \quad (6.1)$$

$$\frac{dx_2}{dt} = \alpha_2 + v_2 \left(\frac{x_1^{h_3}}{k_3^{h_3} + x_1^{h_3}} \right) - d_2 x_2, \quad (6.2)$$

$$\frac{dx_3}{dt} = \alpha_3 + v_3 \left(\frac{x_2^{h_4}}{k_4^{h_4} + x_2^{h_4} \left(1 + \frac{x_4^{h_7}}{\gamma^{h_7}}\right)} \right) - d_3 x_3, \quad (6.3)$$

$$\frac{dx_4}{dt} = \alpha_4 + v_4 \left(\frac{x_3^{h_5}}{k_5^{h_5} + x_3^{h_5}} \right) - d_4 x_4, \quad (6.4)$$

$$\frac{dx_5}{dt} = \alpha_5 + v_5 \left(\frac{x_3^{h_6}}{k_6^{h_6} + x_3^{h_6}} \right) - d_5 x_5, \quad (6.5)$$

where $[CBF1] = x_1$; $[GAL4] = x_2$; $[SWI5] = x_3$; $[GAL80] = x_4$; $[ASH1] = x_5$, Using the parameters in Table A.2 (Scenario 2 column), and a fixed input (Galactose=1), simulation shows the presence of sustained oscillations with natural period (T_N) equal to 110 minutes (Figure 5.2 (C)).

Now we check the response of the network when an oscillatory input, e.g.

6.1.1 Numerical results of the oscillatory version of the network (Scenario 94)

galactose, is given to the cells. We will with test signals with different period, say T_G . The duty-cycle of the input signal we put is fixed to 50 %. In Figures 6.1 and 6.2 we report some simulation results. They show that, depending on the period of the forcing input, entrainment can be lost, and quasi-periodic dynamics can appear. In particular, in Figure 6.1 we observe that, given an input with period T_G equal to 110 minutes (A) or 220 minutes (B), the forced system presents oscillations with period $T=T_G$. In Figure 6.2 we observe that, when the period of the input is equal to 55 minutes (A), 27.5 minutes (B) or 18.33 minutes (C), dynamics of the forced system present, respectively, period 2 oscillations, period 4 oscillations and quasi-periodic oscillations.

6.1.1 Numerical results of the oscillatory version of the network (Scenario 2)

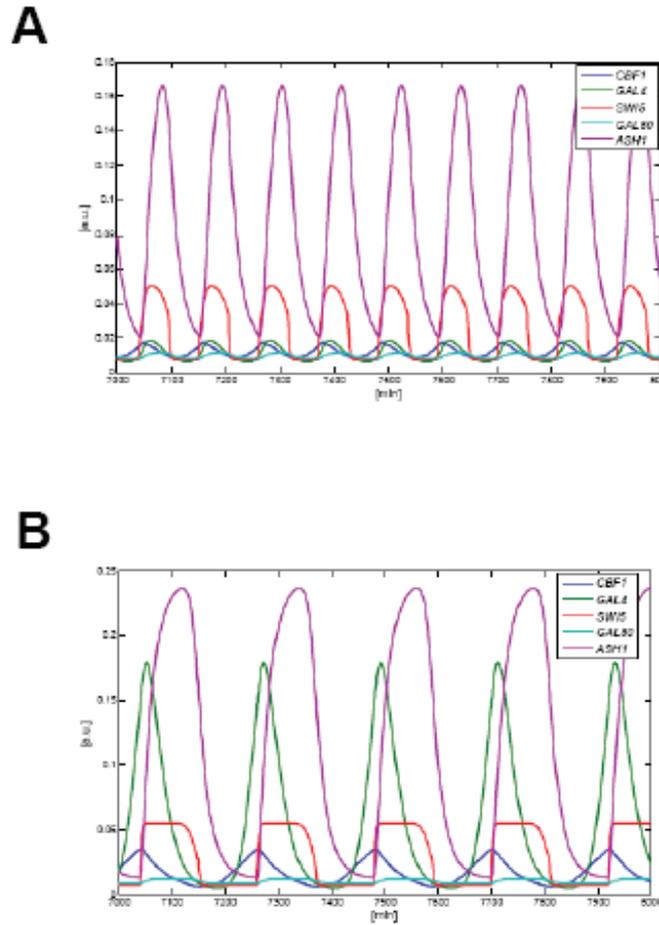


Figure 6.1: **Response of the network to a periodic input.** Periodic dynamics of the oscillating network (Scenario 2), providing a periodic input with (A) $T_G=110$ min; duty=50 %, $T_N/T_G=1$; $T=110$; (B) $T_G=220$ min; duty=50 %, $T_N/T_G=\frac{1}{2}$; $T=220$.

6.1.1 Numerical results of the oscillatory version of the network (Scenario 2)

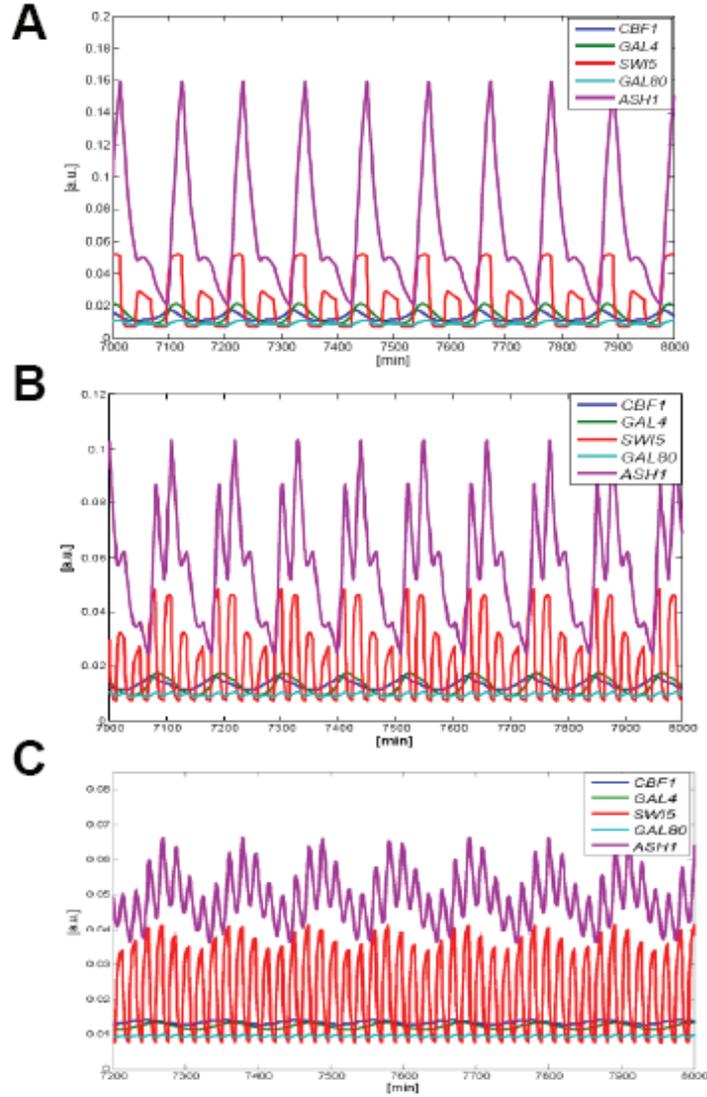


Figure 6.2: **Response of the network to a periodic input** Periodic dynamics of the oscillating network (Scenario 2), providing a periodic input with (A) $T_G=55$ min; duty=50 %, $T_N/T_G=2$; period 2 oscillations; (B) $T_G=27.5$ min; duty=50 %, $T_N/T_G=4$; period 4 oscillations; (C) $T_G=18.33$ min; duty=50 %, $T_N/T_G=6$; quasi-periodic oscillations.

6.1.2 Bifurcation analysis

In order to have a complete picture of the cells output varying input properties, we performed a brute-force two-parameter continuation (see Figure 6.3). In particular, we consider a grid of points in the parameter space (T_G, d) in the parameter range $T_G \in [0 \ 330]$ and $d \in [0 \ 100]$. At each point, the output was simulated and, discarding the transients, the periodicity of the steady state solution was calculated and plotted in different colors (red= period 1, blue= higher period).

Note that the initial conditions of all the variables for simulation are randomly changed at each step in a physically feasible range.

It clearly appears that entrainment (corresponding to the red region in Figure 6.3) is strongly dependent on the properties of the periodic input.

6.1.3 Numerical results on the non-oscillatory version of the network (Scenario 2 modified)

Now we perform a similar simulation analysis applying a forcing oscillatory input to a modified version of topology presented in the above subsection. The model is again composed of equations (6.1)-(6.5); the parameters are ones reported in Table A.2 (Scenario 2 column), with the exception of parameter v_2 and h_3 , which are fixed to the values they have in the non-oscillating version of IRMA (Table A.2, Nominal Value column). Consequently, the topology is again the one in Figure 5.1 (C), but the dynamics of the system, in the presence of a fixed input (Galactose=1), are not oscillatory.

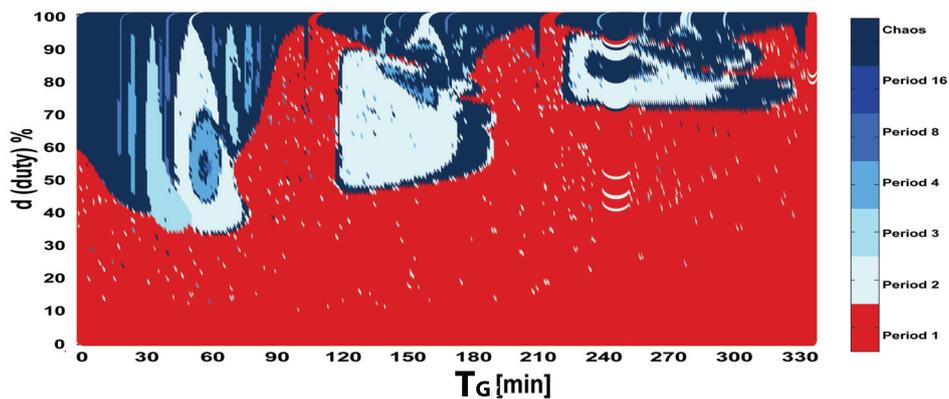


Figure 6.3: **Brute-force two-parameter bifurcation diagram (Scenario 2)**. Entrainment of the oscillating version of IRMA. Colors indicate the period of the forced system oscillations.

Here we directly report simulation results of a brute-force two-parameter continuation diagram (Figure 6.4). In each simulation we change both the period (T_G) and the duty of the periodic input. Entrainment is now always achieved, regardless of the properties of the forcing input.

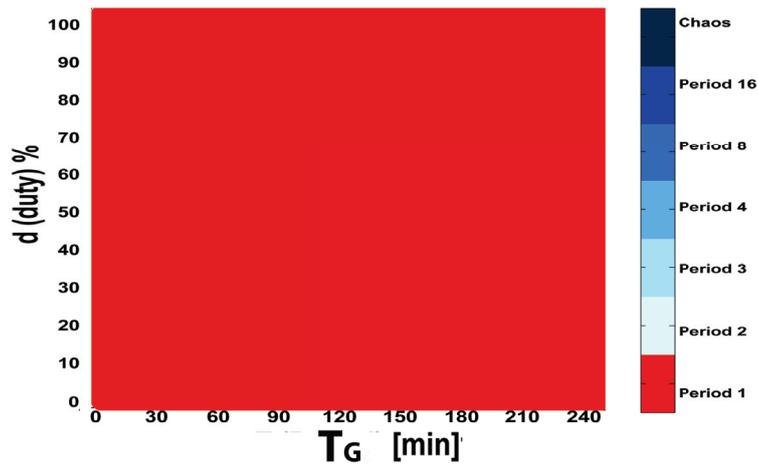


Figure 6.4: **Brute-force two-parameter bifurcation diagram (Scenario 2 modified)**. Entrainment of the non oscillating version of IRMA (Scenario 2 modified). Colors indicate the period of the forced system oscillations.

6.2 Analytical results

6.2.1 Introduction to contraction theory

A most interesting open problem is that of finding analytical conditions for the entrainment to external inputs of biological systems modelled by sets of non-linear differential equations. One approach to analyse the convergence behaviour of non-linear dynamical systems is to use Lyapunov functions.

However, in biological applications, the appropriate Lyapunov functions are not always easy to find and, moreover, convergence is not guaranteed in general in the presence of noise and/or uncertainties. Moreover, such an approach can be hard to apply to the case of non-autonomous systems (that is, dynamical systems directly dependent on time), as is the case when dealing with periodically forced systems.

The above limitations can be overcome if the convergence problem is interpreted as a property of all trajectories, asking that all solutions converge towards one another (contraction). This is the viewpoint of contraction theory [67], and more generally incremental stability methods [54].

Global results are possible, and these are robust to noise, in the sense that, if a system satisfies a contraction property then trajectories remain bounded in the phase space [88].

Contraction theory has been successfully applied to both non-linear control and observer problems [55], and, more recently, to synchronization and consensus problems in complex networks [97]. In [92] it is proposed that contraction can be particularly useful when dealing with the analysis and characterisation of biological networks.

In this Chapter we will use the results presented in [93]. In the paper, Russo et al. provide mathematical conditions that allow one to ensure that biological networks, such as transcriptional systems, can be globally entrained to external periodic inputs. Through the use of contraction theory, it is shown that certain systems driven by external periodic signals have the property that all their solutions converge to a fixed limit cycle.

6.2.2 Contraction results on IRMA

As in [93], starting from the system of interest described by a system of ordinary differential equations $\frac{dx}{dt} = f(x, t)$, we consider a matrix measure of the Jacobian $\mu(J(x, t))$, which is the directional derivative of the matrix norm induced by a vector norm on Euclidian space. The system is infinitesimally contracting on a convex set $C \subseteq R^n$ if there exists some norm in C , with associated matrix measure μ such that, for some constant $c \in R - \{0\}$,

$$\mu(J(x, t)) \leq -c^2, \forall x \in C, \forall t \geq 0. \quad (6.6)$$

In [93], the authors prove that that infinitesimal contractivity implies global contractivity. Also, if the system is contractive, than it can be globally entrained to external periodic inputs.

In our system (equations (6.1)-(6.5)), letting $x_1=x$, $x_2=y$, $x_3=z$, $x_4=v$, $x_5=w$, the Jacobian is given by:

$$J = \begin{bmatrix} -d_1 & 0 & 0 & 0 & J_{15} \\ J_{21} & -d_2 & 0 & 0 & 0 \\ 0 & J_{32} & -d_3 & J_{34} & 0 \\ 0 & 0 & J_{43} & -d_4 & 0 \\ 0 & 0 & J_{53} & 0 & -d_5 \end{bmatrix}$$

where

$$J_{15} = -\frac{v_1 k_2^{h_2} w^{h_2-1} h_2}{(k_2^{h_2} + w^{h_2})^2}, \quad J_{21} = \frac{v_2 h_3 x^{h_3-1} k_3^{h_3}}{(k_3^{h_3} + x^{h_3})^2}, \quad J_{32} = \frac{v_3 h_4 y^{h_4-1} \gamma^{2h_7} k_4^{h_4}}{(k_4^{h_4} \gamma^{h_7} + y^{h_4} \gamma^{h_7} + y^{h_4} v^{h_7})^2},$$

$$J_{34} = -\frac{v_3 y^{2h_4} \gamma^{h_7} v^{h_7-1} h_7}{(k_4^{h_4} \gamma^{h_7} + y^{h_4} \gamma^{h_7} + y^{h_4} v^{h_7})^2}, J_{43} = \frac{v_4 h_5 z^{h_5-1} k_5^{h_5}}{(k_5^{h_5} + z^{h_5})^2}, J_{53} = \frac{v_5 h_6 z^{h_6-1} k_6^{h_6}}{(k_6^{h_6} + z^{h_6})^2}.$$

As matrix measure, we will use the measure μ_P induced by the vector $|\mu_P|$, where P is a suitable non-singular matrix. More specifically, we will pick P diagonal with positive elements; in our case it is a 5x5 matrix. As matrix measure, we will use

$$\mu_{P,\infty}(J) = \mu_\infty(PJP^{-1}) \quad (6.7)$$

We can then calculate

$$PJP^{-1} = \begin{bmatrix} -d_1 & 0 & 0 & 0 & J_{15}^P \\ J_{21}^P & -d_2 & 0 & 0 & 0 \\ 0 & J_{32}^P & -d_3 & J_{34}^P & 0 \\ 0 & 0 & J_{43}^P & -d_4 & 0 \\ 0 & 0 & J_{53}^P & 0 & -d_5 \end{bmatrix}$$

where

$$J_{15}^P = -\frac{p_1 v_1 k_2^{h_2} w^{h_2-1} h_2}{(k_2^{h_2} + w^{h_2})^2 p_5}, J_{21}^P = \frac{p_2 v_2 h_3 x^{h_3-1} k_3^{h_3}}{(k_3^{h_3} + x^{h_3})^2 p_1}, J_{32}^P = \frac{p_3 v_3 h_4 y^{h_4-1} \gamma^{h_7} k_4^{h_4}}{(k_4^{h_4} \gamma^{h_7} + y^{h_4} \gamma^{h_7} + y^{h_4} v^{h_7})^2 p_2},$$

$$J_{34}^P = -\frac{p_3 v_3 y^{2h_4} \gamma^{h_7} v^{h_7-1} h_7}{(k_4^{h_4} \gamma^{h_7} + y^{h_4} \gamma^{h_7} + y^{h_4} v^{h_7})^2 p_4}, J_{43}^P = \frac{p_4 v_4 h_5 z^{h_5-1} k_5^{h_5}}{(k_5^{h_5} + z^{h_5})^2 p_3}, J_{53}^P = \frac{p_5 v_5 h_6 z^{h_6-1} k_6^{h_6}}{(k_6^{h_6} + z^{h_6})^2 p_3}.$$

In order to prove contractivity of the system, we have to prove that it satisfies (6.6), that is finding a set of scalars c_i, p_i , with $i = 1 \dots 5$, that guarantees that the sum of the elements on each column of the above matrix is less or equal to c_i^2 .

In particular, we want to check contractivity for Scenario 2 (parameter values

in Table A.2, Scenario 2 column), and the modified Scenario 2 (same parameters of Scenario 2 with the exception of parameters v_2 and h_3 , which are fixed to the values reported in the Nominal Value column of Table A.2). In what follows, we will calculate the sum on the columns of the matrix PJP^{-1} , and fix the Hill coefficients and the Michalis-Menten coefficients, thus letting free the maximal transcriptional rates and the degradation rates.

1. **Condition on the sum of the elements of the first column of the matrix PJP^{-1} :**

$$-d_1 + \frac{p_2 v_2 h_3 x^{h_3-1} k_3^{h_3}}{(k_3^{h_3} + x^{h_3})^2 p_1} \leq -c_1^2. \quad (6.8)$$

For *Scenario 2 modified* ($h_3=1$) the inequality (6.8) becomes

$$-d_1 + \frac{p_2 v_2 k_3}{(k_3 + x)^2 p_1} \leq -c_1^2. \quad (6.9)$$

It is easy to check that

$$-d_1 + \frac{p_2 v_2 k_3}{(k_3 + x)^2 p_1} \leq -d_1 + \frac{p_2}{p_1} v_2 26.88 \leq -c_1^2. \quad (6.10)$$

For the values of parameters of Scenario 2 modified, such inequality holds $\forall p_1, p_2 > 0$.

For *Scenario 2* ($h_3=4$), from (6.8) we can derive the following inequality:

$$-d_1 + \frac{4 p_2 v_2 x^3 k_3^4}{(k_3^4 + x^4)^2 p_1} \leq -c_1^2. \quad (6.11)$$

Substituting the parameters value of k_3 in Table A.2 (Scenario 2 column), and letting free the degradation rate d_2 and the maximal transcriptional rate v_2 , if we calculate the maximum of the resulting algebraic formula we get:

$$-d_1 + 28.7 v_2 \frac{p_2}{p_1} \leq -c_1^2. \quad (6.12)$$

If we can ensure that

$$\frac{v_2}{d_1} \leq 0.034, \quad (6.13)$$

inequality (6.8) holds $\forall p_1, p_2 > 0$. For the parameter values in Table A.2 (Scenario 2 column), $\frac{v_2}{d_1} = 1.1$, the inequality is not satisfied.

2. Condition on the sum of the elements of the second column of the matrix PJP^{-1} :

$$-d_2 + \frac{p_3 v_3 h_4 y^{h_4-1} \gamma^{2h_7} k_4^{h_4}}{(k_4^{h_4} \gamma^{h_7} + y^{h_4} \gamma^{h_7} + y^{h_4} v^{h_7})^2 p_2} \leq -c_2^2. \quad (6.14)$$

By substituting the values of the Hill coefficient, that are $h_4=h_7=4$ (identical for *Scenario 2* and *Scenario 2 modified*), the above inequality becomes

$$-d_2 + \frac{p_3 v_3 4 y^3 \gamma^8 k_4^4}{(k_4^4 \gamma^4 + y^4 \gamma^4 + y^4 v^4)^2 p_2} \leq -c_2^2. \quad (6.15)$$

Substituting the values of the Michaelis-Menten coefficients and calculating the maximum of the embedded algebraic formula in (6.15) we get:

$$-d_2 + \frac{p_3 v_3 4 y^3 \gamma^8 k_4^4}{(k_4^4 \gamma^4 + y^4 \gamma^4 + y^4 v^4)^2 p_2} \leq -d_2 + 5 \cdot 10^{-10} v_3 \frac{p_3}{p_2} \leq -c_2^2. \quad (6.16)$$

Thus, if we can ensure that

$$\frac{v_3}{d_2} \leq 2.00 \cdot 10^4, \quad (6.17)$$

inequality (6.14) holds $\forall p_2, p_3 > 0$. For the parameter values of both Scenario 2 and Scenario 2 modified $\frac{v_3}{d_2} = 0.42$, the inequality is satisfied.

3. Condition on the sum of the elements of the third column of the matrix PJP^{-1} :

$$-d_3 + \frac{p_4 v_4 h_5 z^{h_5-1} k_5^{h_5}}{(k_5^{h_5} + z^{h_5})^2 p_3} + \frac{p_5 v_5 h_6 z^{h_6-1} k_6^{h_6}}{(k_6^{h_6} + z^{h_6})^2 p_3} \leq -c_3^2. \quad (6.18)$$

By substituting $h_5 = h_6 = 1$ (identical for *Scenario 2* and *Scenario 2 modified*), we get

$$-d_3 + \frac{p_4 v_4 k_5}{(k_5 + z)^2 p_3} + \frac{p_5 v_5 k_6}{(k_6 + z)^2 p_3} \leq -d_3 + \frac{p_4}{p_3} v_4 0.55 + \frac{p_5}{p_3} v_5 16.52 \leq -c_3^2. \quad (6.19)$$

Thus, if

$$\frac{v_4}{d_3} \leq 1.8 \quad (6.20)$$

and

$$\frac{v_5}{d_3} \leq 0.06, \quad (6.21)$$

the inequality (6.18) is satisfied. It is true for the values of parameters of both Scenario 2 and Scenario 2 modified.

4. **Condition on the sum of the elements of the fourth column of the matrix PJP^{-1} :**

$$-d_4 + \frac{p_3 v_3 y^{2h_4} \gamma^{h_7} v^{h_7-1} h_7}{(k_4^{h_4} \gamma^{h_7} + y^{h_4} \gamma^{h_7} + y^{h_4} v^{h_7})^2 p_4} \leq -c_4^2. \quad (6.22)$$

By substituting $h_7 = h_4 = 4$ (identical for *Scenario 2* and *Scenario 2 modified*) we have

$$-d_4 + \frac{p_3 v_3 4 v^3 \gamma^4 y^8}{(k_4^4 \gamma^4 + y^4 \gamma^4 + y^4 v^4)^2 p_4} \leq -c_4^2. \quad (6.23)$$

Proceeding as above we find that, in order to satisfy (6.22) $\forall p_3, p_4 > 0$ we must have

$$\frac{v_3}{d_4} \leq 0.05. \quad (6.24)$$

Such condition is satisfied for the parameter values of both Scenario 2 and Scenario 2 modified.

5. **Condition on the sum of the elements of the fifth column of the matrix PJP^{-1} :**

$$-d_5 + \frac{p_1 v_1 h_2 w^{h_2-1} k_2^{h_2}}{(k_2^{h_2} + w^{h_2})^2 p_5} \leq -c_5^2. \quad (6.25)$$

By substituting $h_2=1$ (identical for *Scenario 2* and *Scenario 2 modified*) we get

$$-d_5 + \frac{p_1 v_1 k_2}{(k_2 + w)^2 p_5} \leq -d_5 + 285v_1 \frac{p_1}{p_5} \leq -c_5^2. \quad (6.26)$$

Thus, if

$$\frac{v_1}{d_5} \leq 0.0035 \quad (6.27)$$

the inequality (6.25) is satisfied $\forall p_1, p_5 > 0$. It is true for the values of parameters values of both Scenario 2 and Scenario 2 modified.

In conclusion, we proved analytically that parameters of Scenario 2 modified, for which the autonomous system dynamics are not oscillatory, verify the conditions required to have global entrainment. On the other hand, for parameters of Scenario 2, entrainment is not guaranteed. In particular, the critical parameters are v_2 and h_3 , which are the ones that differ in Scenario 2 as compared to Scenario 2 modified.

6.3 Discussion

In this Chapter we analysed, both via simulation and analytically, the response of IRMA to an oscillatory input. We showed that only some parameters guarantee that the network can always be entrained to the periodic input. Chaotic attractor can be obtained if the forced system presents autonomous oscillations.

In Dr. Diego di Bernardo Systems and Synthetic biology Lab, we are currently

setting up a novel experimental platform based microfluidics [10] to check *in vivo* the *in silico* predictions. This will allow to reproduce experimentally the presented *in silico* analysis by providing the desired periodic galactose input to both the non-oscillatory and oscillatory version of the network, and to check *in vivo* the appearance of entrainment.

Chapter 7

Mathematical model of a novel synthetic oscillator in mammalian cells

In this Chapter, we present the mathematical modelling of a novel synthetic oscillator that is being developed in the Systems and Synthetic biology Lab of Dr. Diego di Bernardo by Velia Siciliano (for further information see <http://dibernardo.tigem.it>). A full mathematical analysis is developed for two possible topologies of the oscillator. Moreover, we present preliminary *in vivo* results and the relative parameters fitting.

7.1 Aim of the project

The aim of this project is the development of a new synthetic circuit able to express an mRNA/protein of interest independently of the cell cycle or

other oscillatory endogenous signals in a mammalian cell. Most of the studies carried out so far in mammalian cells are based on plasmid transfection, which prevent precise quantitative measurements due to the unpredictable amount of plasmids that enters in each cell, and to the transient nature of transfection. We aim at engineering the synthetic oscillator in a lentiviral vector, so that we can easily transfer the network in living cells and in an *in vivo* animal model. Lentiviral vectors are efficient vehicles for the delivery of genes to both dividing and non-dividing cells *in vitro* and *in vivo* [1]. In view of medical applications, our synthetic circuit will have a significant impact for gene therapy of complex diseases. We are building two versions of the network, in order to understand how the topology affects the oscillator dynamics.

7.2 Description of the biological system: Oscillator topology 1

The first version of the oscillator (Topology 1, presented in Figure 7.1 (A)), consists of an activator, which promotes its own transcription, as well as the transcription of a repressor against itself, in agreement with the design of the natural circadian oscillators [86]. As activator, we are using the tetracycline-controlled transactivator (*tTA*), which is the result of the fusion between the bacterial Tet repressor protein with the VP16 activation domain [100]. *tTA* expression is self-controlled by a *CMV-TET* promoter responsive to the tTA protein. The promoter is inducible by Doxycycline: it is on in absence of the

antibiotic, and off in presence. The same promoter drives the transcription of a destabilized yellow-green variant of enhanced green fluorescent protein (d2EYFP) (Clontech) as readout of the system. The *CMV-TET* promoter regulates also the expression of a microRNA (*miR*) directed against the *tTA*. MicroRNAs are a class of RNA involved in a natural process, the RNA interference, in which there is a sequence-specific RNA-mediated pathway for turning off gene expression [37]. This process has been adapted also for an “artificial” regulation of gene expression via small interfering RNA (siRNA) designed on the mRNA sequence of the gene of interest, and then embedded in the microRNA context. By repressing tTA at the post-transcriptional level, we will overcome any problem due to the leakiness of the promoter, since the gene will be silenced. A red fluorescent protein (mcherry) will enable us to follow the dynamics of the microRNA. Thus, the overall topology is composed by a positive auto-feedback loop and a negative feedback loop between composed of two genes (Figure 7.1 (B)).

7.3 Mathematical model of the network and continuation results

The model we developed is based on non-linear Delay Differential Equations.

We assumed:

- Hill functions to model the rates of gene transcription, including basal activity to describe the leakiness of the promoter (*CMV-TET*);
-

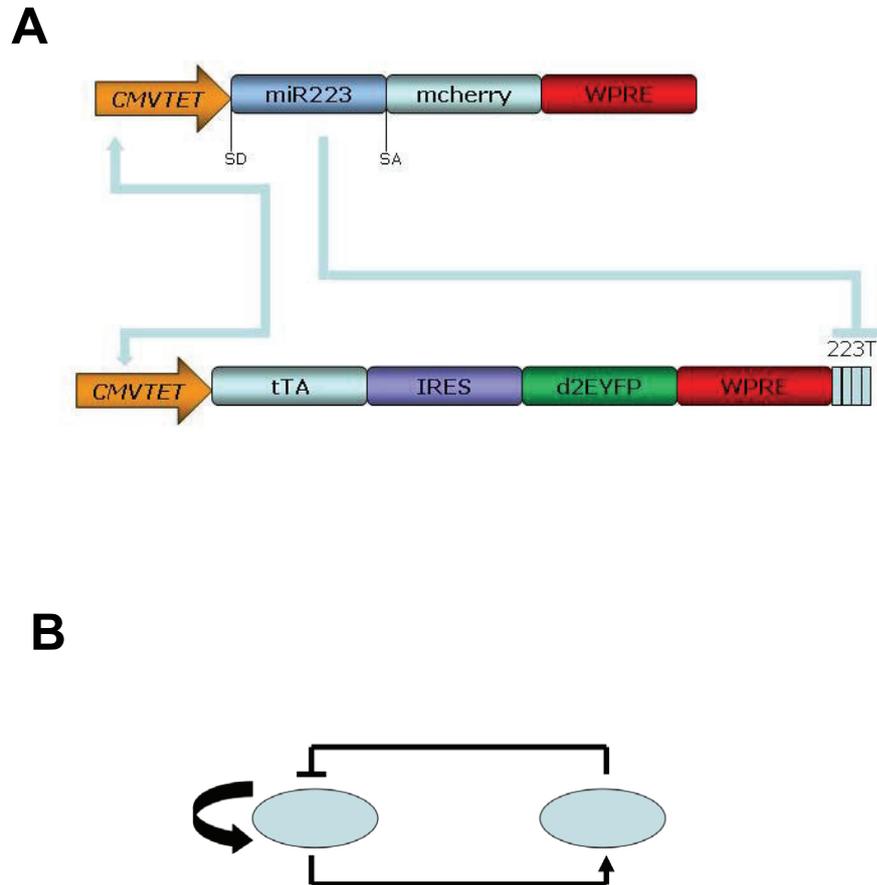


Figure 7.1: **Oscillator in mammalian cells, Topology 1.** In (A) we report the metabolites composing the network, and in (B) a schematic representation of the network topology.

- 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 (modelling approach proposed in [106]). In order to silence its target gene (*tTA*), the microRNA needs to be processed throughout cleavage by Dicer and binding to the RISC complex. In order to take into account these steps, we introduced a fixed time delay in the silencing effect by *miR* on *tTA* (see equation (7.1));
- distinct dynamics for the unfolded (inactive) and folded (active) forms of the reporter proteins.

By setting $x_1 = tTA$ mRNA concentration; $x_2 = tTA$ protein concentration; $x_3 = miR$ mRNA concentration; $x_4 =$ unfolded d2EYFP protein concentration; $x_5 =$ folded d2EYFP protein concentration; $x_6 =$ unfolded mcherry protein concentration; $x_7 =$ folded mcherry protein concentration, we have the following Delay Differential Equations model:

$$\frac{dx_1}{dt} = G_1 v_1 \left(\alpha_1 + (1 - \alpha_1) \frac{\left(\frac{\theta}{\theta+D} x_2\right)^{h_1}}{K_1 + \left(\frac{\theta}{\theta+D} x_2\right)^{h_1}} \right) - d_1 x_1 - \lambda \frac{x_3^{h_2} (t - \tau)}{K_2 + x_3^{h_2} (t - \tau)} x_1, \quad (7.1)$$

$$\frac{dx_2}{dt} = v_2 x_1 - d_2 x_2, \quad (7.2)$$

$$\frac{dx_3}{dt} = G_2 v_1 \left(\alpha_1 + (1 - \alpha_1) \frac{\left(\frac{\theta}{\theta+D} x_2\right)^{h_1}}{K_1 + \left(\frac{\theta}{\theta+D} x_2\right)^{h_1}} \right) - d_3 x_3, \quad (7.3)$$

$$\frac{dx_4}{dt} = v_3x_1 - (d_4 + K_f)x_4, \quad (7.4)$$

$$\frac{dx_5}{dt} = K_fx_4 - d_4x_5, \quad (7.5)$$

$$\frac{dx_6}{dt} = v_3x_3 - (d_4 + K_f)x_6, \quad (7.6)$$

$$\frac{dx_7}{dt} = K_fx_6 - d_4x_7. \quad (7.7)$$

Parameters $d_i, i = 1, \dots, 4$ are the degradation rates, $K_j, j = 1, 2$ are the Michaelis-Menten constants, h_j are the cooperativity constants, K_f is the folding rate, α_1 is the basal activities of the *CMV-TET* promoter, $v_k, k = 1, \dots, 3$ represent the maximal transcription or translation rates, θ models the inducer-*CMV-TET* promoter interaction, τ is the time delay. Parameters G_1 and G_2 are used to model the infection efficiency. In fact, the network is implemented on two separated lentiviral vectors, one containing the cassette *CMVTET-tTA-d2EYFP* and the other the cassette *CMVTET-miR-mcherry*. Of note, in what follows, we will always consider the network in the on condition, thus we will set the concentration of the antibiotic equal to 0.

7.3.1 Identification of the parameters oscillatory region and continuation results

Using the reference parameters values reported in Table A.3 (literature values from [106] and [107]), the dynamics of the network appear monotonically increasing toward a steady-state value.

By properly tuning the parameters values, we found autonomous oscillations (simulation in Figure 7.2 (A), parameters in Table A.3). In simulations, the

time delay was fixed to 60 minutes. In order to achieve oscillatory behaviour, as compared to the literature values, we have:

- increased the efficiency of viral infection of the vector containing the *tTA* with respect of the one the *miR*. It means that we are imposing that the dynamics of the positive feedback loop are faster than the negative ones. This is in agreement with results presented in [27, 46, 116, 84, 36, 109];
- increased the strength and non-linearity of the silencing.

The predicted period of the oscillations is equal to 256 [min].

By using the continuation tool DDE-BIFTOOL [35], we performed one parameter continuation of the limit cycle and two parameters continuation of the Hopf bifurcation on all the combinations of parameters. Continuation results remarked the need of having faster dynamics of the positive feedback loop as compared to the negative loop ones. In particular, the time delay is crucial: it must be at least equal to 52 minutes (see continuation of the limit cycle on the delay reported in Figure 7.2 (B)).

7.4 Model guided re-engineering of the network: oscillator Topology 2

The mathematical analysis presented in the previous section suggested us to implement, in parallel with Topology 1, a second topology (Topology 2 in Figure 7.3 (A)) in which the negative feedback loop encompasses a third

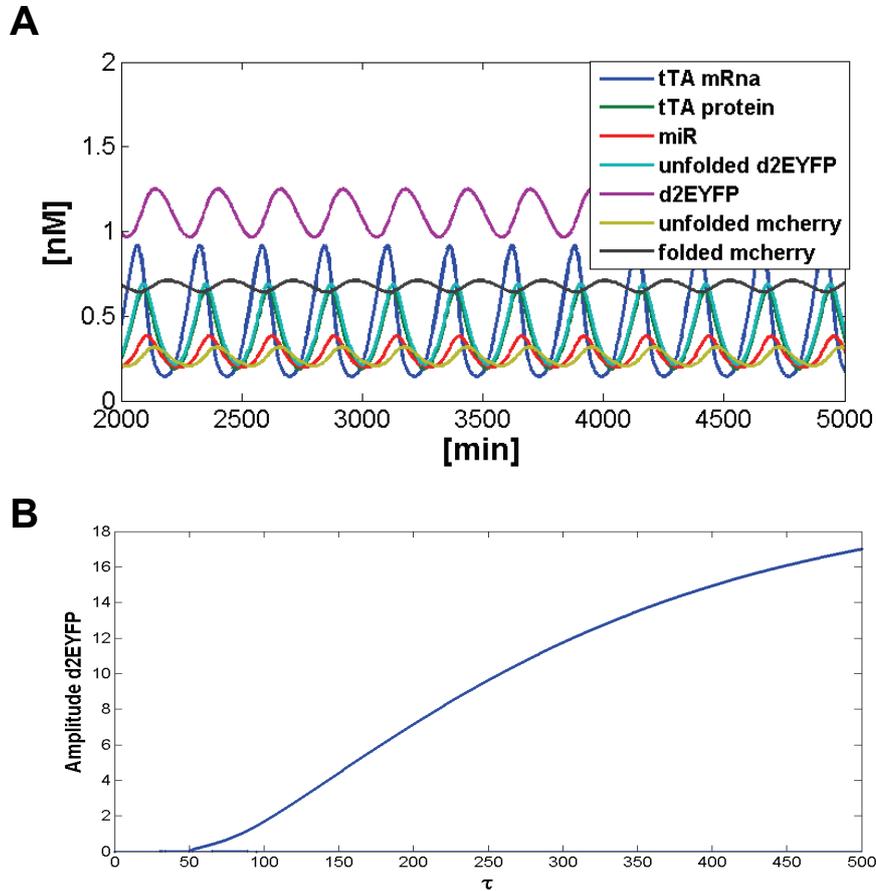


Figure 7.2: **Topology 1, simulation and continuation results.** (A) Model predicted oscillatory dynamics using the parameters reported in Table A.3. (B) Continuation of the Hopf bifurcation on the time delay.

gene, namely *PIT*. It is activated by *tTA* and regulates the expression of the microRNA via the *PPIR* promoter. Note that such promoter is inducible by Pristinamycin (if it is not present in the medium, the promoter is fully active). In this way, even if the process through which the microRNA becomes active (cleavage by Dicer and bounding to the RISC complex) is not slow enough, the dynamics of the overall negative feedback loop should become slower, as it is now composed by three metabolites (Figure 7.3 (B)).

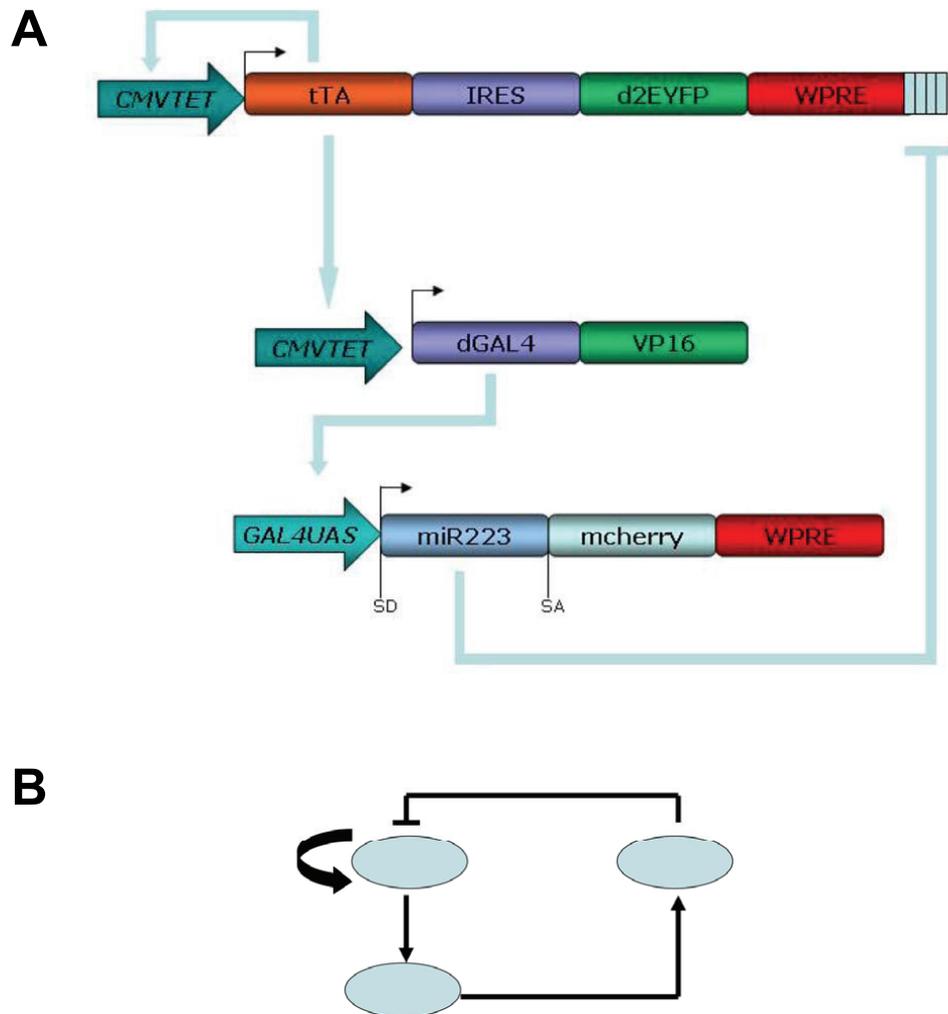


Figure 7.3: **Oscillator in mammalian cells, Topology 2.** In (A) we report the metabolites composing the network, and in (B) a schematic representation of the network topology.

In order to check the validity of the hypothesis that brought us to re-engineer the network, we analysed a DDEs model of the modified network.

7.4.1 Mathematical model, identification of the parameters oscillatory region and continuation results

By setting $x_1 = tTA$ mRNA concentration; $x_2 = tTA$ protein concentration; $x_3 = PIT$ mRNA concentration; $x_4 = PIT$ protein concentration; $x_5 = miR$ concentration; $x_6 =$ unfolded d2EYFP protein concentration; $x_7 =$ folded d2EYFP protein concentration; $x_8 =$ unfolded mcherry protein concentration; $x_9 =$ folded mcherry protein concentration, with the same modelling assumptions presented in Section 7.3, the Delay Differential Equations model of Topology 2 is:

$$\frac{dx_1}{dt} = G_1 v_1 \left(\alpha_1 + (1 - \alpha_1) \frac{\left(\frac{\theta}{\theta+D} x_2\right)^{h_1}}{K_1 + \left(\frac{\theta}{\theta+D} x_2\right)^{h_1}} \right) - d_1 x_1 - \lambda \frac{x_5 (t - \tau)^{h_2}}{K_2^{h_2} + x_5 (t - \tau)^{h_2}} x_1, \quad (7.8)$$

$$\frac{dx_2}{dt} = v_2 x_1 - d_2 x_2, \quad (7.9)$$

$$\frac{dx_3}{dt} = G_2 v_1 \left(\alpha_1 + (1 - \alpha_1) \frac{\left(\frac{\theta}{\theta+D} x_2\right)^{h_1}}{K_1 + \left(\frac{\theta}{\theta+D} x_2\right)^{h_1}} \right) - d_3 x_3, \quad (7.10)$$

$$\frac{dx_4}{dt} = v_3 x_3 - d_4 x_4, \quad (7.11)$$

$$\frac{dx_5}{dt} = G_3 v_4 \left(\alpha_2 + (1 - \alpha_2) \frac{\left(\frac{\gamma}{\gamma+P} x_4\right)^{h_3}}{K_3 + \left(\frac{\gamma}{\gamma+P} x_4\right)^{h_3}} \right) - d_5 x_5, \quad (7.12)$$

$$\frac{dx_6}{dt} = v_5 x_1 - (d_6 + K_f) x_6, \quad (7.13)$$

$$\frac{dx_7}{dt} = K_f x_6 - d_6 x_7, \quad (7.14)$$

$$\frac{dx_8}{dt} = v_5x_5 - (d_6 + K_f)x_8, \quad (7.15)$$

$$\frac{dx_9}{dt} = K_fx_8 - d_6x_9. \quad (7.16)$$

Note that, in this case, the entire network is implemented on three lentiviral vectors, thus in the model we have three G parameters. We identified the parameter region that ensures oscillatory dynamics (Table A.4, time simulation in Figure 7.4 (A)). The time delay was fixed equal to 60 minutes. We found that again, in order to achieve oscillatory behaviour, we need higher efficiency of viral infection of the vectors containing the *tTA* and *PIT* genes with respect of the one the *miR*, and a strong and stiff effect of the silencing. The predicted period of the oscillations is equal to 482 [min].

The most significant continuation results is presented in Figure 7.4 (B). This shows that, due to the presence of a third gene in the negative feedback loop, the time delay can also be null, thus confirming our hypothesis that the proposed re-engineering of the system should be effective to obtain the desired oscillatory dynamics.

7.5 Preliminary *in vivo* data and parameter identification results

In this section, we aim at characterising one part of the circuit, that is the cassette containing the inducible positive feedback loop (*CMV-TET* promoter, responsive to the Tetracycline-controlled transactivator tTA, driving expression of the tTA protein itself, Figure 7.5). We infected HEK 293 cells with

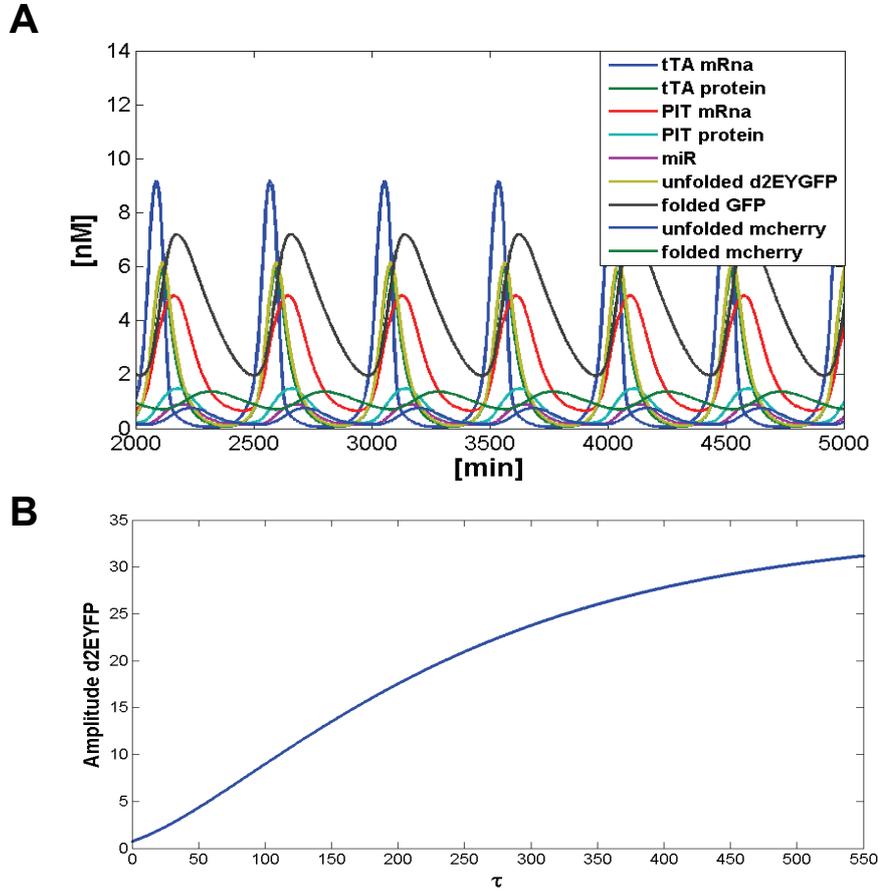


Figure 7.4: **Topology 2, simulation and continuation results.** (A) Model predicted oscillatory dynamics using the parameters reported in Table A.4. (B) Continuation of the Hopf bifurcation on the time delay.

a virus carrying our circuit (Figure 7.5) and imaged them using time-lapse microscopy. We performed two sets of time series experiments. For both the experimental designs, at the first time point, cells were treated with Doxycycline in order to switch off the system. In the first set of experiments (Data set 1) the dynamics were followed for 37h at 37°C (Figure 7.6 (A)), and the data were collected by acquiring an inverted epifluorescence microscope (see Methods section). The second round of experiments (Data set 2) was car-

ried out by lowering the temperature level at 32 °C, as reported [57] in order to limit cell motility and reduce the risk associated to data loss occurring when cells exit the tracked field. In this experimental setup we tested the behaviour of the circuit by treating cells with different amounts of Doxycycline: 100 ng/mL (Figure 7.6 (B)); 1 $\mu g/mL$ (Figure 7.6 (C)); 10 $\mu g/mL$ (Figure 7.6 (D)). The dynamics were followed for 61 h. Details about the experimental procedure, including image acquisition and analysis, are reported in Appendix D.

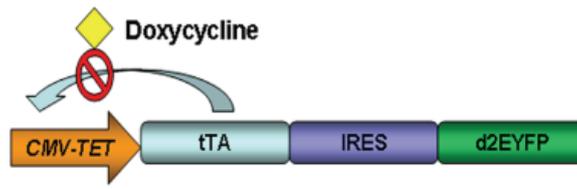


Figure 7.5: Design of the positive feedback loop in mammalian cells.

7.5.1 Mathematical model of the inducible positive feedback loop and parameters fitting

The ODEs model of the cassette containing the positive feedback loop, in agreement with the modelling assumptions presented above, letting x_1 be the $tTA/d2EYFP$ mRNA concentration, x_2 the tTA protein concentration, x_3 the unfolded d2EYFP protein concentration and x_4 the folded d2EYFP

protein concentration, is:

$$\frac{dx_1}{dt} = v_1 \left(\alpha_1 + (1 - \alpha_1) \frac{\left(\frac{\theta}{\theta+D}x_2\right)^{h_1}}{K_1^{h_1} + \left(\frac{\theta}{\theta+D}x_2\right)^{h_1}} \right) - d_1x_1, \quad (7.17)$$

$$\frac{dx_2}{dt} = v_2x_1 - d_2x_2, \quad (7.18)$$

$$\frac{dx_3}{dt} = v_2x_1 - (d_3 + K_f)x_3, \quad (7.19)$$

$$\frac{dx_4}{dt} = K_fx_3 - d_3x_4. \quad (7.20)$$

Of note, as compared to the system of the entire oscillator, the equation of *tTA* mRNA lacks the silencing term.

For the parameter identification, we used the Genetic Algorithm implemented in the Optimization Toolbox of Matlab to minimize the cost function described in Appendix B. In simulations of the “switch off”, the initial values of the metabolites were set to the steady state values predicted by the model in the “on” condition.

The simulations of the fitted model are shown in Figure 7.6, red lines, and the inferred parameters are reported in Table A.5. The inferred model is able to recapitulate the system dynamics in response to different inducer concentrations and experimental settings. We needed to adapt the degradation rate of the reporter protein (d2EYFP) to the different experimental conditions. As mentioned above, in the first set of experiments (Data set 1) the cells were kept at 37°C, while, in the second round of experiments (Data set 2), we used a lower temperature (32°C) in order to limit cell motility. Figure 7.6 (A) and (C) show the response of the system using the same amount of

inducer ($1 \mu\text{g}/\text{mL}$), with cells at 37°C and 32°C , respectively. The dynamics of the “switch off” are faster if the temperature is higher, as the cells metabolism is faster [57]. In the model we captured this behaviour by changing the degradation rate of the reporter protein (parameter d_3 in eq. (7.19) and (7.20)): it was estimated to be lower for Data set 2 as compared to the one fitted using Data set 1, due to the different stability of the protein (Table A.5).

Of note, we varied from the reported literature value is the maximal transcription rate of the *CMV-TET* promoter (Table A.5). The physical meaning is that the strength of the positive-feedback loop is much stronger than previously estimated, at least in the cell-line we used in this experiment (HEK 293). The presence of the auto-regulation is the key to understand the dynamics of the system, because it makes harder for the promoter to be down-regulated by Doxycycline. In Figure 7.6 (E) we analyse how the presence of the positive feedback loop affects the switch-off dynamics: decreasing its strength (green line) or removing it (black line), the system is switched off faster.

7.6 Discussion

In this Chapter, we presented theoretical and preliminary *in vivo* results about the construction of a novel and robust synthetic oscillator in mammalian cells. We highlighted the importance of the mathematical analysis to properly engineer the network. In particular the theoretical analysis showed

that:

- both the oscillator topologies can exhibit oscillatory dynamics;
- for both the topologies the key elements to have oscillations are:
 1. a strong effect of the silencing (the maximum silencing rate must be high, while it is not necessary that the cooperativity coefficient is very higher than its nominal value);
 2. fast dynamics for the positive feedback loop;
 3. slow dynamics for the negative feedback loop (obtainable by adopting Topology 2, or already present in Topology 1 if the silencing process takes enough time). Note that Topology 2 can exhibit oscillations even without an explicit time delay in the silencing term.

The preliminary *in vivo* data-set allowed us to characterise the first cassette of the circuit. Our experimental set-up allows to have data arising from a uniformly infected population of cells, overcoming troubles arising from plasmid transfection techniques. The dynamics of the feedback loop were well defined in the *in vivo* data-set, and correctly captured by the fitted mathematical model. The estimated parameters indicated the dependance of the dynamics on the feedback loop strength. In particular we estimated the positive feedback loop to be faster than how it was estimated in literature; this should ensure us to be in the oscillatory region for the network.

Currently, we aim at characterizing the remaining two cassettes of the circuit. This will allow us to check if the silencing process is strong enough as re-

quired by the mathematical model for both the oscillator topologies, and slow enough for Topology 1 only. Once we will have a complete characterisation of the single network cassettes, we will will perform time-lapse experiments to follow the dynamics of the entire circuit for the both topologies. Finally, it will be of interest to study the problem of synchronization of a population of oscillators.

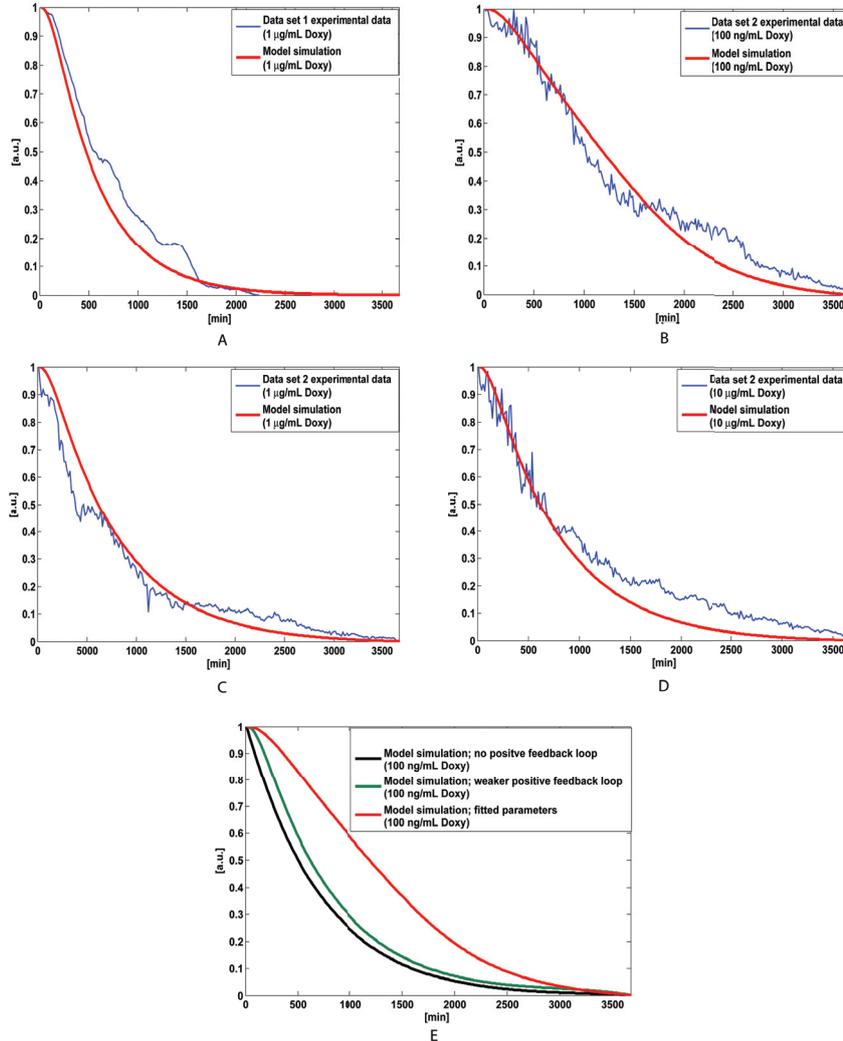


Figure 7.6: **Experimental data and model predictions of the circuit for varying concentrations of Doxycycline (1 $\mu\text{g/mL}$ for (A) and (C), 100 ng/mL for (B) and 10 $\mu\text{g/mL}$ for (D)).** The sample time is equal to 15 min. The cells were treated with the antibiotic at $t=0$ [min]. Model predictions are reported in red while experimental results are represented in blue. In (A) the cells were kept at 37°C and observed up to 37 hours. In (B)-(D) the cells were kept at 32°C and observed up to 61 hours. In (E) we report the comparison of the dynamics of the circuit obtained by varying the strength of the positive feedback loop. Red line=model simulation of the system including the positive feedback loop using the inferred parameter values (Table A.5). Green line=model simulation of the system reducing the strength of the positive feedback loop. Black line=model simulation of the system removing the positive feedback loop.

Chapter 8

Concluding remarks and future work

In this Thesis, modelling and analysis of gene regulatory networks has been discussed. Firstly, we presented an overview about gene regulatory networks and modelling strategies . We then moved to the analysis of a novel synthetic network in yeast. Finally, we presented preliminary results about a synthetic oscillator in mammalian cells.

More specifically, in Chapter 3 we introduced IRMA, the synthetic network we built in yeast to benchmark modelling approaches. In this Chapter we presented the design, topology and construction of the network.

In Chapter 4 we detailed the differential equations based mathematical modelling of IRMA. We illustrated all the steps required: model derivation, experimental design, parameter identification and model validation. Of note, it was necessary to go through iterative refinement steps both in the model and

in the experimental data-set. The mathematical modelling was fundamental to design *ad hoc* experiments to clarify the behaviour of the network. During the modelling process, we needed to simplify some aspects of the model and to increase the level of detail of others, always taking into account the amount and quality of experimental data. The framework we presented can be applied equally well to naturally occurring networks in the cell, thus transforming the drawing of a biological pathway into a computational model.

Chapter 5 dealt with the need to modify a synthetic network after its biological implementation, that is common practice in Synthetic biology. In particular, we showed how to use novel tools from numerical bifurcation theory (e.g. DDE-BIFTOOL [35], able to deal with delayed systems), together with recent results on the link between the dynamics and topology of networks, in order to redesign IRMA. The aim was to understand if and how IRMA could be turned into a robust and tunable synthetic oscillator or a bistable switch. IRMA showed great flexibility. Its topology can be re-engineered in a number of ways in order to achieve the desired dynamical behaviour. We analysed three topologies for the oscillator case and one for the switch case. The major conclusion we drew from our results is that, aiming at constructing a robust and tunable oscillator, the best option is to include in the topology both a delayed negative feedback loop and a fast positive one. In addition, we discovered that, by reducing the topology of the network to a single positive feedback loop, IRMA can be turned into a bistable system. Of note, all the proposed changes are viable *in vivo*.

Chapter 6 focused on the response of the non-oscillatory and the oscillatory

version of the yeast synthetic network to an external periodic input. Such forcing can lead to entrainment, that means that the period of the forced oscillator is exactly the one of the external signal. We analysed the entrainment both via simulation and analytically, using recent contraction theory results. We found that only the non-oscillatory version of the network can be globally entrained to the periodic input. Chaotic attractor can be obtained if the forced system presents autonomous oscillations, depending on the period and duty of oscillations of the input.

Finally, in Chapter 7 we presented the modelling and synthesis of a novel oscillator in mammalian cells based on microRNA and lentiviral infection techniques. The original topology is composed by two genes, and consists of a positive auto-feedback loop and a negative one. By performing bifurcation and continuation analysis of a delay differential equation model, we detected the oscillatory parameter region, and studied the robustness of the oscillator. Our analysis confirmed that the circuit can indeed behave as an oscillator if the dynamics of the positive feedback loop are enough faster than the negative loop ones, and if the silencing is strong. The analysis also suggested to encompass in the negative feedback loop a third gene, in order to slow down its dynamics. Moreover, we presented preliminary *in vivo* data-set that allowed us to characterise the positive feedback loop of the circuit. The dynamics of the feedback loop were well defined in the *in vivo* data-set, and correctly captured by the fitted mathematical model. The estimated parameters indicated the dependence of the dynamics on the positive feedback loop strength. In particular we estimated the positive feedback loop to be faster

than how it was estimated in literature; this should ensure us to be in the oscillatory region for the network.

8.1 Future work

8.1.1 Yeast synthetic network

As showed in Chapter 4, we had some problems in capturing the steady state levels of the genes composing IRMA. We attributed them to the unmodelled effect of protein dynamics, which have been removed from the original model due to the lack of experimental measurements. To further address this issue, we are currently modifying the network in order to include a fluorescent tag for all the genes.

The re-engineering of IRMA in order to turn it into an oscillator currently is under experimental investigation. In particular, following the model's suggestions presented in Chapter 5, we are replacing the *HO* promoter with the *ANB1* promoter, the *ASH1* promoter with the *EGT2* promoter and the *ASH1* gene with the *ROX1* gene. Moreover, we are currently setting up a novel experimental platform based microfluidics [10]. Once the oscillator will be implemented, this will allow to provide it various kinds of periodic inputs, and check *in vivo* the presence of entrainment we analysed *in vitro* in Chapter 6.

8.1.2 Synthetic oscillator in mammalian cells

The first version (we named as Topology 1) of the synthetic oscillator discussed in Chapter 7 has been already synthesized and currently is under experimental investigation. We now aim at characterizing the second cassettes of the circuit (as we reported in Chapter 7, the first cassette, that is the one containing the *tTA* gene, has already been characterised). This will allow us to understand if the silencing process is strong and slow enough as required by the mathematical model to obtain oscillations. Once we will have a complete characterisation of the single network cassettes, we will perform time-lapse experiments to follow the dynamics of the entire circuit, and check the validity of model predictions.

In parallel, we are synthesising the cassette needed to implement the re-engineered version of the oscillator.

Finally, it will be of interest to study the problem of synchronization of a population of oscillators.

Appendix A

Parameters of the mathematical models

In this Appendix we report the parameters of the mathematical models presented in Chapters 4 (derived and identified models of the IRMA), 5 (models of the re-engineering of IRMA in order to turn it into an oscillator, or a bistable switch), 6 (analysis of the response of IRMA to a external periodic input) and 7 (models of the two topologies of the synthetic oscillator in mammalian cells).

Parameter	Model B1	Model B2	Model B3	Model B4	Model C	Model D	Model D ref.	Experim. id.
k_1 [a.u.]	0.329	9.637	1.757	10	1.884	1	1	1
k_2 [a.u.]	8.027	0.002	0.071	0.001	30	0.035	0.035	0.035
k_3 [a.u.]	3.387	0.711	0.886	0.240	0.229	0.037	0.037	0.037
k_4 [a.u.]	4.003	0.853	1.011	0.133	0.216	0.09 Glu 0.01 Gal	0.9 Glu 0.1 Gal	0.09 Glu 0.01 Gal
k_5 [a.u.]	1.750	1.972	7.375	1.313	0.16	1.884	1.884	1.884
k_6 [a.u.]	0.951	0.107	7.191	0.116	0.160	1.884	1.884	1.884
α_1 [a.u. min ⁻¹]	0	0	0	0	0	0	0	—
α_2 [a.u. min ⁻¹]	0	0	0	0	$1.10 \cdot 10^{-4}$	$1.49 \cdot 10^{-4}$	$1.49 \cdot 10^{-4}$	—
α_3 [a.u. min ⁻¹]	0	0	0	0	$3.2 \cdot 10^{-3}$	$3 \cdot 10^{-3}$	$3 \cdot 10^{-3}$	—
α_4 [a.u. min ⁻¹]	0	0	0	0	0	$7.4 \cdot 10^{-4}$	$7.4 \cdot 10^{-4}$	—
α_5 [a.u. min ⁻¹]	0	0	0	0	$7.37 \cdot 10^{-5}$	$6.1 \cdot 10^{-4}$	$6.1 \cdot 10^{-4}$	—
v_1 [a.u. min ⁻¹]	1	1	1	1	0.065	0.04	0.04	—
v_2 [a.u. min ⁻¹]	1	1	1	1	0.002	$8.82 \cdot 10^{-4}$	$8.82 \cdot 10^{-4}$	—
v_3 [a.u. min ⁻¹]	1	1	1	1	0.025	0.002 Glu 0.020 Gal	0.017 Glu 0.155 Gal	$v_{3 \text{ Glu}} / v_{3 \text{ Gal}}$ 9
v_4 [a.u. min ⁻¹]	1	1	1	1	0.007	0.014	0.014	—
v_5 [a.u. min ⁻¹]	1	1	1	1	0.002	0.018	0.018	—
v_{tr} [a.u. min ⁻¹]	—	—	—	—	—	—	0.080	—
d_1 [min ⁻¹]	6.632	9.946	0.964	10	0.033	0.022	0.022	—
d_2 [min ⁻¹]	0.273	1.268	0.013	0.124	0.042	0.047	0.047	—
d_3 [min ⁻¹]	0.109	0.640	0.001	0.297	0.047	0.421	0.590	—
d_4 [min ⁻¹]	1.712	1.335	0.405	2.228	0.141	0.098	0.098	—
d_5 [min ⁻¹]	1.186	8.644	0.133	9.885	0.018	0.050	0.050	—
d_{pr} [min ⁻¹]	—	—	—	—	—	—	0.0144	—
h_1	1	1	1	1	1	1	1	1
h_2	1	1	1	1	1	1	1	1
h_3	1	1	2	2	1	1	1	1
h_4	1	1	2	2	1	4	4	4
h_5	1	1	1	1	1	1	1	1
h_6	1	1	1	1	1	1	1	1
h_7	—	—	—	—	1	4	4	4
K_1 [a.u. ⁻¹ min ⁻¹]	100	100	100	100	—	—	—	—
K_2 [min ⁻¹]	1	1	1	1	—	—	—	—
K_3 [nM ⁻¹ min ⁻¹]	0.1	0.1	0.1	0.1	—	—	—	—
K_4 [min ⁻¹]	1	1	1	1	—	—	—	—
β_1 [min ⁻¹]	—	—	—	—	0.223	0.201	0.201	—
β_2 [min ⁻¹]	—	—	—	—	0.285	0.167	0.167	—
γ [a.u.]	—	—	—	—	10^{-4} Glu 5.55 Gal	0.2 Glu 0.6 Gal	0.2 Glu 0.6 Gal	0.2 Glu 0.6 Gal
τ [min]	—	—	—	—	100	100	100	—
GAL [nM]	$5.55 \cdot 10^7$	$5.55 \cdot 10^7$	$5.55 \cdot 10^7$	$5.55 \cdot 10^7$	—	—	—	—
J (cost function)	4.37	7.951	2.83	6.819	16.79	21.83	22	—

Table A.1 Parameters of the IRMA models.

Parameter	Nominal Value	Scenario 1 (A, B)	Scenario 2 Scenario 3	Scenario 4 (A, B)
k_1 [a.u.]	1	1	—	0.0477 ↓
k_2 [a.u.]	0.035	0.035	0.00035 ↓	—
k_3 [a.u.]	0.037	0.037	0.037	0.037
k_4 [a.u.]	0.01	0.01	0.01	0.01
k_5 [a.u.]	1.884	1.884	1.884	1.884
k_6 [a.u.]	1.884	0.0477 ↓	0.0477 ↓	1.884
α_1 [a.u. min. ⁻¹]	0	0	0	0
α_2 [a.u. min. ⁻¹]	$1.49 \cdot 10^{-4}$	$1.49 \cdot 10^{-4}$	$1.49 \cdot 10^{-4}$	$1.49 \cdot 10^{-4}$
α_3 [a.u. min. ⁻¹]	$3 \cdot 10^{-3}$	$3 \cdot 10^{-3}$	$3 \cdot 10^{-3}$	$3 \cdot 10^{-3}$
α_4 [a.u. min. ⁻¹]	$7.4 \cdot 10^{-4}$	$7.4 \cdot 10^{-4}$	$7.4 \cdot 10^{-4}$	$7.4 \cdot 10^{-4}$
α_5 [a.u. min. ⁻¹]	$6.1 \cdot 10^{-4}$	$6.1 \cdot 10^{-4}$	$6.1 \cdot 10^{-4}$	$6.1 \cdot 10^{-4}$
v_1 [a.u. min. ⁻¹]	0.04	0.04	0.04	0.04
v_2 [a.u. min. ⁻¹]	$8.82 \cdot 10^{-4}$	0.026 ↑ (A) 0.001 ↑ (B)	0.026 ↑	0.026 ↑ (A) $8.82 \cdot 10^{-4}$ (B)
v_3 [a.u. min. ⁻¹]	0.020	0.020	0.020	0.020
v_4 [a.u. min. ⁻¹]	0.014	0.014	0.014	0.014
v_5 [a.u. min. ⁻¹]	0.018	0.018	0.018	0.018
d_1 [min. ⁻¹]	0.022	0.022	0.022	0.022
d_2 [min. ⁻¹]	0.047	0.047	0.047	0.047
d_3 [min. ⁻¹]	0.421	0.421	0.421	0.421
d_4 [min. ⁻¹]	0.098	0.098	0.098	0.098
d_5 [min. ⁻¹]	0.050	0.050	0.050	0.050
h_1	1	1	—	4↑
h_2	1	4↑	1	—
h_3	1	1	4↑	4 (A) 1 (B)
h_4	4	4	4	4
h_5	1	1	1	1
h_6	1	4↑	4↑	1
h_7	4	4	4	4
γ [a.u.]	0.6	0.6	0.6	0.6
τ [min]	100	100	0	0

Table A.2 Parameters of the actual version of IRMA and of the re-engineered versions.

Parameter	Reference Value	Topology 1 Model 1	Topology 1 Model 2
K_1 [nM]	3	3	3
K_2 [nM]	—	0.1	0.3
α_1 [nM min. ⁻¹]	0.085	0.085	0.085
v_1 [nM min. ⁻¹]	0.055	0.055	0.055
v_2 [min. ⁻¹]	0.02	0.02	0.02
v_3 [min. ⁻¹]	0.02	0.02	0.02
K_f [min. ⁻¹]	0.0154	0.0154	0.0154
K_D [min. ⁻¹]	0.00005	0.00005	—
d_1 [min. ⁻¹]	0.0173	0.0173	0.0173
d_2 [min. ⁻¹]	0.0231	0.0231	0.0231
d_3 [min. ⁻¹]	0.054	0.054	0.054
d_4 [min. ⁻¹]	0.0058	0.0058	0.0058
d_5 [min. ⁻¹]	0.00008	0.00008	—
h_1	2	2	2
h_2	—	4	4
λ [min. ⁻¹]	—	2.8	2.8
G_1	—	100	50
G_2	—	5	2
τ [min]	—	—	60
D [nM]	—	0	0

Table A.3 Parameters of the oscillator in mammalian cells, Topology 1.

Parameter	Reference Value	Topology 2 Model 1	Topology 2 Model 2
K_1 [nM]	3	3	3
K_2 [nM]	—	0.1	0.3
K_3 [nM]	3	3	3
α_1 [nM min. ⁻¹]	0.085	0.085	0.085
α_2 [nM min. ⁻¹]	0.056	0.056	0.056
v_1 [nM min. ⁻¹]	0.055	0.055	0.055
v_2 [min. ⁻¹]	0.02	0.02	0.02
v_3 [min. ⁻¹]	0.02	0.02	0.02
v_4 [nM min. ⁻¹]	0.055	0.055	0.055
v_5 [min. ⁻¹]	0.02	0.02	0.02
K_f [min. ⁻¹]	0.0154	0.0154	0.0154
K_D [min. ⁻¹]	0.00005	0.00005	—
d_1 [min. ⁻¹]	0.0173	0.0173	0.0173
d_2 [min. ⁻¹]	0.0231	0.0231	0.0231
d_3 [min. ⁻¹]	0.0173	0.0173	0.0173
d_4 [min. ⁻¹]	0.0658	0.0658	0.0658
d_5 [min. ⁻¹]	0.054	0.054	0.054
d_6 [min. ⁻¹]	0.0058	0.0058	0.0058
d_7 [min. ⁻¹]	0.0008	0.0008	—
h_1	2	2	2
h_2	—	4	4
h_3	2	2	2
λ [min. ⁻¹]	—	2.8	2.8
G_1	—	100	30
G_2	—	5	2
G_3	—	5	2
τ [min]	—	—	60
D [nM]	—	0	0
P [min]	—	0	0

Table A.4 Parameters of the oscillator in mammalian cells, Topology 2.

Parameter	Reference value	Estimated value
K_1 [nM]	3	3
α_1 [nM min ⁻¹]	0.085	0.085
v_1 [nM min ⁻¹]	0.055	0.35
v_2 [min ⁻¹]	0.02	0.02
d_1 [min ⁻¹]	0.017	0.017
d_2 [min ⁻¹]	0.023	0.023
d_3 [min ⁻¹]	-	0.0020 (Data set 1) 0.0014 (Data set 2)
h_1	2	2
θ [nM]	-	90
K_f [min ⁻¹]	0.015	0.015

Table A.5 Parameters of the inducible positive feedback loop in mammalian cells.

Appendix B

Hybrid Genetic Algorithm

B.1 Problem statement

We can formulate the problem of estimating parameter values in the model derived above as a non-linear programming problem (NLP) with differential-algebraic constraints:

$$\begin{aligned} \min_{\gamma} J(x, u; \gamma) \\ \text{s.t. :} \\ \dot{x} = f(x, u; \gamma) \\ g(x(t), u(t); \gamma) \leq 0 \\ h(x(t), u(t); \gamma) = 0 \end{aligned} \tag{B.1}$$

where x is the state vector, u is the vector of inputs acting on the systems (e.g. galactose in our case) and γ is the vector of all parameter to be identified. Moreover, J is an appropriate cost function to be minimized, f is the vector

field of the system dynamics, g and h represent constraints on the variable (e.g. the non-negativeness of the state variables).

B.2 Description of the algorithm

A Genetic Algorithm (GA) randomly initializes a population of individuals (parameters belonging to the space of potential solutions) and then let its members evolve towards better and better regions of the search space by the iterative application of a randomized process of recombination, mutation and selection. An evaluation function, related to the objective function to be minimized, determines the quality of each search point at each iteration step. The population is then ordered according to the evaluation criteria. On the basis of this ordering, a fitness is associated to each individual which determines the probability that the individual will be selected to become the parent of new individuals; the fittest individuals, yielding the best values of the objective function, will have the highest probabilities to be selected.

Although GAs are a quite classical optimization technique in the engineering area, they have not been extensively tested in Biochemistry. A survey on modelling and identification in the framework of molecular biology can be found in [21] and [78]. In particular a hybrid algorithm or HGA is presented that is a combination of GA and non-linear least-square method. The main idea is to merge the global-search properties of GAs with the fast local convergence of Least Square (LS) methods.

Specifically, the HGA works as follows. Let R , N be the sets of real number

and natural number respectively. Let n be the number of parameters to be identified, $I \in R^n$ the hypercube of their possible values; we mark with a bold character each n -tuple in I , as for example $\mathbf{p} = (p_1, \dots, p_n)$, and call it *individual*. Let $S \in I \in R^n$ be the set of the 2^n vertices of I . Moreover:

- $P^*(k) = (\mathbf{p}_1(k), \dots, \mathbf{p}_\mu(k)) \in I^\mu$ is the population at generation k , consisting of μ individuals $\mathbf{p}_i(k) \in I$, with μ a parameter to be chosen;
 - $\bar{P} = (\bar{\mathbf{p}}_1, \dots, \bar{\mathbf{p}}_\rho) \in S^\rho$ is the vertices population, consisting of ρ individuals $\bar{\mathbf{p}}_i \in S$, with ρ chosen as the minimum between the maximum number of vertices (2^n) and a properly fixed percentage of population dimension ($\alpha\mu$), with $\alpha \in [0, 1]$;
 - $P(k) = (\mathbf{p}_1(k), \dots, \mathbf{p}_\mu(k), \bar{\mathbf{p}}_1, \dots, \bar{\mathbf{p}}_\rho) \in I^{\mu+\rho}$ is the enlarged population at generation k on which recombination and mutation operators work. It is obtained by merging the population at generation k (i.e., $P^*(k)$) and the vertices population (\bar{P}), thus consisting of $\mu + \rho$ individuals;
 - $J : I \rightarrow R$ is the evaluate function, in this paper coinciding with the objective function to be optimized;
 - $\beta : I^{\mu+\rho} \rightarrow N^{\mu+\rho}$ is the fitness function providing an estimate of the appropriateness of the population on the basis of the value of function J at each individual;
 - λ is the offspring population size, *i.e.* the number of individuals created at each generation;
-

- $s : I^{\lambda+\mu} \rightarrow I^\mu$ is the selection operator which selects the parent population at the next generation on the basis of the parents and the offspring population at the current generation.

Further we will use two operators, recombination r_{Θ_r} and mutation m_{Θ_m} : the former one produces a new population of λ individuals by randomly combining the characteristics of $\mu + \rho$ individuals; the latter one produces λ individuals by randomly modifying the characteristics of λ individuals. The algorithm starts by defining the initial population, generation 0, which is determined selecting randomly μ individuals $\mathbf{p}_i(0) \in I$, $i = 1, \dots, \mu$, and ρ individuals $\bar{\mathbf{p}}_i \in S$, $i = 1, \dots, \rho$. A fitness is then computed for each member of the initial population. After this initialization, the algorithm can evolve. An inner cycle simulates the evolution of the population using the genetic algorithm: the recombination operator r_{Θ_r} creates λ new individuals from the enlarged population at generation k ; the mutation operator m_{Θ_m} modifies these new individuals obtaining the offspring population; finally the parent population for the next generation is selected and a fitness is associated to each individual. In the outer cycle, after the genetic algorithm is interrupted on the basis of the inner termination criterion, a non-linear least square method is applied to the best individuals of the current population in order to locally optimize the corresponding solutions. The improved individuals are then replaced in the current population. The hybrid algorithm ends when the outer termination criterion returns the true value. The output of the algorithm is then computed applying the non-linear least square procedure to the best individual. The outline of the algorithm is reported below:

```

% Definition of the initial population
let k := 0;
    P*(0) := (p1(0), ..., pμ(0));
    P̄ := (p̄1, ..., p̄ρ);
    P(0) := (P*(0), P̄);
compute β(P(0)) := (β1, ..., βμ+ρ);
% Hybrid algorithm cycle
while not(outer termination criterion) do
    % Genetic algorithm cycle
    while not(inner termination criterion) do
        recombine : P'(k) := rΘr(P(k));
        mutate : P''(k) := mΘm(P'(k));
        evaluate : (P''(k));
        select : P*(k + 1) := s(P''(k), P*(k));
        insert : P(k + 1) := (P*(k + 1), P̄);
        compute β(P(k + 1)) := (β1, ..., βμ+ρ);
        k := k + 1;
    od
    nonlinear_least_square(best individuals);
    put the NLS optimized individuals into the population P(k);
od
% Result
nonlinear_least_square(best individual).

```

Having established the fundamental principles of the GA structure and operational procedure, a more detailed explanation of the algorithm operators is possible.

(a) *Parent Selection.* In each cycle of evolution, a subsequent generation is created from individuals of the current population. This requires that a group of individuals, generally called *parents*, are randomly selected via a specific selection routine based on their own fitness value: the individuals with the best fitness value have the highest probability of becoming parents. In this work a Montecarlo based technique, the *Roulette Wheel Parent Selection Technique*, is used.

(b) *Reproduction Operator.*

- *Recombination*

The recombination operator r_{Θ_r} creates one new individual mixing the characteristics of two or more parents. In our work, the new individual $\mathbf{p}' = (p'_1, \dots, p'_n)$ is generated component-wise from $2n$ parents; in particular for each component i ($i = 1, \dots, n$), a pair of parents, say $p_{S_i,i}$ and $p_{T_i,i}$, are selected and their i -th components blended according to a convex combination

$$p'_i = p_{S_i,i} + \chi_i(p_{T_i,i} - p_{S_i,i}),$$

where $\chi_i \in [0, 1]$ is a uniformly distributed random variable.

- *Mutation*

The mutation operator m_{Θ_m} modifies randomly the components of an individual \mathbf{p}' to obtain another one \mathbf{p}'' . Here we use

$$p_i'' = p_i' + \mathcal{N}(0, \sigma),$$

for $i = 1, \dots, n$, where $\mathcal{N}(0, \sigma)$ is a gaussian stochastic variable with mean 0 and standard deviation σ . We usually let this standard deviation decrease from one generation to the next one in order to induce clustering around the ‘good’ regions of the parameter space.

During the creation of a new generation, recombination and mutation occur with certain probabilities whose choice is a delicate point. Increasing the recombination probability allows the mix of genetic information, but it may also destroy the new information previously introduced by mutation. On the other hand, when the mutation probability increases, the genetic search is transformed into a random one, but this may help re-introduction of lost genetic material. In our approach the recombination probability decreases over generations while the mutation probability increases.

(c) *Selection.*

The algorithm implements a deterministic selection function; the final step in the production of a generation evaluating new individuals through the objective function and then selecting the best μ individ-

uals among the old and the new ones. The selection does not allow duplication in the resulting population.

(d) *Insert Vertices.*

Despite the choice of the mutation operators, an early population clustering may always happen during evolution causing the non-complete exploration of the parameter space. To avoid this drawback, at each evolution step the GA injects into the population fresh information located far away from the clustering by including some of the parameter space vertices as new population members.

(e) *Termination Test.*

Different criteria can be selected for terminating the GA. We usually specify a maximum number of generations.

(f) *Evaluate Function.*

An appropriate evaluate function has to be specified depending on the optimization problem.

Appendix C

Matlab code of the mathematical models

C.1 Models of IRMA

Listing C.1: Model B1 of IRMA

```
1 function dy = model_b1(t,y)
2
3 u= 55500000;
4 K= [0.3297 8.0278 6.6325 3.3874 0.2732 4.0031 0.1091 1.7501 1.7122 0.9515
      1.1861 100 1 0.1000 1];
5
6 % CBF1
7 dy(1)= (y(3)/(K(1)+y(3)))*(K(2)/(K(2)+y(5)))-K(3)*y(1);
8 % GAL4
9 dy(2)= (y(1)/(K(4)+y(1)))-K(5)*y(2);
10 % SWI5
11 dy(3)= (y(2)-y(6))/(K(6)+(y(2)-y(6)))-K(7)*y(3);
12 % GAL80
13 dy(4)= (y(3)/(K(8)+y(3)))-K(9)*y(4);
```

```
14 % ASH1
15 dy(5)= (y(3)/(K(10)+y(3)))-K(11)*y(5);
16 % GalGal80
17 dy(6)= K(12)*(y(2)-y(6))*(y(4)-y(6)-y(7))-K(13)*y(6);
18 % GALGal80
19 dy(7)= K(14)*u*(y(4)-y(6)-y(7))-K(15)*y(7);
20 dy=dy';
21
22 return
```

Listing C.2: Model B2 of IRMA

```

1 function dy = model_b2(t,y)
2
3 u= 55500000;
4 K= [9.6379 0.0022 9.9461 0.7112 1.2683 0.8537 0.6404 1.9727 1.3355 0.1072
      8.6448 100 1 0.1 1];
5
6 % CBF1
7 dy(1)= (y(3)/(K(1)+y(3)))+(K(2)/(K(2)+y(5)))-K(3)*y(1);
8 % GAL4
9 dy(2)= (y(1)/(K(4)+y(1)))-K(5)*y(2);
10 % SWI5
11 dy(3)= (y(2)-y(6))/(K(6)+(y(2)-y(6)))-K(7)*y(3);
12 % GAL80
13 dy(4)= (y(3)/(K(8)+y(3)))-K(9)*y(4);
14 % ASH1
15 dy(5)= (y(3)/(K(10)+y(3)))-K(11)*y(5);
16 % Gal4Gal80
17 dy(6)= K(12)*(y(2)-y(6))*(y(4)-y(6)-y(7))-K(13)*y(6);
18 % GALGal80
19 dy(7)= K(14)*u*(y(4)-y(6)-y(7))-K(15)*y(7);
20 dy=dy';
21
22 return

```

Listing C.3: Model B3 of IRMA

```

1 function dy = model_b3(t,y)
2
3 u= 55500000;
4 K= [1.7572 0.0712 0.9643 0.8867 0.0137 1.0117 0.0019 7.3755 0.4056 7.1914
      0.1339 100 1 0.1 1];
5
6 % CBF1
7 dy(1)= (y(3)/(K(1)+y(3)))*(K(2)/(K(2)+y(5)))-K(3)*y(1);
8 % GAL4
9 dy(2)= (y(1)^2/(K(4)^2+y(1)^2))-K(5)*y(2);
10 % SWI5
11 dy(3)= (y(2)-y(6))^2/(K(6)^2 +(y(2)-y(6))^2)-K(7)*y(3);
12 % GAL80
13 dy(4)= (y(3)/(K(8)+y(3)))-K(9)*y(4);
14 % ASH1
15 dy(5)= (y(3)/(K(10)+y(3)))-K(11)*y(5);
16 % Gal4Gal80
17 dy(6)= K(12)*(y(2)-y(6))*(y(4)-y(6)-y(7))-K(13)*y(6);
18 % GALGal80
19 dy(7)= K(14)*u*(y(4)-y(6)-y(7))-K(15)*y(7);
20 dy=dy';
21
22 return

```

Listing C.4: Model B4 of IRMA

```

1 function dy = model_b4(t,y)
2
3 u= 55500000;
4 K= [9.999 0.0016 9.9999 0.2401 0.1241 0.1333 0.2975 1.3132 2.2287 0.1167
      9.8858 100 1 0.1 1];
5
6 % CBF1
7 dy(1)= (y(3)/(K(1)+y(3)))+(K(2)/(K(2)+y(5)))-K(3)*y(1);
8 % GAL4
9 dy(2)= (y(1)^2/(K(4)^2+y(1)^2))-K(5)*y(2);
10 % SWI5
11 dy(3)= (y(2)-y(6))^2/(K(6)^2+(y(2)-y(6))^2)-K(7)*y(3);
12 % GAL80
13 dy(4)= (y(3)/(K(8)+y(3)))-K(9)*y(4);
14 % ASH1
15 dy(5)= (y(3)/(K(10)+y(3)))-K(11)*y(5);
16 % Gal4Gal80
17 dy(6)= K(12)*(y(2)-y(6))*(y(4)-y(6)-y(7))-K(13)*y(6);
18 % GALGal80
19 dy(7)= K(14)*u*(y(4)-y(6)-y(7))-K(15)*y(7);
20 dy=dy';
21
22 return

```

Listing C.5: Model C of IRMA

```

1 function dy = model_c(t,y,Z)
2
3 g= 1*10^-2;
4 u= 5.55*10^2*( (sign(t-3010)+1)/2 - (sign(t-7000)+1)/2 ) ;
5 K= [1.8840 30 0.0523 0.2294 14.6060 0.2160 2.8550 0.1600 19.6000 0.1602
      9.7484 77.7600 39.6000 0.1031 1.1615 0.6378 0.0028 0.0167 1.5120 0.0019
      0 0.0384 0.0196 0.0072 0 0.9896 100 0.8 1 1];
6
7 % CBF1
8 dy(1,1)= K(16)*(K(14)*(Z(3)^K(26)/((K(1)^K(26)+Z(3)^K(26))*(1+(y(5)^K(29)/K
      (2)))))-K(3)*y(1)+K(21));
9 % GAL4
10 dy(2,1)= K(17)*(K(30)*(y(1)/(K(4)+y(1)))-(K(5)-(K(12)*((sign(t-3000)-1)/2-(
      sign(t-3010)-1)/2)))*y(2)+K(22));
11 % SWI5
12 dy(3,1)= K(18)*(K(19)*(y(2)/((1+((K(27)*y(4))/(g+u))*K(6)+y(2))))-K(7)*y
      (3)+K(23));
13 % GAL80
14 dy(4,1)= K(24)*((y(3)^K(28)/(K(8)^K(28)+y(3)^K(28)))-(K(9)-(K(13)*((sign(t
      -3000)-1)/2-(sign(t-3010)-1)/2)))*y(4)+K(25));
15 % ASH1
16 dy(5,1)= K(20)*(K(15)*(y(3)^K(28)/(K(10)^K(28)+y(3)^K(28)))-K(11)*y(5)+K(22)
      );
17
18 end

```

Listing C.6: Model D of IRMA

```

1 function dy=model_d(t,y,Z)
2
3 % Time delay = 100[min]
4 K= [0 0.0404 1 0.0356 0.0221 0.0001 0.0008 0.0372 0.0477 0.2013 0.0029
      0.0022 0.2000 0.0937 0.4216 0.0007 0.0146 1.8140 0.0980 0.1676 0.0006
      0.0181 1.8140 0.0500 9 3 9];
5 % input (galactose) is given from t=3010 to t=6000.
6 u= ((sign(t-3010)+1)/2 -(sign(t-6000)+1)/2);
7
8 % CBF1
9 dy(1,1)= K(1)+ K(2)*(Z(3)/((K(3)+Z(3))*(1+(y(5)/K(4)))))-K(5)*y(1);
10 % GAL4
11 dy(2,1)= (K(6)+K(7)*(y(1)/(K(8)+y(1)))-(K(9)-(K(10)*((sign(t-3000)-1)/2-(
      sign(t-3010)-1)/2)))*y(2);
12 % SWI5 (note that the values of 3 parameters change depending on the
13 % medium)
14 dy(3,1)= K(11)+(K(12)*(1-u)+u*(K(12)*K(25)))*(y(2).^4./((K(14)*(1-u)+u*(K
      (14)/K(27))).^4+y(2).^4.*(1+(y(4).^4./((K(13)*(1-u)+u*(K(13)*K(26))).^4)
      ))))-K(15)*y(3);
15 % GAL80
16 dy(4,1)= K(16)+K(17)*(y(3)/(K(18)+y(3)))-(K(19)-(K(20)*((sign(t-3000)-1)/2-(
      sign(t-3010)-1)/2)))*y(4);
17 % ASH1
18 dy(5,1)= K(21)+ K(22)*(y(3)/(K(23)+y(3)))-K(24)*y(5);
19
20 end

```

C.2 Models of the re-engineered versions of IRMA

Listing C.7: Oscillator model Scenario 1 A

```

1 function dy = scenario1a(t,y,Z)
2
3 % Time delay = 100 [min]
4 K= [0 0.0404 1 0.0356 0.0221 0.0001 0.0240 0.0372 0.0477 0.2013 0.0029
      0.0020 0.2000 0.0937 0.4210 0.0007 0.0146 1.8140 0.0980 0.0167 0.0006
      0.0181 0.0477 0.0500 9 3 9];
5 u= 1;
6
7 % CBF1
8 dy(1,1)= K(1)+ K(2)*(Z(3)/((K(3))+Z(3)))*((K(4)*1)^4/((K(4)*1)^4+y(5)^4))-K
      (5)*y(1);
9 % GAL4
10 dy(2,1)= K(6)+K(7)*(y(1)/(K(8)+y(1)))-K(9)*y(2);
11 % SWI5
12 dy(3,1)= K(11)+(K(12)*(1-u)+u*(K(12)*K(25)))*(y(2)^4/((K(14)*(1-u)+u*(K(14)
      )/K(27)))^4+y(2)^4))-K(15)*y(3);
13 % GAL80
14 dy(4,1)= K(16)+K(17)*(y(3)/(K(18)+y(3)))-K(19)*y(4);
15 % ASH1
16 dy(5,1)= K(21)+ K(22)*(y(3)^4/(K(23)^4+y(3)^4))-K(24)*y(5);
17
18 end

```

Listing C.8: Oscillator model Scenario 1 B

```

1 function dy = scenario1b(t,y,Z)
2
3 % Time delay = 100 [min]
4 K= [0 0.0404 1 0.0356 0.0221 0.0001 0.0015 0.0372 0.0477 0.2013 0.0029
      0.0020 0.2000 0.0937 0.4210 0.0007 0.0146 1.8140 0.0980 0.0167 0.0006
      0.0181 0.0477 0.0500 9 3 9];
5 u= 1;
6
7 % CBF1
8 dy(1,1)= K(1)+ K(2)*(Z(3)/((K(3))+Z(3)))*((K(4)*1)^4/((K(4)*1)^4+y(5)^4))-K
      (5)*y(1);
9 % GAL4
10 dy(2,1)= K(6)+K(7)*(y(1)/(K(8)+y(1)))-K(9)*y(2);
11 % SWI5
12 dy(3,1)= K(11)+(K(12)*(1-u)+u*(K(12)*K(25)))*(y(2)^4/((K(14)*(1-u)+u*(K(14)
      )/K(27)))^4+y(2)^4))-K(15)*y(3);
13 % GAL80
14 dy(4,1)= K(16)+K(17)*(y(3)/(K(18)+y(3)))-K(19)*y(4);
15 % ASH1
16 dy(5,1)= K(21)+ K(22)*(y(3)^4/(K(23)^4+y(3)^4))-K(24)*y(5);
17
18 end

```

Listing C.9: Oscillator model Scenario 2

```

1 function dy = scenario2(t,y)
2
3 K= [0 0.0404 1 0.00035 0.0221 0.0001 0.0260 0.0372 0.0477 0.2013 0.0029
      0.0020 0.2000 0.0937 0.4210 0.0007 0.0146 1.8140 0.0980 0.0167 0.0006
      0.0181 0.0477 0.0500 9 3 9];
4 u= 1;
5
6 % CBF1
7 dy(1,1)= K(1)+ K(2)*(K(4)/(K(4)+y(5)))-K(5)*y(1);
8 % GAL4
9 dy(2,1)= K(6)+K(7)*(y(1)^4/(K(8)^4+y(1)^4))-K(9)*y(2);
10 % SWI5
11 dy(3,1)= K(11)+(K(12)*(1-u)+u*(K(12)*K(25)))*(y(2)^4/((K(14)*(1-u)+u*(K(14)
      )/K(27)))^4+y(2)^4*(1+(y(4)^4/(K(13)*(1-u)+u*(K(13)*K(26)))^4)))))-K
      (15)*y(3);
12 % GAL80
13 dy(4,1)= K(16)+K(17)*(y(3)/(K(18)+y(3)))-K(19)*y(4);
14 % ASH1
15 dy(5,1)= K(21)+K(22)*(y(3)^4/(K(23)^4+y(3)^4))-K(24)*y(5);
16
17 end

```

Listing C.10: Oscillator model Scenario 3

```

1 function dy = scenario3(t,y)
2
3 K= [0 0.0404 1 0.00035 0.0221 0.0001 0.0260 0.0372 0.0477 0.2013 0.0029
      0.0020 0.2000 0.0937 0.4210 0.0007 0.0146 1.8140 0.0980 0.0167 0.0006
      0.0181 0.0477 0.0500 9 3 9];
4 u= 1;
5
6 % CBF1
7 dy(1,1)= K(1)+ K(2)*(K(4)/(K(4)+y(5)))-K(5)*y(1);
8 % GAL4
9 dy(2,1)= K(6)+K(7)*(y(1)^4/(K(8)^4+y(1)^4))-K(9)*y(2);
10 % SWI5
11 dy(3,1)= K(11)+(K(12)*(1-u)+u*(K(12)*K(25)))*(y(2)^4/((K(14)*(1-u)+u*(K(14)
      )/K(27)))^4+y(2)^4*(1+(y(4)^4/((K(13)*(1-u)+u*(K(13)*K(26)))^4)))))+ K
      (22)*(y(3)^4/(K(23)^4+y(3)^4))-K(15)*y(3);
12 % GAL80
13 dy(4,1)= K(16)+K(17)*(y(3)/(K(18)+y(3)))-K(19)*y(4);
14 % ASH1
15 dy(5,1)= K(21)+K(22)*(y(3)^4/(K(23)^4+y(3)^4))-K(24)*y(5);
16
17 end

```

Listing C.11: Bistable switch model Scenario 4 A

```

1 function dy = scenario4a(t,y)
2
3 K=[0 0.0404 0.0477 0.0356 0.0221 0.0001 0.0260 0.0372 0.0477 0.2013 0.0029
      0.0020 0.2000 0.0937 0.4210 0.0007 0.0146 1.8140 0.0980 0.0167 0.0006
      0.0181 0.0477 0.0500 9 3 9];
4 u=1;
5
6 % CBF1
7 dy(1,1)= K(1)+ K(2)*(y(3)^4/((K(3))^4+y(3)^4))-K(5)*y(1);
8 % GAL4
9 dy(2,1)= K(6)+K(7)*(y(1)/(K(8)+y(1)))-K(9)*y(2);
10 % SWI5
11 dy(3,1)= K(11)+(K(12)*(1-u)+u*(K(12)*K(25)))*(y(2)^4/((K(14)*(1-u)+u*(K(14)
      )/K(27)))^4+y(2)^4))-K(15)*y(3);
12 % GAL80
13 dy(4,1)= K(16)+K(17)*(y(3)/(K(18)+y(3)))-K(19)*y(4);
14 % ASH1
15 dy(5,1)= K(21)+ K(22)*(y(3)^4/(K(23)^4+y(3)^4))-K(24)*y(5);
16
17 end

```

Listing C.12: Bistable switch model Scenario 4 B

```

1 function dy = scenario4b(t,y)
2
3 K= [0 0.0404 0.0477 0.0356 0.0221 0.0001 0.0008 0.0372 0.0477 0.2013 0.0029
      0.0020 0.2000 0.0937 0.4210 0.0007 0.0146 1.8140 0.0980 0.0167 0.0006
      0.0181 0.0477 0.0500 9 3 9];
4 u= 1;
5
6 % CBF1
7 dy(1,1)= K(1)+ K(2)*(y(3)^4/((K(3))^4+y(3)^4))-K(5)*y(1);
8 % GAL4
9 dy(2,1)= K(6)+K(7)*(y(1)/(K(8)+y(1)))-K(9)*y(2);
10 % SWI5
11 dy(3,1)= K(11)+(K(12)*(1-u)+u*(K(12)*K(25)))*(y(2)^4/((K(14)*(1-u)+u*(K(14)
      )/K(27)))^4+y(2)^4))-K(15)*y(3);
12 % GAL80
13 dy(4,1)= K(16)+K(17)*(y(3)/(K(18)+y(3)))-K(19)*y(4);
14 % ASH1
15 dy(5,1)= K(21)+ K(22)*(y(3)^4/(K(23)^4+y(3)^4))-K(24)*y(5);
16
17 end

```

C.3 Models of the synthetic oscillator in mammalian cells

Listing C.13: Oscillator in mammalian cells, Topology 1, DDEs model

```

1 function dy = topology1(t,y,Z)
2
3 % To reproduce simulation results presented in the Thesis, fix the delay to
4 % 60 [min].
5 D= 0;
6 K= [0.0550 0.0850 2 3 0.01730 2.8500 4 0.3000 0.0200 0.0231 0.0540 0.0500
      0.0154 0.0058 0.00005];
7 G= [50 2];
8
9 % TTA mRNA
10 dy(1,1)= G(1)*K(1)*(K(2)+(1-K(2))*(((K(12)/(K(12)+D))*y(2))^K(3)/(K(4)+((K
      (12)/(K(12)+D))*y(2))^K(3))))-K(5)*y(1)-K(6)*(Z(3)^K(7)/(K(8)^K(7)+Z(3)^
      K(7)))*y(1);
11 % TTA protein
12 dy(2,1)= K(9)*y(1)-K(10)*y(2);
13 % MIR
14 dy(3,1)= G(2)*K(1)*(K(2)+(1-K(2))*(((K(12)/(K(12)+D))*y(2))^K(3)/(K(4)+((K
      (12)/(K(12)+D))*y(2))^K(3))))-(K(11)+K(15))*y(3);
15 % Unfolded dGFP
16 dy(4,1)= K(9)*y(1)-K(13)*y(4)-K(14)*y(4);
17 % Folded dGFP
18 dy(5,1)= K(13)*y(4)-K(14)*y(5);
19 % Unfolded mcherry
20 dy(6,1)= K(9)*y(3)-K(13)*y(6)-K(14)*y(6);
21 % Folded mcherry
22 dy(7,1)= K(13)*y(6)-K(14)*y(7);
23
24 end

```

Listing C.14: Oscillator in mammalian cells, Topology 2, DDEs model

```

1 function dy = topology2(t,y,Z)
2
3 % To reproduce simulation results presented in the Thesis, fix the delay to
4 % 60 [min].
5 D= 0;
6 K= [0.0550 0.0850 2 3 0.01730 2.850 4 0.3000 0.0200 0.0231 0.0540 0.0500
      0.0154 0.0058 0.0173 0.0200 0.0658 0.0550 0.0560 3 0.00005];
7 G= [30 2 2];
8
9 % TTA mRNA
10 dy(1,1)= G(1)*K(1)*(K(2)+(1-K(2)))*(((K(12)/(K(12)+D))*y(2))^K(3)/(K(4)+((K
      (12)/(K(12)+D))*y(2))^K(3)))-K(5)*y(1)-K(6)*(Z(5)^K(7)/(K(8)^K(7)+Z(5)^
      K(7)))*y(1);
11 % TTA protein
12 dy(2,1)= K(9)*y(1)-K(10)*y(2);
13 % PIT mRNA
14 dy(3,1)= G(2)*K(1)*(K(2)+(1-K(2)))*(((K(12)/(K(12)+D))*y(2))^K(3)/(K(4) +((K
      (12)/(K(12)+D))*y(2))^K(3)))-K(15)*y(3);
15 % PIT protein
16 dy(4,1)= K(16)*y(3)-K(17)*y(4);
17 % MIR
18 dy(5,1)= G(3)*K(18)*(K(19)+(1-K(19)))*(((K(12)/(K(12)+D))*y(4))^K(3)/(K(20)
      +((K(12)/(K(12)+D))*y(4))^K(3)))-((K(11)+K(21))*y(5));
19 % Unfolded dGFP
20 dy(6,1)= K(9)*y(1)-K(13)*y(6)-K(14)*y(6);
21 % Folded dGFP
22 dy(7,1)= K(13)*y(6)-K(14)*y(7);
23 % Unfolded mcherry
24 dy(8,1)= K(9)*y(5)-K(13)*y(8)-K(14)*y(8);
25 % Folded mcherry
26 dy(9,1)= K(13)*y(8)-K(14)*y(9);
27
28 end

```

Listing C.15: Positive feedback loop in mammalian cells

```

1  function dy = positiveloop(t,y)
2
3  % Note that the value of Doxycycline must be changed depending on the
4  % experiment. The parameters values reported here are valid for Data set 2;
5  % for Data set 1, parameter K(15) is equal to
6  % 0.002.
7  D= 2.25*10^3*( (sign(t-3000)+1)/2 - (sign(t-8000)+1)/2 ) ;
8  K= [0.0550 0.0850 2 3 0.0173 0 0 0 0.0200 0.0231 0.0540 0.0500 0.0154 0.0058
9      0.3500 90 0.0014];
10 % TTA mRNA
11 dy(1,1)= K(15)*(K(2)+(1-K(2))*(((G(2)/(G(2)+D))*y(2))^K(3)/(K(4)^K(3)+((K
12      (16)/(K(16)+D))*y(2))^K(3))))-K(5)*y(1);
13 % TTA protein
14 dy(2,1)= K(9)*y(1)-K(10)*y(2);
15 % Unfold dEYFProtein
16 dy(3,1)= K(9)*y(1)-K(13)*y(3)-K(17)*y(3);
17 % Fold dEYFProtein
18 dy(4,1)= K(13)*y(3)-K(17)*y(4);
19 end

```

Appendix D

Experimental procedures

D.1 Experiments performed on the yeast synthetic network

D.1.1 Construction of *S. cerevisiae* strains

To construct the IRMA containing strain, sequential PCR-based genomic integrations were made with the cassettes described in the text below. All the integrations were confirmed by PCR.

At first, two HA epitopes were cloned in pAG32 [43] among Hind III and Bgl II sites. The 2xHA-hphMX4 cassette was amplified by PCR and inserted in front of the stop codon of *ASH1* gene in YM4271 strain resulting in P278 strain.

To generate P280 strain *MET16* promoter (−446 to −1, ATG = +1) was amplified from W303 and cloned in YIplac128 between Hind III and Sac

I; *GAL4*ORF was then cloned between Sac I and Nde I thus resulting in p*MET16pGAL4*. The *MET16pGAL4-LEU2* cassette was integrated in *SHE2* locus (-11 to +751).

*CBF1*ORF was amplified from W303 and cloned among Bam HI and Pac I of pFA6a-GFP(S65T)-kanMX6 [113]. Then, the *CBF1-GFP-kanMX6* cassette was integrated downstream of the *HO* promoter (between -1 to +1758) of P280 strain, obtaining P324.

ASH1 promoter (-591 to -1) was cloned in Pst I and Bam HI of YIplac211 where the *GAL80-3xFLAG* was then inserted between Bam HI and Sac I. The *ASH1pGAL80-3XFLAG-URA3* was integrated in *SWI5* locus (-50 to +2299) thus yielding P326. In this strain, *ACE2* gene was also deleted (from -345 to +2314) by integrating natMX4 cassette from pAG25 [43].

To build *GAL10pSWI5_{AAA}-MYC9-KITRP1*, the *SWI5_{AAA}* locus was tagged at the C-terminus with nine Myc epitopes in K2072 strain that was kindly provided by K. Nasmyth [79]. *SWI5_{AAA} -MYC9-KITRP1* was then amplified by PCR from the resulting strain and cloned in YIplac204 between Eco RI and Aat II. The *GAL10* promoter (-523 to -1) was cloned in YIplac204 between Hind III and Eco RI yielding the vector containing the entire integrated cassette.

Finally, *GAL10pSWI5_{AAA}-MYC9-KITRP1* was integrated in *CBF1* locus (-1 to +1464) resulting in IRMA containing strain (P340).

Strains used for promoter strength measurements were constructed by inte-

grating the “ad hoc” built promoters containing cassettes. The *kanMX4 – MET25p* cassette was amplified by PCR from plasmid pYM-N34 (EUROSCARF) and integrated upstream of the starting codon of *GAL4* (in P265, a wild type strain), and upstream of the ATG of *ASH1* and *SWI5* (in P358, a *she2Δace2Δ* strain) to obtain respectively P549, P362 and P364 strains.

In order to obtain strains which express the *CBF1* TF at different levels, we integrated at the 5’ of this gene constitutive promoters of variable strength (*CYC1*, *ADH1*, *TEF*, *GPD* promoters) and the *CUP1* inducible promoter, obtaining P353, P351, P360, P354 and P365 strains, respectively. The promoters were amplified (together with the *kanMX4* resistance cassette) from plasmids pYM-N10, pYM-N6, pYM-N18, pYM-N14, pYM-N1 (EUROSCARF). The promoters of the network, chosen in such a way that for each of them a single transcription factor (TF) is sufficient and essential to activate transcription, were assembled upstream of non-self gene coding sequences. Further details can be found in [15]. All data presented refer to mRNA levels of the five IRMA genes and were measured by quantitative real-time RT-PCR (q-PCR).

D.1.2 Data collection

For the preliminary “switch-on” data-set (used for the identification of Models B1, B2, B3 and B4), we collected samples every 20 minutes up to 5 hours in four independent experiments and we measured mRNA levels of the five IRMA genes by quantitative real-time RT-PCR (q-PCR). Out of the four time-series, three were 3 hours long, and one lasted 5 hours. The averaged

data-set is shown in Fig. 4.2 (A) and Fig. 4.3. We then performed one additional 5 hour “switch-on” time-series. The experimental set up is identical, but we included as the first point of the time-series the expression level of the network genes after growing cells overnight in glucose. The new averaged “switch-on” data-set was used for the identification of Models C and D and is shown in Fig. 4.2 (B).

The “switch-off” data-set (Fig. 4.2 (C)) is the average of 4 experiments performed by shifting cells from galactose to glucose and collecting samples every 10 minutes up to 3 hours [15].

For the “Galactose steady-state” and “Glucose steady-state” data-set (Fig. 4.4 (A), (C)), the over-expression of each gene was performed in cells grown either in glucose, or in galactose. The steady-state expression levels of IRMA genes were measured by q-PCR.

For the promoters data-set, each of the transcription factor genes was stably expressed at different levels in a wild-type strain, and the transcription of the corresponding promoter genes were measured by q-PCR at steady-state, for a total of 165 data points [15].

D.2 Preliminary experiments performed on the synthetic oscillator in mammalian cells

D.2.1 Construction of the inducible positive feedback loop

To implement the gene circuit in a lentiviral vector, we used the ViraPower Promotrless 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 synthesized 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 and a reverse primer containing an EcoRV recognition sequence. 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 linearized 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 linearized 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. The lentivirus was then produced in 293FT cells as described in the instructions provided by Invitrogen.

D.2.2 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). Hek 293 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 and 1% antibiotic/antimycotic solution (GIBCO BRL).

To transduce cells with the virus produced, 500000 HEK293 cells were plated and incubated overnight. The day of transduction the medium was removed and 1mL of the virus was added to the cells together with polybrene (In-

vitrogen) to a final concentration of 6 μ g/mL. After an overnight incubation the medium containing the virus was removed and replaced with complete culture medium containing the blasticidin (Sigma) to a final concentration of 3 μ g/mL to select for stably transduced cells. For the switch off experiment 500 stably-integrated-HEK293 cells were plated in chamber slide (lab-Tek) and treated with Doxycycline (Clontech) to a final concentration ranging from 100 ng/mL to 10 μ g/mL).

Image acquisition and analysis

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 photo-bleaching while optimizing 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

built in Autofocus routine was employed 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.

Image segmentation was carried out in Mathworks Matlab R2008b (Mathworks Inc.); the algorithm we used to quantify fluorescence was meant to distinguish the foreground (living cells) from the background in the phase-contrast field. A new binary image was then built in order to find areas of the fluorescence images where cells were located and compute the mean d2EYFP intensity over those pixels only.

D.2.3 Data processing

Data from the quantification algorithm have been processed in order to reduce the impact of noise on the following steps. In particular a Savitzky-Golay smoothing filter [85] has been applied to the fluorescence signals of Figure 3 (A) using a first order interpolating polynomial and a 15 samples window. Data were further normalised in the range $[0, 1]$ in order to standardise the fitting process.

Bibliography

- [1] Laurie Ailles and Luigi Naldini. Hiv-1-derived lentiviral vectors. *Curr.Top. Microbiol. Immunol*, 261:31–52, 2002.
 - [2] 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.
 - [3] Alexander Anders, Hauke Lilie, Kathlen Franke, Lutz Kapp, Jörg Stelling, Ernst D. Gilles, and Karin D. Breunig. The galactose switch in *kluveromyces lactis* depends on nuclear competition between gal4 and gal1 for gal80 binding. *Journal of Biological Chemistry*, 281(39):29337–29348, 2006.
 - [4] James Anderson and Antonis Papachristodoulou. On validation and invalidation of biological models. *BMC Bioinformatics*, 10, 2009.
 - [5] David Angeli, James E. Ferrell, and Eduard D. Sontag. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *PNAS*, 101:1822–1827, 2004.
 - [6] Sylvain Arlot and Alain Celisse. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79, 2010.
 - [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*, 113:597–607, 2003.
-

-
- [8] Grgory Batt, Delphine Ropers, Hidde De Jong, Johannes Geiselman, Radu Mateescu, and Dominique Schneider. Validation of qualitative models of genetic regulatory networks by model checking: Analysis of the nutritional stress response in escherichia coli. *Bioinformatics*, 21:19–28, 2005.
- [9] Lewin Benjamin. *Genes*. Jones and Bartlett, 2007.
- [10] Matthew R. Bennett, Wyming L. Pang, Natalie A. Ostroff, Bridget L. Baumgartner, Sujata Nayak, Lev S. Tsimring, and Jeff. Hasty. Metabolic gene regulation in a dynamically changing environment. *Nature*, 454:1119–1122, 2008.
- [11] Leena T. Bhoite, Yaxin Yu, and David J. Stillman. The swi5 activator recruits the mediator complex to the ho promoter without rna polymerase II. *Genes Dev*, 15:2457–2469, 2001.
- [12] Olga Boubaker and Ada Fourati. Structural identifiability of non linear systems : An overview. *Industrial Technology, 2004, IEEE ICIT*, 3:1224–1248, 2004.
- [13] Cullen Bryan R. Rnai the natural way. *Nature*, 37:1163 – 1165, 2005.
- [14] Laurence Calzone, Denis Thieffry, John J. Tyson, and Bela Novak. Dynamical modeling of syncytial mitotic cycles in Drosophila embryos. *Mol Sys Bio*, 3:131–141., 2007.
- [15] Irene Cantone, Lucia Marucci, Francesco Iorio, Maria A Ricci, Vincenzo Belcastro, Mukesh Bansal, Mario di Bernardo, Stefania Santini, Diego di Bernardo, and Maria Pia Cosma. A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches. *Cell*, 137:171–181, 2009.
- [16] Richard W. Carthew and Erik J. Sontheimer. Origins and Mechanisms of miRNAs and siRNAs. *Cell*, 136(4):642–655, February 2009.
- [17] Srikrupa Chandrasekaran, Andrew E. Deffenbaugh, David A. Ford, Eric Bailly, Neal Mathias, and Dorota Skowyra. Destabilization of binding to cofactors and SCF^{Met30} is the ratelimiting regulatory step in degradation of polyubiquitinated Met4. *Molecular cell*, 24:689–699, 2006.
-

-
- [18] 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.
- [19] Robert Copeland. *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis, 2nd Edition*. Wiley-VCH, New York, 2000.
- [20] Maria Pia Cosma, Tomoyuki Tanaka, and Kim Nasmyth. Ordered Recruitment of Transcription and Chromatin Remodeling Factors to a Cell Cycle and Developmentally Regulated Promoter. *Cell*, 97:299–311, 1999.
- [21] Edmund J. Crampin, Patrick E. McSharry, and Santiago Schnell. Extracting biochemical reaction kinetics from time series data. *Knowledge-Based Intelligent Information and Engineering Systems, Pt 2, Proceedings*, 3214:329–336, 2004.
- [22] Marie E. Csete and John C. Doyle. Reverse Engineering of Biological Complexity. *Science*, 295(5560):1664–1669, 2002.
- [23] Giulia Cuccato, Giusy Della Gatta, and Diego di Bernardo. Systems and synthetic biology: tackling genetic networks and complex diseases. *Heredity*, 102:527532, 2009.
- [24] Hazel M Davey and Douglas B Kell. Flow cytometry and cell sorting of heterogeneous microbial populations—the importance of single-cell analyses. *Microbiological Reviews*, 60:641–696, 1996.
- [25] Hidde de Jong. Modeling and simulation of genetic regulatory systems: A literature review. *J. Comp. Biol.*, 9:67–103, 2002.
- [26] Hidde De Jong. Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 9:67–103, 2002.
- [27] Domitilla Del Vecchio. Design and Analysis of an Activator-Repressor Clock in *E. Coli*. *American Control Conference, 2007. ACC '07*, 2007.
- [28] Diego di Bernardo, Michael J. Thompson, Timothy S. Gardner, Sarah E. Chobot, Erin L. Eastwood, Andrew P. Wojtovich, Sean J. Elliott, Scott E. Schaus, and
-

- James J. Collins. Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. *Nature Biotechnology*, 23(3):377–383, March 2005.
- [29] Barbara Di Ventura, Caroline Lemerle, Kostantinos Michalodimitrakis, and Luis Serrano. From *in vivo* to *in silico* biology and back. *Nature*, 443:527–533, 2006.
- [30] Markus Eiswirth, Albrecht Freund, and John Ross. Operational procedure toward the classification of chemical oscillators. *J. Phys. Chem*, 95:1294–99, 1991.
- [31] Tom Ellis, Xiao Wang, and James J. Collins. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat Biotech*, 27(5):465–471, 2009.
- [32] Michael B. Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, January 2000.
- [33] Lukas Endler, Nicolas Rodriguez, Nick Juty, Vijayalakshmi Chelliah, Camille Laibe, Chen Li, and Nicolas Le Novère. Designing and encoding models for synthetic biology. *Journal of the Royal Society, Interface / the Royal Society*, 6 Suppl 4(Suppl 4):S405–S417, 2009.
- [34] Koen Engelborghs, Tatyana Luzyanina, Karel J. in t Hout, and Dirk Roose. Collocation methods for the computation of periodic solutions of delay differential equations. *SIAM J. Sci. Comput.*, 22:1593–1609, 2001.
- [35] Koen Engelborghs, Tatyana Luzyanina, and Giovanni Samaey. DDE-BIFTOOL v. 2.00: a Matlab package for numerical bifurcation analysis of delay differential equations. *Report TW, vd330, Department of Computer Science, K.U. Leuven, Leuven, Belgium*, 2001.
- [36] Christopher P Fall, Eric S Marland, John M Wagner, and John J Tyson. *Computational cell biology*. New York: Springer-Verlag, 2002.
- [37] Andrew Fire, Siqun Xu, Mary K. Montgomery, Steven A. Kostas, Samuel E. Driver, and Craig C. Mello. Potent and specific genetic interference by double-stranded rna in *caenorhabditis elegans*. *Nature*, 391(6669):806–811, February 1998.
-

-
- [38] Eileen Fung, Wilson W Wong, Jason K Suen, Thomas Bulter, Sun-gu Lee, and James C Liao. A synthetic gene metabolic oscillator. *Nature*, 435:118–122, 2005.
- [39] 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.
- [40] Gillespie, Daniel T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, December 1976.
- [41] Gillespie, Daniel T. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1-3):404–425, September 1992.
- [42] Edward Giniger and Mark Ptashne. Cooperative dna binding of the yeast transcriptional activator gal4. *Proc. Natl. Acad. Sci.*, 85:382–386, 1988.
- [43] Alan L Goldstein and John H McCusker. Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast*, 15:1541–53, 1999.
- [44] Didier Gonze, Samuel Bernard, Christian Waltermann, Achim Kramer, and Hanspeter Herzel. Spontaneous synchronization of coupled circadian oscillators. *Biophysical Journal*, 89(11):120–129, 2005.
- [45] Anthony J. F. Griffiths, Susan R. Wessler, Richard C. Lewontin, and Sean B. Carroll. *Introduction to Genetic Analysis (Introduction to Genetic Analysis (Griffiths))*. W. H. Freeman, 9 edition, February 2007.
- [46] Raul Guantes and Juan F. Poyatos. Dynamical Principles of Two-Component Genetic Oscillators. *PLoS Computational Biology*, 2:0188–0197, 2006.
- [47] Peter Hanggi. Stochastic resonance in biology. How noise can enhance detection of weak signals and help improve biological information processing. *Chemphyschem*, 3:285–290, 2002.
- [48] Jeff Hasty, David McMillen, and J. J. Collins. Engineered gene circuits. *Nature*, 420(6912):224–230, November 2002.
-

-
- [49] Vassily Hatzimanikatis and Kelvin H. Lee. Dynamical analysis of gene networks requires both mRNA and protein expression information. *Metabolic Engineering*, 1(4):275–281, 1999.
- [50] Peter Hemmerich, Tanja Stoyan, Gerhard Wieland, Marianne Koch, Johannes Lechner, and Stephan Diekmann. Interaction of yeast kinetochore proteins with centromere-protein transcription factor cbf1. *Proc Natl Acad Sci U S A.*, 97(23):12583–12588, 2000.
- [51] Werner Horsthemke and Rene Lefever. *Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology*. New York: Springer-Verlag, 1984.
- [52] Steven Johnson. *Emergence*. New York: Scribner, 2001.
- [53] Ghil Jona, Mordechai Choder, and Opher Gileadi. Glucose starvation induces a drastic reduction in the rates of both transcription and degradation of mrna in yeast. *Biochimica et biophysica acta*, 1491:37–48, 2000.
- [54] Jerome Jouffroy and Thor I Fossen. A tutorial on incremental stability analysis using contraction theory. *Modeling, Identification and Control*, 31(3):93–106, 2010.
- [55] Jrme Jouffroy and Jean-Jacques E. Slotine. Methodological remarks on contraction theory. In *IEEE Int. Conf. on Decision and Control*, 2004.
- [56] Mads Kaern, William J. Blake, and James J. Collins. The engineering of gene regulatory networks. *Annu Rev Biomed Eng*, 5:179–206, 2003.
- [57] Hitto Kaufmann, Xenia Mazur, Martin Fussenegger, and James E Bailey. Influence of low temperature on productivity, proteome and protein phosphorylation of CHO cells. *Biotechnology and Bioengineering*, 63(5):573–582, June 1999. PMID: 10397813.
- [58] Yiannis Kaznessis. Models for synthetic biology. *BMC Systems Biology*, 1(1), 2007.
- [59] Branislav Kovacech, Kim Nasmyth, and Tillman Schuster. EGT2 gene transcription is induced predominantly by Swi5 in early G1. *Mol. Cell. Bio.*, 16:3264–3274, 1996.
-

-
- [60] Beat P. Kramer, Alessandro 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.
- [61] Aditya Kumar and Prodromos Daoutidis. *Control of nonlinear differential algebraic equation systems: with application to chemical processes*. CRC Press, 1999.
- [62] Laurent Kuras, Regine Barbey, , and Dominique Thomas. Assembly of a bZIP-bHLH transcription activation complex: formation of the yeast Cbf1-Met4-Met28 complex is regulated through Met28 stimulation of Cbf1 DNA binding. *The EMBO journal*, 16:2441–2451, 1995.
- [63] Yuri A. Kuznetsov. *Elements of Applied Bifurcation Theory (Applied Mathematical Sciences)*. Springer, Berlin, 3rd edition, 2004.
- [64] Tong Ihn Lee, Nicola J Rinaldi, Francois Robert, Duncan T Odom, Ziv Bar-Joseph, Georg K Gerber, Nancy M Hannett, Christopher T Harbison, Craig M Thompson, Itamar Simon, Julia Zeitlinger, Ezra G Jennings, Heather L Murray, D Benjamin Gordon, Bing Ren, John J Wyrick, Jean-Bosco Tagne, Thomas L Volkert, Ernest Fraenkel, David K Gifford, and Richard A Young. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 298(5594):799–804, Oct 2002.
- [65] Lennart Ljung. *System Identification: Theory for the User (2nd Edition)*. Prentice Hall PTR, December 1998.
- [66] 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.
- [67] Winfried Lohmiller and Jean-Jacques E Slotine. On contraction analysis for nonlinear systems. *Automatica*, 34:683–696, 1998.
- [68] Jane L. Lubischer. *The Cell Cycle, Principles of Control*. New Science Press, 2007.
-

-
- [69] Tatyana Luzyanina and Dirk Roose. Numerical stability analysis and computation of Hopf bifurcation points for delay differential equations. *J. Comput. Appl. Math.*, 72, 1996.
- [70] Shmoolik Mangan and Uri Alon. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A*, 100(21):11980–11985, 2003.
- [71] Philippe Marguet, Frederick Balagadde, Cheemeng Tan, and Lingchong You. Biology by design: reduction and synthesis of cellular components and behaviour. *Journal of the Royal Society, Interface / the Royal Society*, 4(15):607–23, 2007.
- [72] Lucia Marucci, David A. W. Barton, Irene Cantone, Maria Aurelia Ricci, Maria Pia Cosma, Stefania Santini, Diego di Bernardo, and Mario di Bernardo. How to turn a genetic circuit into a synthetic tunable oscillator, or a bistable switch. *PLoS ONE*, 4(12):e8083, 12 2009.
- [73] Lucia Marucci, Stefania Santini, Mario di Bernardo, and Diego di Bernardo. Derivation, identification and validation of a computational model of a novel synthetic regulatory network in yeast. *Journal of Mathematical Biology*, June 2010.
- [74] Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal Sheffer, and Uri Alon. Superfamilies of evolved and designed networks. *Science*, 303(5663):1538–1542, 2004.
- [75] Milo, Ron, Shen-Orr, Shai , Itzkovitz, Shalev, Kashtan, Nadav, Chklovskii, Dmitri, and Alon, Uri. Network Motifs: Simple Building Blocks of Complex Networks. *Science*, 298(5594):824–827, 2002.
- [76] Melanie Mitchell. *An Introduction to Genetic Algorithms (Complex Adaptive Systems)*. The MIT Press, February 1998.
- [77] 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.
-

-
- [78] C.G. Moles, P. Mendes, and J.R. Banga. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.*, 13:2467–2474, 2003.
- [79] T. Moll, G Tebb, U. Surana, H. Robitsch, and K. Nasmyth. The role of phosphorylation and the CDC28 protein kinase in cell cycle-regulated nuclear import of the *S. cerevisiae* transcription factor SWI5. *Cell*, 66:743–58, 1991.
- [80] Harold J. Morowitz. *The emergence of everything : how the world became complex*. Oxford University Press, 2002.
- [81] M. Muratani, C. Kung, K.M. Shokat, and W.P. Tansey. The f box protein dsgr1/mdm30 is a transcriptional coactivator that stimulates gal4 turnover and co-transcriptional mrna processing. *Cell*, 120:887–899, 2005.
- [82] James Dickson Murray. *Mathematical Biology, Vol. 1: An Introduction*. Springer-Verlag, 2008.
- [83] Oliver Nelles. *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Springer, 1 edition, December 2000.
- [84] Béla Novák and John J. Tyson. Design principles of biochemical oscillators. *Nature Reviews Molecular Cell Biology*, 9(12):981–991, 2008.
- [85] Sophocles Orfanidis. *Introduction to Signal Processing*. Prentice Hall, US ed edition, August 1995.
- [86] V Paetkau, R Edwards, and R Illner. A model for generating circadian rhythm by coupling ultradian oscillators. *Theoretical Biology and Medical Modelling*, 3:3–12, 2006.
- [87] F. Pengcheng and S. Panke. *Systems Biology and Synthetic Biology*. Wiley-Interscience, 2009.
- [88] Quang-Cuong Pham, Nicolas. Tabareau, and Jean-Jacques E Slotine. A contraction theory approach to stochastic incremental stability. *IEEE T Automat Contr*, 2009.
-

-
- [89] Christopher V. Rao, Denise M. Wolf, and Adam P. Arkin. Control, exploitation and tolerance of intracellular noise. *Nature*, 420:231–237, 2002.
- [90] Andreas Raue, Clemens Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and Jens Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.
- [91] 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, 2006.
- [92] Giovanni Russo and Mario Di Bernardo. How to synchronize biological clocks. *Journal of computational biology : a journal of computational molecular cell biology*, 16(2):379–393, 2009.
- [93] Giovanni Russo, Mario di Bernardo, and Eduardo D. Sontag. Global entrainment of transcriptional systems to periodic inputs. *PLoS Comput Biol*, 6(4):e1000739, 04 2010.
- [94] 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.
- [95] Odeniel Sertil, Rachna Kapoor, Brian D Cohen, Natalia Abramova, and Charles V Lowry. Synergistic repression of anaerobic genes by Mot3 and Rox1 in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 31:5831–5837, 2003.
- [96] Rudiger Seydel. *Practical Bifurcation and Stability Analysis: From Equilibrium to Chaos*. Springer, 1994.
-

-
- [97] Jean-Jacques E. Slotine, Wei Wang, and Khalid El Rifai. Contraction analysis of synchronization in networks of nonlinearly coupled oscillators. In *16th International Symposium on Mathematical Theory of Networks and Systems*, 2004.
- [98] 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.
- [99] David Sprinzak and Michale B. Elowitz. Reconstruction of genetic circuits. *Nature*, 438(7067):443–8, November 2005.
- [100] Craig A. Strathdee, Marilyn R. McLeod, and Jennifer R. Hall. Efficient control of tetracycline-responsive gene expression from an autoregulated bi-directional expression vector. *Gene*, 229(1-2):21 – 29, 1999.
- [101] 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.
- [102] Zoltan Szallasi, Jörg Stelling, and Vipul Periwal. *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts*. The MIT Press, 2006.
- [103] Thattai, Mukund and van Oudenaarden, Alexander. Stochastic Gene Expression in Fluctuating Environments. *Genetics*, 167(1):523–530, May 2004.
- [104] R. Thomas. *Quantum Noise*. Springer Series in Synergetics; ed. Gardiner, C. W. (Springer, Berlin), 1981.
- [105] Xiao-Jun Tian, Xiao-Peng Zhang, Feng Liu, and Wei Wang. Interlinking positive and negative feedback loops creates a tunable motif in gene regulatory networks. *Physical Review E*, 80(1), 2009.
- [106] Marcel Tigges, Nicolas Dénervaud, David Greber, Jörg Stelling, and Martin Fussenegger. A synthetic low-frequency mammalian oscillator. *Nucleic acids research*, March 2010.
-

-
- [107] Marcel Tigges, Tatiana T. Marquez-Lago, Jörg Stelling, and Martin Fussenegger. A tunable synthetic mammalian oscillator. *Nature*, 457(7227):309–312, 2009.
- [108] Ana Traven, Branka Jelacic, and M. Sopta. Yeast gal4: a transcriptional paradigm revisited. *EMBO Rep*, 7:496–499, 2006.
- [109] Tony Y. Tsai, Yoon S. Choi, Wenzhe Ma, Joseph R. Pomerening, Chao Tang, and James E. Ferrell. Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science*, 321(5885):126–129, July 2008.
- [110] John J. Tyson, Attila Csikasz-Nagy, and Bela Novak. The dynamics of cell cycle regulation. *Bioessays*, 24(12):1095–1109, December 2002.
- [111] Malkhey Verma, Paik J Bhat, and Kareenhalli V Venkatesh. Quantitative analysis of gal genetic switch of *saccharomyces cerevisiae* reveals that nucleocytoplasmic shuttling of gal80p results in a highly sensitive response to galactose. *Journal of Biological Chemistry*, 278:48764–48769, 2004.
- [112] José M G Vilar, Hao Y Kueh, Naama Barkai, and Stanislas Leibler. Mechanisms of noise-resistance in genetic oscillators. *Proc Natl Acad Sci U S A*, 99(9):5988–5992, 2002.
- [113] Achim Wach, Arndt Brachat, Christina Alberti-Segui, Corinne Rebischung, and Peter Philippsen. Heterologous HIS3 marker and GFP reporter modules for PCR-targeting in *Saccharomyces cerevisiae*. *Yeast*, 13:1065–75, 1997.
- [114] Eric Walter and Luc Pronzato. *Identification of Parametric Models from Experimental Data*. Springer, 1997.
- [115] Wilfried Weber, Jörg Stelling, Markus Rimann, Bettina Keller, Marie Daoud-El Baba, Cornelia C Weber, Dominique Aubel, and artin. Fussenegger. A synthetic time-delay circuit in mammalian cells and mice. *Proc Natl Acad Sci U S A.*, 104(8):2643–2648, 2007.
- [116] Stefanie Widder, Josef Schicho, and Peter Schuster. Dynamic patterns of gene regulation I: simple two-gene systems. *Journal of theoretical biology*, 246:395–419, 2007.
-

-
- [117] Yeager-Lotem, Esti, Sattath, Shmuel, Kashtan, Nadav, Itzkovitz, Shalev, Milo, Ron, Pinter, Ron Y., Alon, Uri, and Margalit, Hanah. Network motifs in integrated cellular networks of transcription regulation and protein-protein interaction. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16):5934–5939, April 2004.
-