Dynamics and biological information processing: from gene circuits to cellular populations

Marçal Gabaldà Sagarra

TESI DOCTORAL UPF / ANY 2016

DIRECTOR DE LA TESI
Jordi Garcia Ojalvo

Departament CEXS

The important thing is not to stop questioning. Curiosity has its own reason for existing. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvelous structure of reality. It is enough if one tries merely to comprehend a little of this mystery every day. Never lose a holy curiosity.

Albert Einstein

Als meus pares

Agraïments

Durant aquest camí he tingut la immensa sort d'estar acompanyat d'un bon grapat de bona gent. Si aquesta tesi ha estat possible és només gràcies al seu recolzament.

Jordi, moltes gràcies per TOT. Gràcies tant pel suport i la confiança, com per la dedicació i la paciència. Gràcies per tantes coses noves apreses. I moltes gràcies pel teu optimisme i bon humor pràcticament inamovibles. Amb tota franquesa, em sento afortunat d'haver pogut treballar amb tu aquests quatre anys.

Muchas gracias a Nara, por darme la mano mientras daba mis primeros pasos en el mundo de la investigación. Lo recuerdo con mucho cariño. Moltíssimes gràcies gràcies al Pau, la Belén, l'Elena, la Marta, la Letícia, l'Alessandro, el Maciek, el Bernat, la Rosa, la Lara, el Carlos, el Nico, el Josep, l'Elba i Jon, per un ambient estimulant, divertit i molt acollidor, dins i fora la ciència. Heu estat un gran recolzament i m'heu ensenyat moltes coses, i podeu estar segurs que m'enduc un tros de vosaltres allà on vagi o em quedi.

Thanks to Gurol and Jintao for a thrilling journey to the biofilm world.

Thanks also to Michael Stumpf for giving me the opportunity to visit your group. Thanks to Adam and Aaron for the discussions and insightful explanations, even at late hours. Also thanks a lot to them and to Robert,

Virginia, Ezter, Anderson, Nick, Ann, Ravneet and Delphine for a very warm environment.

Moltes gràcies al Marc, la Berta i la Laura per acollir-me, per passejar-me, i per assegurar-vos, d'una manera certament pragmàtica, que no passés gana. I en general, per fer que a Londres sortís més el sol.

Gràcies als Primos pel continu bon l'humor i per la teràpia de grup. Han estat molt importants!

Moltes gràcies als de la Crew, per ser una via de despressurització tant necessària de tant en tant. Però també per animar-me a superar els meus límits, i per vigilar, cervesa en mà, que no me'n posi de nous.

Moltíssimes gràcies Sergi per fer-me costat, per animar-me a confiar en mi mateix, que ja saps que costa. I per seguir fent-me de germà gran.

Gràcies als meus pares, per transmetre'm que no hem vingut aquí a patir, però al mateix temps que l'esforç acostuma a donar els seus fruits. I gràcies per un suport incondicional sempre i a pesar del que sigui.

Claudia, moltes gràcies pel recolzament, els ànims, i la comprensió. I per la paciència, que no ha estat poca! I moltes gràcies per compartir-ho tot.

Un cop més, moltes gràcies a tots.

Abstract

Living beings strive to survive and reproduce. To do so they need to monitor and adapt to a dynamic environment, taking advantage of the opportunities and dodging the threats that they may encounter. This thesis is devoted to analyse, using a mathematical and computational approach, a variety of mechanisms of anticipation, adaptation, and resistance to changes in the external conditions. To do so, we have studied regulatory systems across different scales. First, we analysed the genetic circuit that controls the galactose utilization by *Saccharomyces cerevisiae*, which is composed by a small set of genes, with the goal of establishing the tools to explore the epistatic interactions of mutations within this system. Second, we propose a framework to understand how cellular regulatory networks, composed by hundreds or thousands of elements, can process time-dependent information in a state-dependent manner. Finally, we discovered and characterised the mechanism of growth oscillations in biofilms of *Bacillus subtilis* cells. We found that these oscillations are caused by the metabolic co-dependence between distant groups of cells, and are crucial to ensure the resistance of the colony to external aggressions.

Resum

Els éssers vius malden per sobreviure i reproduir-se. Per aconseguir-ho necessiten monitoritzar i adaptar-se a un ambient dinàmic, aprofitant les oportunitats i esquivant les amenaces que els sobrevenen. L'objectiu d'aquesta tesis ha estat analitzar, mitjançant eines matemàtiques i computacionals, alguns dels mecanismes cel·lulars d'anticipació, adaptació i resistència a canvis en el medi. Per fer-ho hem estudiat sistemes de regulació en diferents escales. En primer lloc, hem analitzat el circuit genètic, format per uns pocs gens, que controla el consum de galactosa per part de *Saccharomyces cerevisiae*, i hem desenvolupat les eines necessàries per explorar les interaccions epistàtiques entre mutacions dins aquest sistema. En segon lloc, proposem un marc teòric per explicar com les xarxes de regulació cel·lular, que consten de centenars o milers d'elements, poden processar informació temporal codificant-la en la seva pròpia dinàmica. Finalment, hem descobert i caracteritzat el mecanisme d'oscil·lacions en el creixement de biofilms de *Bacillus subtilis*. Aquestes oscil·lacions són conseqüència de l'acoblament metabòlic entre regions distants del biofilm, i juguen un paper crucial per assegurar la resistència de la colònia a agressions externes.

Preface

Life is a dynamic process that takes place within a dynamic environment. Living beings strive to survive and reproduce in an habitat with scarce resources and fierce competition. In this context, there is an arms race to develop mechanisms to adapt to the ever fluctuating external conditions. We devote this thesis to the study of some of such mechanisms, their dynamical properties and their computational capabilities.

This thesis begins with an introduction to the systems studied and their dynamical properties. First, we discuss the different scales at which biological systems are regulated: small gene circuits, large-scale regulatory networks, and cellular populations. Second, we deal with the dynamic regimes exhibited by the systems studied in the thesis, namely bistable and oscillatory behaviours, and the complex dynamics observed in networks. Finally, we discuss the relevance of information-processing processes in biological systems, paying special attention to the use of temporally structured information to anticipate fluctuations in external conditions.

We next discuss our study of the gene regulatory circuit that controls the galactose utilisation by *Saccharomyces cerevisiae*. Under certain conditions this circuit displays memory, which can be lost due to stochastic transitions in a manner that depends on galactose concentration. Using experimental data obtained by our collaborators Aaron New and Ben Lehner at the Centre for Genomic Regulation, we characterise the system's response to external cues, and develop two models to explore how this response is affected by mutations.

Next we propose the use of a functional network framework known as *reservoir computing* to understand how cellular regulatory networks

could process temporal information. This framework has been developed in the fields of computational neuroscience and machine learning. With a purely theoretical work, but using the real regulatory networks of five distant organisms, we discuss the possibility that biological systems function as reservoir-computing instances, encoding the recent history in their dynamics.

Finally, in close collaboration with experimentalists in the group led by Gürol Süel at University of California San Diego, we study how colonies of millions of cells coordinate and balance conflicting interests of growth and protection. Specifically, we identify and characterise growth oscillations in *Bacillus subtilis* biofilms, and determine that they are caused by collective metabolic regulation. The study of these collective dynamics allows us to understand one of the mechanisms of biofilm resilience to external attacks, and suggests new strategies to control this bacterial resistance phase.

Contents

List of figures **xviii**

List of tables **xix**

1 INTRODUCTION **1**

1.1 Biology at multiple scales 1

1.1.1 Regulatory circuits 1

1.1.2 Regulatory networks 2

1.1.3 Cellular populations 3

1.2 Dynamics 7

1.2.1 Bistability 8

1.2.2 Oscillations 17

1.2.3 Network dynamics 23

1.3 Cellular information processing 24

1.3.1 Anticipating the environment 26

1.3.2 Reservoir computing 29

2 DISSECTION OF A STOCHASTIC MEMORY CIRCUIT **33**

2.1 Population dynamics 35

2.1.1 Population model 38

2.1.2 Fitting the model parameters with experimental data 39
2.2 Simplified model of the GAL network 43
2.2.1 Model description 43
2.2.2 Computing the switching rates 46
2.2.3 Interaction of genetic mutations 50
2.3 Mechanistically detailed model of the GAL network . . . 53
2.3.1 Deterministic model 53
2.3.2 Computing the switching rate with stochastic simulations 57
2.3.3 Missing information 60
2.3.4 Interaction of genetic mutations 68
2.4 Discussion 69

3 STATE-DEPENDENT COMPUTATION IN GENE REGULATORY NETWORKS 73
3.1 Reservoir computing structure in gene regulatory networks 76
3.2 Dynamical encoding of information in the biological reservoirs 81
3.2.1 Reservoir dynamics 83
3.2.2 NARMA task 85
3.2.3 Critical memory capacity 89
3.3 Encoding biologically relevant inputs 90
3.4 Evolvability of a functional readout 94
3.5 Discussion 97

4 COLLECTIVE OSCILLATIONS IN CELLULAR COMMUNITIES 101
4.1 Cooperation versus competition 101
4.1.1 Growth oscillations 102

4.1.2 Metabolic regulation of biofilm growth 104
4.2 Mathematical model of metabolic codependence 109
4.3 Model perturbations 116
4.3.1 Addition of glutamine 116
4.3.2 Addition of glutamate 118
4.3.3 Addition of ammonium 119
4.4 Model predictions 122
4.4.1 Growth in the interior after an external attack . . 122
4.4.2 GDH overexpression 124
4.4.3 Biofilm elimination: *divide et impera* 127
4.5 Oscillation onset and system size 128
4.5.1 Relative size of interior and periphery 129
4.5.2 Death in the interior 133
4.5.3 Effect of the subpopulation size on the oscillator 134
4.6 Discussion 137

5 CONCLUSIONS 141

A BOW-TIE STRUCTURE OF BIOLOGICAL NETWORKS 147

Contributions 153

Bibliography 153

List of Figures

1.1 Undomesticated *Bacillus subtilis* biofilm growing in a Petri dish 4
1.2 *Cooperative bacteria*: educational video project about biofilms (Video) 5
1.3 Localised cell death facilitates wrinkle formation in *B. subtilis* biofilms 6
1.4 Hysteresis and irreversibility in bistable signalling circuits 9
1.5 Topology and dynamics of positive and double negative feedback circuits 11
1.6 GAL regulatory network of *Saccharomyces cerevisiae* . . 14
1.7 History-dependent activation of GAL network 16
1.8 Effect of delay on a negative-feedback loop 19
1.9 Types of biochemical oscillators according to the source of delay 20
1.10 Two possible mechanisms of loss of collective oscillations at low cell density 22
1.11 Comparison of regulatory approaches to respond to an environmental perturbation 26
1.12 Architectures of neural networks 29

1.13 Recurrent neuronal networks project stimuli as a trajectory into a multidimensional space 31

2.1 The distribution of expression of Gal1p is bimodal and sensitive to previous growth conditions 36
2.2 Variability of the parameter values of the best candidate solutions tried during the fitting process 42
2.3 Bootstrap of the rates fitted with experimental data . . . 43
2.4 Creation and destruction rates and energy potential of the *GAL3* feedback loop 48
2.5 Experimental and simulated activation rates with respect to galactose concentration 51
2.6 Systematic exploration of the activation rates with respect to changes in galactose concentration and other physiological parameters 52
2.7 Bifurcation diagram of the GAL regulatory network with respect to α_{gal} 56
2.8 Stochastic simulation of the time evolution of the protein copy number of the GAL regulatory proteins 60
2.9 Dependence of the activation rate on V and α_{gal} with respect to the experimental observations 62
2.10 Analysis of the Pearson correlation coefficients of the RNA-seq gene expression data 65
2.11 Distribution of predicted protein copy number for each GAL protein using a population of reference proteins . . 67
2.12 Results of the iterative fitting of volume and galactose to α_{gal} 69
2.13 Systematic exploration of two parameter changes in the stochastic mathematical model 70

3.1 Structural and functional organization of reservoir computing 77
3.2 Degree distribution of the gene regulatory networks . . . 80
3.3 Relative size of functional groups of a reservoir for each network 82
3.4 Setup to test the memory of a network 84
3.5 Representative time series of the test phase of a 10th order NARMA prediction task 86
3.6 Predicting performance of reservoirs with topologies from the biological cores compared to control topologies 88
3.7 Critical memory capacity of different structures of the GRN 91
3.8 Topology and critical memory capacity of the readout structures of the gene regulatory networks 92
3.9 Efficiency of signalling pathways feeding an input signal to the reservoir 94
3.10 Training a readout through an evolutionary process . . . 96

4.1 Biofilms must balance nutrient access and protection . . 102
4.2 Biofilms grown in microfluidic devices show oscillations in colony expansion 104
4.3 Stable oscillations in biofilm growth (Video) 105
4.4 Onset of oscillations in biofilm growth (Video) 105
4.5 Biofilm growth depends specifically on extracellular ammonium availability 107
4.6 The model describes the dynamics of two populations in a moving frame 109
4.7 A metabolic negative feedback loop as the cause of the growth oscillations 111
4.8 Metabolic oscillations produced by the matematical model 114

4.9 Systematic sensitivity analysis of the parameters of the model 117
4.10 Modelling results for glutamine supplementation 119
4.11 Modelling results for glutamate supplementation 120
4.12 Modelling results for ammonium supplementation 121
4.13 External and internal attacks on biofilms 125
4.14 Glutamate dehydrogenase overexpression breaks the metabolic coupling 127
4.15 Oscillations start when the biofilm is large enough and become slower over time 129
4.16 Modelling the biofilm as two concentric circular regions of constant density 130
4.17 Effect of biofilm size on growth oscillations 136
4.18 Metabolic coupling ensures the viability of the whole colony 138

A.1 Scheme of the bow-tie structure classification of nodes . 148

List of Tables

3.1 General properties of the gene regulatory networks studied and their recurrent cores 79
3.2 Connected components in the gene regulatory networks . 79
3.3 Number of nodes by category in the largest connected component of each network 81
3.4 Number of nodes affected by stress type for each evidence threshold 93

4.1 Parameters of the mathematical model of the biofilm growth oscillations 115

A.1 Sizes of the different groups of nodes defined by bow-tie classification 149

1 Introduction

1.1 Biology at multiple scales

Biological research has been dominated by a reductionist approach that has produced a great amount of knowledge about individual cellular components and their functions. Yet, there is growing evidence that individual molecules can only rarely account for discrete biological functions. Instead, it is the interactions among the myriad of cellular constituents — e.g. proteins, DNA, RNA, and small molecules— and even among populations of cells that give rise to most characteristics of living systems.

1.1.1 Regulatory circuits

Proteins are the main effectors of cellular processes, from metabolism or synthesis to structural elements that shape the cell. However, many of them do not perform an actual enzymatic or structural role, but rather regulate the action of other proteins. Others are, at the same time, effectors and regulators. All these regulators implement non-trivial logical operations in the cellular response to stimulus to adapt to a changing environment. The combination of such regulators increases even further the complexity of the repertoire of responses that the cell can orchestrate.

Cellular functions, including the regulatory ones, are often performed by highly modular circuits (Hartwell et al., 1999). In general, modularity refers to a group of physically or functionally linked molecules that work together to perform a relatively distinct function (Hartwell et al., 1999; Wall et al., 2004; Alon, 2003). Some of the many examples of modularity found in biology are the invariant protein-protein and protein-RNA complexes that are at the core of numerous basic biological functions — from nucleic-acid synthesis to protein degradation (Alberts, 1998)—, or the temporally co-regulated groups of molecules that control the different stages of the cell cycle (Simon et al., 2001; McAdams and Shapiro, 2003), or that convey signals in bacterial chemotaxis or yeast pheromone response. These modules, or circuits, are the main specialised regulatory units that give rise to specific cellular functions (Alon, 2007b).

Notoriously, even relatively small circuits can give rise to a surprisingly wide variety of dynamical behaviours. For instance, a bacterial module of two proteins that regulate each other activity can display monostable, bistable, excitable or oscillatory dynamics depending on the inputs it receives (Espinar et al., 2013).

1.1.2 Regulatory networks

The analysis of small regulatory circuits and modules discussed above is motivated by the growing evidence that most cellular processes are regulated by groups of molecules within functional modules (Hartwell et al., 1999). However, thanks to the development of high-throughput data-collection techniques it is becoming increasingly clear that these modules are not isolated from each other. Instead, these circuits exhibit a high degree of connectivity among them, interact and frequently overlap (Danial et al., 2003), giving rise to various types of interaction webs, or networks.

These can include protein–protein interaction, metabolic, signalling and transcriptional networks. As an example, in the *Escherichia coli* transcriptional regulatory network most motifs overlap, generating distinct homologous clusters, in which the specific motifs are no longer clearly separable (Dobrin et al., 2004). Besides, these networks often have an inherent scale-free hierarchy, in which the achievable dynamical range is constrained by the underlying topology (Ravasz et al., 2002; Almaas et al., 2004). All in all, these circuits are not isolated from each other, and their study needs to be addressed not only with a reductionist, motif-based approach, but also from a global perspective (Bardwell et al., 2007; Al-Shyoukh et al., 2011).

1.1.3 Cellular populations

In addition, cells rarely live in isolation, but rather inside cellular communities. Within these cellular communities a network of interactions among the individual cells arises. In many situations, cells just compete for the resources available (Hibbing et al., 2010; Oliveira et al., 2014), but in many other cases cellular populations exhibit a remarkable degree of cooperative behaviour. This is evident in the case of multicellular organisms, from nematodes to humans, but it also appears to apply widely among single-celled organisms such as bacteria, fungi, and amoeba. In many cases, the label 'single-celled' applies to only part of the life cycle of these organisms. This is the case of unicellular organisms such as the soil bacterium *Bacillus subtilis*, which can assemble into multicellular structures even starting from a single cell (Branda et al., 2001; Webb et al., 2003).

When initially deposited on a surface, either organic or not, *B. subtilis* cells adhere to it and start growing and dividing to form a colony.

They subsequently secrete a scaffolding matrix, mainly composed of proteins and polysaccharides, known as extracellular polymeric substance or EPS. The EPS embeds the cells, providing a substrate for their growth and maintaining them together in a macroscopic aggregate (Webb et al., 2003) visible to the naked eye (Figure 1.1). This structure is a *biofilm*.

Figure 1.1: Undomesticated *Bacillus subtilis* biofilm growing in a Petri dish. Image courtesy of Munehiro Asally.

Biofilms are sessile resistance phases of bacteria that are found in surfaces from the ocean floor to human teeth, and are present almost ubiquitously across the prokaryotic phylogenetic tree, dating at least 3.2 billions of years (Rasmussen, 2000; Westall et al., 2001). The biofilm structure not only serves as mechanical support for cell growth, but protects the colony against a wide range of environmental challenges, such as UV exposure (Espeland and Wetzel, 2001), metal toxicity (Teitzel and Parsek, 2003), acid exposure (McNeill and Hamilton, 2003), dehydration and salinity (Le Magrex-Debar et al., 2000), phagocytosis (Leid et al., 2002) and several antibiotics and antimicrobial agents (Stewart and Costerton,

2001; Gilbert et al., 2002; Mah and O'Toole, 2001). As an example, *Pseudomonas aeruginosa* cells growing inside a biofilm are resistant to antibiotic concentrations 100-1000 times larger than planktonic cells (Stewart, 2002; Høiby et al., 2010). This resilience makes them hard to eradicate, causing problems ranging from persistent infections in wounds and medical implants (Lynch and Robertson, 2008) to high maintenance costs in industrial installations or increased fuel consumption in ships (Mattila-Sandholm and Wirtanen, 1992; Flemming, 2002; Chambers et al., 2006). An educational explanation of the biofilm formation process and the impact of this bacterial growth form can be found in Figure 1.2.

Figure 1.2: *Cooperative bacteria*: educational video project about biofilms (Video). This video was a project for the *Visual Science* course, which addresses the use of images and video in science communication. The course was offered as part of the PRBB Intervals programme of interdisciplinary education.
https://vimeo.com/96884263

In any case, the biofilm organisation poses a conflict between the needs of individual and the colony fitness. This conflict is specially relevant since, unlike in a multicellular organism, cells within the biofilm have full replicative potential, and in the wild biofilms are often multispecies consortia (Watnick and Kolter, 2000). As a matter of fact, although biofilms increase the fitness of most of their inhabitants, this may come at a price of a loss of fitness by part of the colony.

A paradigmatic example of loss of fitness that produces a common good is given by the formation process of the intricately wrinkled structures that the external scaffold adopts (Figure 1.1). These structures are thought to improve the transport of soluble nutrients and waste (Wilking

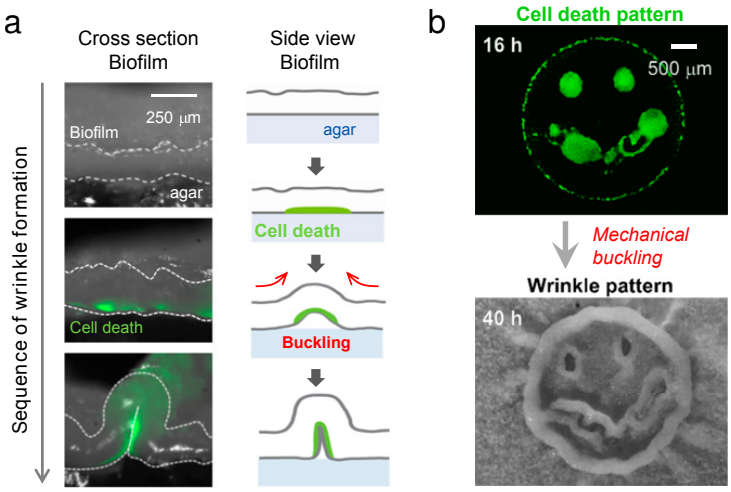

Figure 1.3: Localised cell death facilitates wrinkle formation in *B. subtilis* biofilms. a, Cross-sectional images and schematic of the wrinkle formation process. Cell death is marked in green. b, Artificial smiley face in cell death pattern (top) produced by locally increasing the cell density, and the wrinkle pattern generated (bottom). From Asally et al. (2012).

et al., 2013), and prevent the community from permeation by fluids (Epstein et al., 2011). However, as Asally et al. (2012) show, these wrinkles are generated as a consequence of localised cell death patterns. These regions where cell death occurs detach from the surface, facilitating the folding of the biofilm scaffold (Figure 1.3).

We devoted Chapter 4 to study the mechanisms that allow biofilms formed by millions of cells to balance the fitness of individual cells with colony resilience. Specifically, we characterised the metabolic co-dependence that occurs between the sheltered interior cells and the fast-growing peripheral cells. We discovered that this co-dependence gives rise to metabolic oscillations and causes periodic halting of the

biofilm growth, preventing starvation of the interior cells and protecting the colony from external aggressions.

1.2 Dynamics

Dynamical behaviour is ubiquitous in gene regulatory processes, and is governed by the architecture of the underlying genetic networks. Feedback-loop structures, for instance, can strongly influence the dynamics of a regulatory circuit, expanding the set of possible biological properties, including robustness to uncertainty (Astrom and Murray, 2012), and can produce single-cell phenotypic heterogeneity in a uniform environment (Mitrophanov and Groisman, 2008). The dynamic consequences of each type of feedback loop, namely activatory (positive) and inhibitory (negative), have been described, and include signal amplification and bistability for positive-feedback loops (Isaacs et al., 2003; Brandman and Meyer, 2008; Novák and Tyson, 2008), and noise control, improvement of response time, and oscillations for negative-feedback loops (Rosenfeld et al., 2002; Becskei and Serrano, 2000; Novák and Tyson, 2008).

On the other hand, in large-scale regulatory networks the number of interlocked feedbacks and recurrences might be so large that it becomes difficult, if not impossible, to distinguish the individual contributions of each one of them to the global dynamics. Indeed, in large and interconnected networks new dynamical properties emerge (Bhalla and Iyengar, 1999).

In the following sections we will describe two of the most salient dynamics that can arise from positive and negative-feedback loops. Then, we will review some examples of complex dynamics observed in bio-

logical networks and the potential of networks as information-processing systems.

1.2.1 Bistability

While most of the biochemical reactions involved in cell signalling and regulation are reversible (conformational changes, phosphorylation, or even synthesis and degradation), many biological transitions are essentially irreversible. For example, under most circumstances cells remain differentiated even after the stimulus that triggered their differentiation ceases (Ferrell and Machleder, 1998; Xiong and Ferrell, 2003). The self-sustained nature of the changes induced by the reversible activation of signalling pathways is a consequence of the bistable dynamics of the underlying regulatory system.

A bistable dynamical system is characterised by having two coexisting stable steady states. In other words, the system will be in either one of two discrete states, and intermediate states are possible only transiently. In that case, which stable state the system adopts will depend on the initial concentrations of the reacting species. Consequently, if the state of the system is affected by an external signal in a way that it undergoes bifurcations entering and exiting the bistable regime depending on the strenght of the signal, different profiles of activity —i.e. stimulus/response curves— will be obtained when the stimulus is increased and decreased (Figure 1.4a). The limit point of signal strength at which the system will transition from one stable branch to the other will differ from the signal strength at which the system will return to the original branch. For intermediate intensities of the signal the state of the system will depend on its history, as it will just remain in the branch it has been

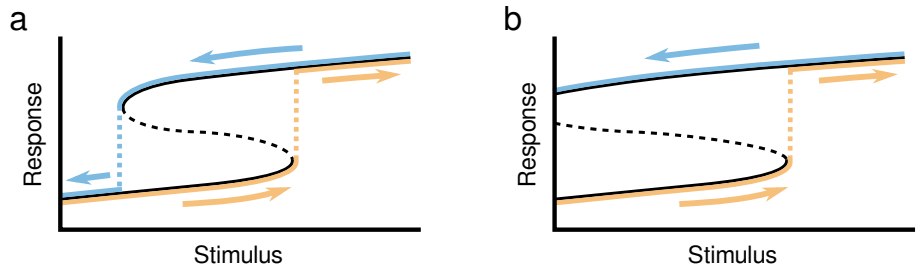

Figure 1.4: Hysteresis and irreversibility in bistable signalling circuits. Bifurcation diagram of two arbitrary bistable systems, showing the value of system variable at the steady state (response) as a function of the value of an input (stimulus). Solid (dashed) black lines are stable (unstable) branches, and orange (blue) lines are the stimulus/response curves obtained when increasing (decreasing) the stimulus strength. a, Hysteresis. The limit point where the system jumps from the lower to the upper branch for increasing signal strengths is different than where it jumps from the upper to the lower branch for decreasing signals. Between the two limit points the system will be either on the upper or the lower branch depending on its history. b, Irreversibility. For certain parameter values, and in particular if the feedback in a bistable circuit is strong enough, the system may become irreversible. In this case, the saddle-node bifurcation where the system would reach the left limit point and jump from the upper to the lower branch lays on the negative domain of signal strength and, thus, it is not accessible. Once the system has jumped to the upper stable branch it will remain there even if the stimulus is removed completely.

since the last transition. This time-based dependence on present and past inputs is called *hysteresis*, which is a form of cellular memory.

On the other hand, under some conditions the transition from one of the stable branch to the other might not be physically possible. In such cases, the system can transition from one branch of steady states to the other but cannot undergo the reverse transition when the signal strength varies back and forth across its domain. Then, the transition is said to be irreversible (Ferrell and Machleder, 1998). One of the transitions can be-

come inaccessible in many ways (Guidi and Goldbeter, 1997), for example one of the two limit points can move towards an inaccessible domain (e.g. negative values) of signal strength (Figure 1.4b).

Such irreversible switches presumably play major roles in developmental processes characterised by a point-of-no-return. The maturation process of frog oocytes of the genus *Xenopus* is a well-known example (Ferrell and Machleder, 1998; Xiong and Ferrell, 2003). When these oocytes are exposed to progesterone, the intracellular activity of the mitogen-activated protein kinase (MAPK) shoots up, and does not cease even after the progesterone signal is gone, triggering the maturation of the cell. Similarly, irreversible bistable dynamics have been proposed to describe the regulation of apoptosis (Chang et al., 2002; Legewie et al., 2006), and for the activation of the lytic cycle of the λ phage of *Escherichia coli* (Tian and Burrage, 2004; Oppenheim et al., 2005). On the other hand, nice examples of two-way switches displaying hysteresis include the regulation of the *lac* operon in *E. coli* (Laurent and Kellershohn, 1999; Santillán et al., 2007), the activation of M-phase-promoting factor (MPF) in frog egg extracts (Novák and Tyson, 1993), the conversion of normal prion protein to its pathogenic form (Kellershohn and Laurent, 2001), and the signalling pathway mediated by Cdc42 that controls polarization of budding yeast (Brandman et al., 2005). Furthermore, artificial genetic networks with bistable dynamics have been designed and built both in prokaryotic (Collins et al., 2000) and eukaryotic (Becskei et al., 2001) cells.

The key feature of a system required for bistability is some sort of feedback loop with a net positive effect. Essentially, this can be a positive feedback loop (figure 1.5a) where a regulator directly or indirectly promotes its own activity, or a double negative feedback loop (figure 1.5b)

where the regulator inhibits one of its inhibitors (Ferrell, 2002). These feedback interactions can involve any kind of regulation such as synthesis, degradation, translocation, or any post-transcriptional modification. Besides the regulatory feedback, at least one of the regulatory interactions is required to be highly non-linear so that it displays ultrasensitivity (Ferrell and Xiong, 2001). Finally, the strength of the different regulatory steps that form the feedback loop must be properly balanced for the system to be bistable (Ferrell, 2002).

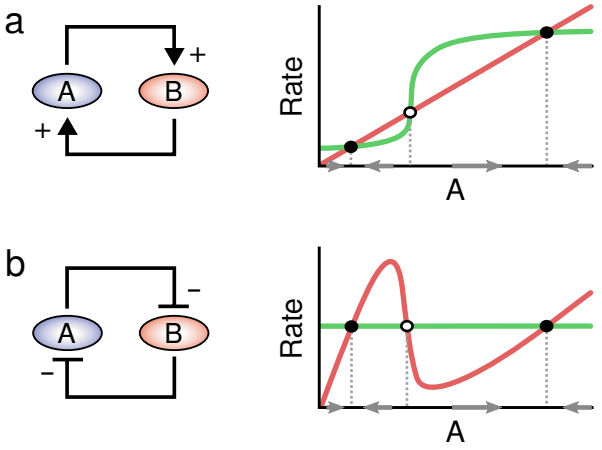

Figure 1.5: Topology and dynamics of positive and double negative feedback circuits. Interaction diagrams of simple instances of bistable circuits, and their respective synthesis or activation (green) and degradation or inhibition (red) rates of the element A as a function of its concentration. In the rate plots, full (empty) dots indicate stable (unstable) steady states, and the arrows on the horizontal axis indicate the direction of change of protein concentration. a, Positive feedback loop: A activates B and B activate A. As a result there could be a stable steady state with both variables lowt and another with both variables high. b, Double-negative feedback loop: A inhibits or represses B and B inhibits or represses A. As a results there could be a stable steady state with A high and B lowt and another with A lowt and B high. Both circuits could exhibit hysteresis or irreversible dynamics. Circuit cartoons adapted from Ferrell (2002).

On top of that, biochemical processes such as bursting activity of the transcriptional machinery, or uneven partitioning of the pool of molecules during cellular division, introduce stochasticity in the dynamics of regulatory systems. In bistable systems, one consequence of this stochasticity is that noise is expected to occasionally induce transitions between the two stable states, effectively corrupting the memory encoded in the state of the system. For sufficiently weak noise, the frequency of such stochastic transitions depends exponentially on the stability of the fixed point the system is in, measured as the potential energy barrier between the stable steady state and the unstable steady state (Hänggi et al., 1990). This potential energy barrier, in turn, is determined of the biochemical properties of the system. In a population of cells with such bistable circuits, these stochastic transitions will originate a bimodal distribution of activity of the bistable system: a fraction of cells will be in one metastable state and the rest in the other one. This is the case of the GAL regulatory network of *Saccharomyces cerevisiae*, a well-characterised regulatory circuit with bistable dynamics. We now review in detail the structure of this regulatory network. Furthermore, in Chapter 2 we characterise the dependence of the rate of stochastic transitions on the galactose concentration, to develop the necessary tools to investigate how the interaction of genetic variants affect a dynamic phenotype.

GAL regulatory network

The GAL system of *Saccharomyces cerevisiae* consists of a set genes (*GAL1*, *GAL2*, *GAL7* and *GAL10*) that encode the enzymes required to import and metabolise galactose into a metabolically useful form, glucose-6-phosphate, and the regulatory proteins (*GAL1*, *GAL3*, *GAL4* and *GAL80*) that control their expression (Figure 1.6). When yeast cells

grow in absence of galactose most of these genes are transcriptionally inert. In such conditions, Gal4p is constitutively synthesised, dimerises, and cooperatively binds to various Upstream Activating Sequences (UAS) of the rest of genes of the GAL system (Giniger et al., 1985). However, its role as transcription activator is repressed by Gal80p, which can translocate to the nucleus and block the activation domain of Gal4p (Platt and Reece, 1998). If galactose becomes available Gal3p binds to it (acting as a ligand sensor), increasing its affinity for Gal80p and blocking the activity of the latter in a galactose- and ATP-dependent manner (Bhat and Hopper, 1992; Zenke et al., 1996; Platt and Reece, 1998; Egriboz et al., 2011). As a consequence Gal4p is released and can recruit the RNA polymerase, increasing up to 1000-fold the transcription rate of the rest of the GAL genes (Lohr et al., 1995).

One of the enzymes of the GAL network upregulated by Gal4p is Gal1p, a paralog of Gal3p that has a dual role: on the one hand it is the first enzyme of the metabolic pathway that transforms galactose to glucose-6-phosphate, and on the other it plays a regulatory role similar to Gal3p. Gal3p binds to Gal80p with higher affinity, which is crucial during the initial stages of the activation of the GAL network. But Gal1p is much more abundant, which becomes relevant once the GAL network is fully active (Zacharioudakis et al., 2007; Abramczyk et al., 2012). Besides, another upregulated GAL gene is *GAL2*, which encodes for a membrane-bound permease transporter with specificity for galactose that significantly increases the rate of galactose uptake from the extracellular environment (Bhat and Hopper, 1992).

Gal1p, Gal2p, Gal3p, and Gal80p directly or indirectly modulate the activity of Gal4p, which in turn transcriptionally regulates them. This defines a series of feedback loops. Specifically, Gal2p and Gal3p participate

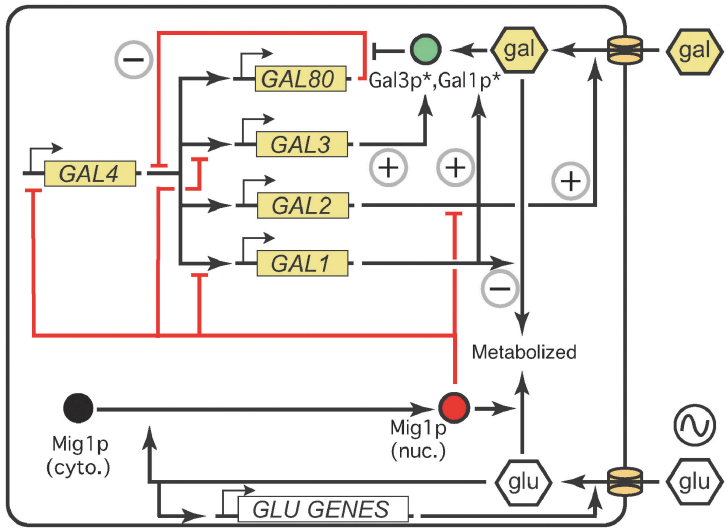

Figure 1.6: GAL regulatory network of *Saccharomyces cerevisiae*. The GAL network is controlled by interlocking positive (+) and negative (-) feedback loops. Red arrows indicate interactions that decrease the level of expression of the GAL network. The regulatory interactions of the network are discussed in detail in the text. From Stockwell et al. (2015)

in two positive feedback loops, as their upregulation leads to an increase of their own expression, and Gal80p creates a negative feedback loop, as its activity represses its own expression. Finally, Gal1p defines both a positive feedback loop with its regulatory role, by blocking Gal80p and increasing transcription, and a negative-feedback loop with its enzyme role, by removing galactose from the system.

The *S. cerevisiae* GAL network is also affected by the glucose concentration in the medium. The energy return of metabolising glucose is higher than of galactose, and thus, cells prioritise its use. When glucose is present, Mig1 transcriptionally represses *GAL1*, *GAL3* and *GAL4* (Bryant et al., 2008), triggers the degradation of Gal2p (Horak and Wolf, 1997),

and accelerates the decay of *GAL1* and *GAL3* transcripts (Bennett et al., 2008). Additionally, when the GAL network is repressed and Gal2p is not present, galactose and glucose bind competitively to hexose transporters, blocking the transport of galactose (and the activation of the GAL network) depending on its ratio with respect to glucose (Escalante-chong et al., 2014).

Among all the regulatory loops, the positive ones dominate the system, producing bistable dynamics in the regulation of the GAL network (Venturelli et al., 2012). Consequently, the network exhibits an all-or-none response, where depending on the current and previous concentrations of galactose and glucose available the system can be either fully repressed (OFF) or completely activated (ON). Under certain conditions, this bistable dynamics, together with the stochastic transitions, generates a bimodal distribution of levels of *GAL1* expression in the population. Specifically, for intermediate galactose concentrations a bimodal distribution of *GAL1* expression is obtained from cells that were initially OFF (Figure 1.7, blue). On the other hand, when the initial population is composed of ON cells, no bimodal distribution is produced: either the galactose concentration is low enough to trigger the repression of the GAL networks or all cells remain active (Figure 1.7, red). This is a consequence of the fact that the rate of stochastic transitions from the ON to the OFF state is rather insignificant in comparison to the rate of OFF to ON transitions (Venturelli et al., 2015). It has been suggested that the amount of Gal80p present also determines the stability of both metastable states, so that when Gal80p is more abundant stochastic transitions are less frequent (Acar et al., 2005), which increases the memory of the system.

Not surprisingly, the behaviour of the GAL network differs from strain to strain. Peng et al. (2015) investigated the effect on the galactose-

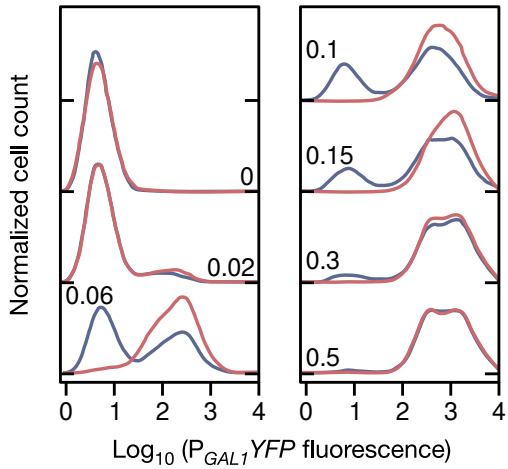

Figure 1.7: History-dependent activation of GAL network. The activity of the GAL promoters is analysed monitoring, with flow cytometry, the expression of a yellow fluorescent protein (YFP) driven by the *GAL1* promoter. Expression distributions are plotted as a function of log(YFP fluorescence). Blue and red distributions denote cells that were initially grown for 12 h without galactose (but with 2% raffinose) and with 2% galactose (and 2% raffinose), respectively. Thus, these populations were initially repressing (blue) or expressing (red) the GAL genes. After this initial incubation cells were grown for a further 27h in various concentrations of galactose as specified in the numbers next to each plot. Adapted from Acar et al. (2005)

dependent activation dynamics of combinations of variants of the promoters of the network regulator genes. Specifically, they tested the effect of each one of the promoters of *GAL2*, *GAL3*, *GAL4* and *GAL80* from *Saccharomyces paradoxus* in a *S. cerevisiae* strain. However, they did not focus on the bistable regime of the system, but on the general profile of GAL network activation as a function of galactose available. In Chapter 2 we develop the tools needed to investigate how mutations affecting the dynamics of the GAL regulatory network combine, what is the resulting phenotype in terms of the robustness of the system in the bistable domain,

and how this phenotype affects the population dynamics that originate the bimodality.

1.2.2 Oscillations

Oscillatory dynamics are found in many contexts in biological systems, and play significant regulatory roles. Some of the best known examples include the cyclin protein circuits that control the cell cycle (Richard, 2003; Chen et al., 2007), the circadian rhythms that adjust the physiological state of cells and organisms to the day-night cycle (Golden et al., 1997; Tyson et al., 1999), and the somitogenesis clock driving the development of vertebrate embryos (Lewis, 2003; Monk, 2003).

Given the ubiquity of regulatory circuits with oscillatory dynamics, it is not surprising that we ask ourselves what the necessary characteristics of a system are for it to display an oscillatory behaviour. As it turns out, there are four essential requirements of a biochemical oscillator: a negative feedback, significant nonlinearities, enough time delay and a proper balancing of the time scales of the dynamics of the elements involved (Novák and Tyson, 2008). First, the dynamics of the system must be dominated by a negative feedback, so that the oscillator can revert its state back to the oscillation 'starting point'. Then, it is necessary that nonlinearities in the kinetic rate laws of the reactions destabilise the system dynamics, and that the effect of negative feedback suffers a delay, so that the system cannot reach a steady state.

To illustrate the relevance of this time delay, let us consider a protein that represses its own expression and that is affected by enzymatic

degradation. The dynamics of this system are described by

$$\frac{dY(t)}{dt} = k_1 S \frac{K_d^p}{K_d^p + Y(t-\tau)} - k_2 E_T \frac{Y(t)}{K_m + Y(t)} \tag{1.1}$$

where its degradation (second term) is enzymatic and follows a Michaelis-Menten dynamics, and its synthesis (first term) is proportionally increased by an activator S and inhibited by itself in a cooperative manner. This inhibition is modelled using the Hill factor $K_d^p / \left(K_d^p + Y(t-\tau)\right)$, where p expresses the degree of cooperativity and K_d the dissociation constant of Y to the upstream regulatory region of its own promoter. Importantly, τ defines a delay in the self-repression of Y, so that when $\tau > 0$ the rate if protein expression depends on a past concentration of Y. As shown in Figure 1.8, this simple system can reach a steady state or produce oscillations depending on the value of the delay τ.

This time delay can have multiple natures. For instance, delays may be caused by physical constrains, such as the time required for transcription and translation, or for translocation of species between cellular compartments. This is specially relevant in eukaryotic cells, where many maturation and translocation steps might be involved from the moment a gene of a transcription factor is transcribed until the actual protein exerts its regulatory function. Similarly, there might be a long chain of intermediates participating in the loop, which would also introduce a delay in the feedback loop (Figure 1.9a). These systems are referred to as *delayed negative-feedback loops*. Delayed negative-feedback mechanisms have been suggested to explain the oscillatory dynamics of the *PER-TIM* circadian clock of fruit flies (Tyson et al., 1999), the p53 response to ionizing radiation (Lev Bar-Or et al., 2000; Ma et al., 2005; Monk, 2003), and the

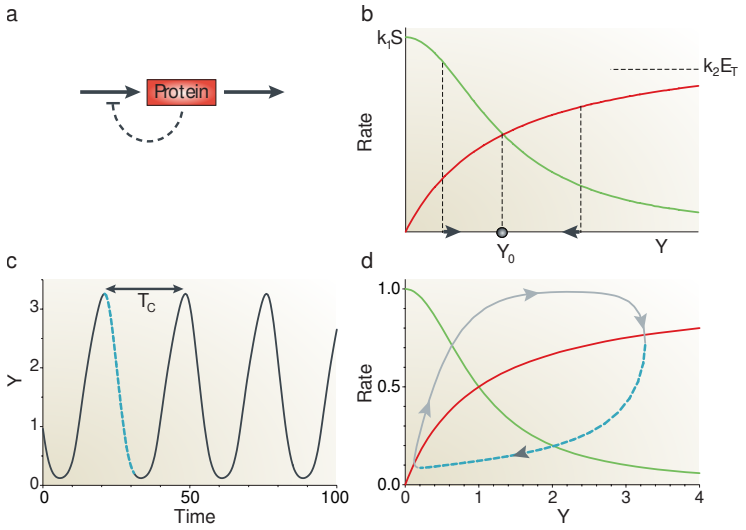

Figure 1.8: Effect of delay on a negative-feedback loop. a, Regulation scheme of the protein dynamics described by Eq. (1.1). The protein inhibits its own synthesis and it is degraded enzymatically. b, Protein synthesis (green) and degradation (red) rates as a function of its concentration. For $\tau = 0$, the system always tends towards the stable state Y_0 (the arrows indicate the direction of change of protein concentration). c, For $\tau = 10\,\text{min}$, the system is in an oscillatory regime. The period T_C of the oscillation is 27.2 min. d, The delayed rate of protein synthesis as a function of current protein concentration is shown in grey. The blue portion of the curve corresponds to the blue portion of the time trace in c, and has corresponds to a duration of τ time units. The value of the rest of parameters is given by $p = 2$, $K_m/K_d = 1$, $S/K_d = 1$, $k_1 = k_2E_T/K_d = 1\,\text{min}^{-1}$. Adapted from Novák and Tyson (2008).

NF-κB response to tumour necrosis factor (Monk, 2003; Nelson et al., 2004; Cheong et al., 2008), to name a few.

On the other hand, hysteresis (caused by a positive feedback in the reaction mechanism, see Section 1.2.1) can also introduce the delay needed for the system to oscillate. In this case, the positive-feedback loop forces the system to overshoot and undershoot the steady state continuously

(Figure 1.9b). This type of mechanism, known as *amplified negative feedback*, has been used to describe the oscillations controlling the timing of cellular divisions in frog eggs (Pomerening et al., 2005). Furthermore, the synthetic oscillator developed by Stricker et al. (2008) also falls into this category.

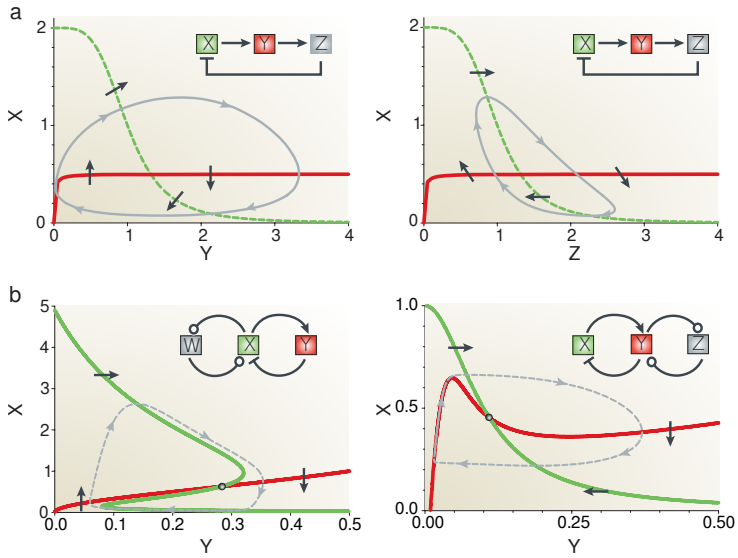

Figure 1.9: Types of biochemical oscillators according to the source of delay. Interaction diagrams of instances of oscillators of different types, and state-space representations of their dynamics. a, Two different projections (XY) and (XZ) of a three-element delayed negative feedback loop. b, Two different flavours of amplified negative-feedback loops, where the activator (left) or the inhibitor (right) is amplified, as it forms part of an additional positive-feedback loop. In the diagrams pointy arrowheads mean activation, flat arrowheads mean inhibition, and circular arrowheads mean either activation or inhibition. In any case, multiple circular arrowheads in the same diagram must be of the same type. The state-space diagrams display the nullclines of X (green) and Y (red) and the projection of the limit cycle onto the XY plane. Adapted from Novák and Tyson (2008).

When multiple oscillators interact, they tend to adapt their periods in a process known as *entrainment* (Roenneberg et al., 2005). The period of entrained oscillators often match in a 1:1 ratio, which is the case we will consider here, but different ratios are also possible (Grasman, 1984; Jensen and Krishna, 2012). Entrainment and synchronisation of a population of cellular oscillators produce emergent dynamics at the population level. In other words, when most oscillators synchronise, the oscillations become observable at the population scale (Winfree, 2002). Furthermore, if the interaction between oscillators is mediated by diffusive molecules through the extracellular medium, the cellular density will strongly influence the strength of the coupling. Together, this would generate oscillations at the population level that are dependent on cellular density (Figure 1.10a).

Another possible cause for emergent population-level oscillations that depend on cellular density is that intracellular dynamics depends itself on population density, so that the oscillations are lost within each cell if the population becomes too sparse (Figure 1.10a). An example of such collective oscillatory mechanisms is the collective glycolytic oscillations in yeast cells. These oscillations are found in the time dynamics of NAD(P)H in starved, anaerobic yeast cell suspensions (Duysens and Amesz, 1957; Betz and Chance, 1965; Richard, 1996). De Monte et al. (2007) analysed the decay dynamics of the oscillation with respect to cellular densities, determining that the combined release of the coupling molecule (acetaldehyde) by a large enough number of cells was essential not only for synchronization, but for the oscillations themselves. Indeed, the coupling exerted on the oscillations is strong enough so that whenever the oscillations are possible, they are synchronised. Interestingly, in this case, the dynamical state of the cell reflects the population density.

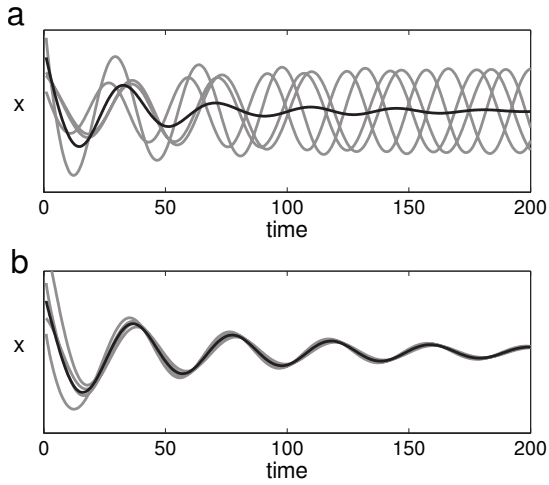

Figure 1.10: Two possible mechanisms of loss of collective oscillations at low cell density. The grey lines show the dynamics of four arbitrary oscillators within a population 25 times larger, and the black line represents the average over the population, which is the macroscopic observable. a, Incoherence: cells progressively loose their mutual entrainment due to phase drift, as a consequence of noise or frequency differences, producing an algebraic (i.e. non-exponential) decay of the oscillation amplitude. b, Dynamic quorum sensing: cells have a coherent motion and stop oscillating in synchrony. In this case, the amplitude of the oscillations decays exponentially. From De Monte et al. (2007).

In Chapter 4 we study global oscillations observed in the growth dynamics of *Bacillus subtilis* biofilms. In that case, the population level oscillations are not caused by synchronization of individual oscillations: the metabolic coupling between different regions of the biofilm introduces the negative-feedback loop responsible for the oscillations, and the finite diffusion of metabolites between those regions sets the time delay. Thus, it is rather that the oscillatory dynamics itself emerges when the biofilm becomes large enough.

1.2.3 Network dynamics

In the previous sections we have seen how the interaction between small groups of elements could give rise to dynamics richer than what those elements could offer by themselves. Indeed, we cannot stress enough the fact that it is the interactions between the elements what causes oscillations, bistability, as well as other dynamics, such as ultra-sensitivity and perfect adaptation often found in biological systems. When considering these circuits in isolation, we are able to understand and pinpoint the mechanisms that shape their dynamics. Yet, regulatory circuits are often highly interconnected within the cellular machinery, forming large networks (Bhalla and Iyengar, 1999; Danial et al., 2003; Helikar et al., 2008). Although our knowledge of the dynamic determinants of larger and more interacting systems is not as complete, we expect this increase in the size and connectivity to originate new and more complex dynamics. Similarly, it is well known that large neuronal networks display emergent dynamical properties that cannot appear in individual cells or small circuits (Chialvo, 2010).

An insightful example of specific dynamical properties of large biological networks is given by Rué et al. (2010). They analyse the dynamics of the signal transduction network of human fibroblasts using a Boolean model developed by Helikar et al. (2008). Boolean network models simplify the dynamics of the elements of the network trying to isolate the essential effect of the biochemical interactions of the system, and have been successfully applied to describe gene regulatory networks (Mendoza et al., 1998; Covert et al., 2004; Fauré et al., 2006; Davidich and Bornholdt, 2008), cellular differentiation (Klemm and Bornholdt, 2005), evolution (Bornholdt and Sneppen, 2000), and signal transduction in cells

(Saez-Rodriguez et al., 2007; Helikar et al., 2008; Samaga et al., 2009; Bauer et al., 2010).

Rué et al. (2010) show that the network displays an enormous richness of attractor states, both fixed and periodic, which can be related to the capacity to store information in the network. Furthermore, the network has a characteristic relaxation time scale, which is not altered by the presence of noise in the input signals nor can be attributed to the typical shortest path length. Interestingly, this property determines how the signal transduction network responds to time varying input signals. Finally, they show that the network displays a variety of nontrivial frequency responses (high-pass, band-pass and low-pass) for time-varying input signals.

All in all, their results illustrate that the dynamical properties of networks of interacting elements is much more complex than what we can expect from the small-circuit descriptions. Inspired by this possibility, in Chapter 3 we propose that cellular regulatory networks can integrate signals and encode memory in a decentralised manner, not relying on specialised regulatory circuits but on the complexity of the dynamics of the whole network.

1.3 Cellular information processing

All living beings, from the simplest unicellular organism to humans, survive by constantly processing the information that they survey from their environment. The ability to sense the presence of sustenance, reproductive opportunities, and imminent danger is, as such, the primary physiological requirement across all domains and stages of life. Consequently, a requirement of life is the complex biological machinery that has evolved to process such information, translating relevant signals from the environ-

ment into appropriate responses and behaviours (Lenski and Travisano, 1994). Furthermore, while multicellular organisms can have cells specialised in sensory and information-processing functions, microbes need to resort to their signal-transduction, regulatory and metabolic networks in order to perceive and process their environment and articulate the corresponding action, with a classical example being a cell computing an external ligand concentration by time-averaging (Berg and Purcell, 1977).

Until recently, homoeostasis has been the *de facto* framework for understanding cellular behaviour —although it was originally proposed in the context of human physiological adaptation (Cannon, 1932). In its most essential form, it implies that cells maintain their internal state in some acceptable range by directly responding to perturbations resulting from environmental fluctuations (e.g., expression of heat shock proteins in response to high-temperature stress). This homoeostasis-based framework has been extremely useful to understand a wide range of cellular behaviour. Yet, over the past years novel findings have pointed towards more complex cellular behaviours than those predicted by the reactive homoeostatic framework. For instance, by focusing on the functional role of noise in biological systems (Eldar and Elowitz, 2010; Raj and van Oudenaarden, 2008; Swain et al., 2002), and switching the attention from population to single-cell dynamics (Kalisky and Quake, 2011; Lidstrom and Konopka, 2010; Locke and Elowitz, 2009; Raser, 2005; Taniguchi et al., 2010), it has been shown (Eldar and Elowitz, 2010) that microbes can improve their resilience to environmental fluctuations by exploiting the heterogeneity of the population (Figure 1.11).

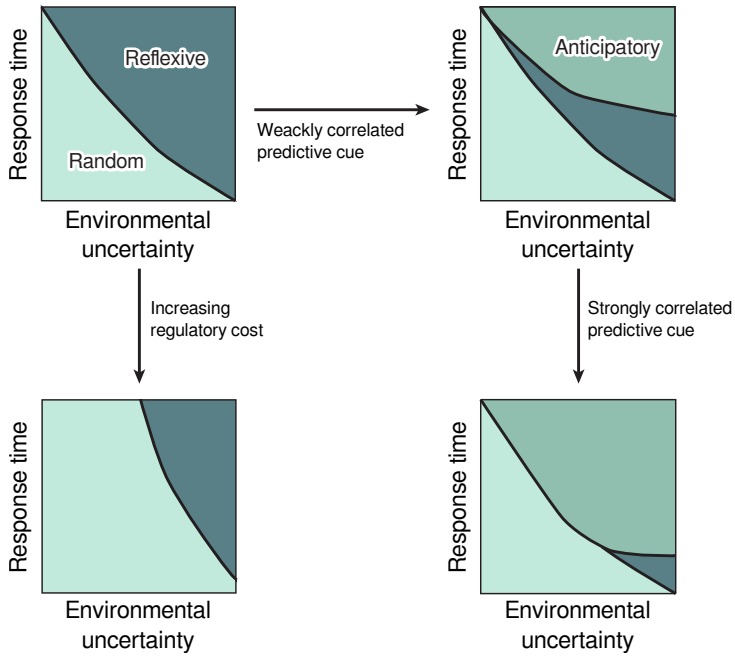

Figure 1.11: Comparison of regulatory approaches to respond to an environmental perturbation. Qualitative scheme showing the most efficient regulatory approach as a function of the time needed to orchestrate the response and of the predictability of the environment fluctuations. The possible regulatory approaches considered are reflexively responding to the environment perturbations (reflexive), randomly activating the response mechanism to ensure that part of the population is prepared for the perturbation when it arrives (random), and predicting the perturbation from additional cues in the environment (anticipatory). From Freddolino and Tavazoie (2012).

1.3.1 Anticipating the environment

Cellular responses rely on biochemical processes that are not instantaneous, such as signal transduction, gene expression, transport, and protein degradation. Hence, cells need a certain amount of time to build a response to a given external change (Alon, 2007a). Consequently, when

such fluctuations are essentially random —and thus unpredictable—, it may become difficult for the cell to activate the appropriate response program. Nonetheless, the native habitats of cells are highly structured, so that there are strong correlations in environmental perturbations that make them a bit more predictable. As a consequence, there is a potential benefit for cells in being capable to process certain temporal information and inferring future external determinants (Figure 1.11). In other words, the ability to efficiently anticipate changes in the environment represents a significant improvement in the capacity of adaptation of the cell.

If the environment is perfectly predictable, such as day-night cycles, cells can orchestrate a pre-emptive response. An example of such predictive behaviour is given by Tagkopoulos et al. (2008), who show that undomesticated strains of *E. coli* sense the temperature rise as they are ingested by a mammal, in order to anticipate the depletion of oxygen and start adapting their metabolism to the upcoming anaerobic environment that cells will find in the host gut. Similar cross-regulations of unrelated stressors in yeast have been evolved in a laboratory setting, where the presence of one of the stressors carries information about the likelihood of the appearance of the other (Dhar et al., 2013). These associations between an stimulus and an apparently unrelated response are learnt at evolutionary scales, and are hardcoded into the regulatory networks of cells. On the other hand, Sorek et al. (2013) propose a regulatory circuit that would be able to capture, store and use the association, or lack thereof, between a biologically relevant signal and multiple additional stimuli. Although the proposed associative learning circuit could infer associations at a much faster time scale, it also requires a circuit specialised at storing the presence or absence of correlation for each of the stimuli.

Besides relying on complex temporal structures, anticipation requires memory: it is necessary to store the relevant period of the cell history in order to extract the necessary information from it. In that direction, Wolf et al. (2008) showed that organisms as simple as soil bacteria possess a short-term and a long-term memory that influence cellular decisions. They documented how clonal cultures of *Bacillus subtilis* grown in different conditions, when subjected to a common stress, develop different responses depending on the pre-stress conditions.

Even though some of the environment-anticipation mechanisms mentioned above may involve only a handful of cellular elements, the large complexity of interactions among diverse types of molecules such as DNA, RNA, proteins and metabolites is what makes the cell able to accurately adapt to the environment conditions (Mattick, 2001; Stelling et al., 2002). Just as neuronal networks have enormous computational capabilities in comparison with a single neuron (Chialvo, 2010), larger regulatory networks have much larger information-processing potential than isolated circuits of a few elements (Bhalla and Iyengar, 1999).

In that sense, in Chapter 3, we show that the network of interactions that regulate the cell can encode not only the direct response mechanisms of the cell, but also the whole paradigm that the cell uses to integrate and process environmental signals (Barabási and Oltvai, 2004). Specifically, we are interested in the capability of cellular regulatory networks to integrate complex inputs, and specially in their ability to process complex temporal information. In that context, we propose that cellular regulatory networks —and gene regulatory networks in particular— can encode temporal information in their transient dynamics, following the *reservoir computing* paradigm.

1.3.2 Reservoir computing

Reservoir computing is a framework for computation that was described independently in the fields of neuroscience and machine learning under the names of *liquid state machine* (Maass et al., 2002) and *echo state networks* (Jaeger, 2001), respectively. This framework tries combine recurrent neural networks (RNN) and feed-forward networks (FFN) to reap the information-processing benefits of RNN, while using feed-forward networks to overcome their limitations.

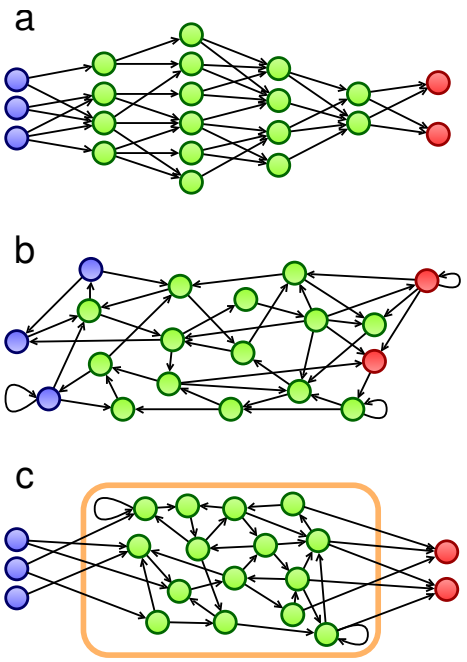

Figure 1.12: Architectures of neural networks. a, Feed-forward network. b, Recurrent neural network. c, Reservoir computing. The different types of nodes are input nodes (blue) that are the entry point of the external signals into the network, the hidden nodes (green) that are internal to the network, and the readout nodes (red) that must give the appropriate response of the system.

Feed-Forward neural networks (FFN), also known as *multi-layer perceptrons*, were one of the first neural network architectures with machine learning applications (Figure 1.12a). The processing elements, called neurons, are organised in hierarchical layers, letting the information travel unidirectionally from one layer to the next one. This structure is neatly modular, allowing the use of different types of neurons and simplifying the process of training the weights of the connections for a given task. On the other hand, it is equivalent to a mathematical function and it cannot represent temporal structures: the output of the system computed from a given input does not depend on previous nor following inputs. In summary, these networks can be efficiently trained to process spatially complex inputs but cannot process temporal information (Buonomano and Maass, 2009; Lukoševičius and Jaeger, 2009).

Recurrent neural networks (RNN), in comparison with FFN, are much more robust, can process temporal information, and are able to model highly nonlinear systems (Verstraeten et al., 2007). In these networks there are no hierarchies in the way neurons are organised, with no topological distinction between input, output or intermediate nodes (Figure 1.12b). As a consequence, there are recurrences in the information flow inside the network and, thus, virtually any node can affect any other. Indeed, these networks behave like a high-dimensional dynamical system: an input signal will produce a complex dynamical perturbation of the current state of the system that will only gradually disappear, and the dynamics of the network will depend both on the external inputs and the internal state of the network. Consequently, the recent history of the networks is projected in the multidimensional space of their dynamics. On the other hand, the main drawback of RNNs is also a consequence of the recurrence of its connections: the potential

influence of any connection on the global dynamics of the network makes the training of the network a very inefficient process (Buonomano and Maass, 2009).

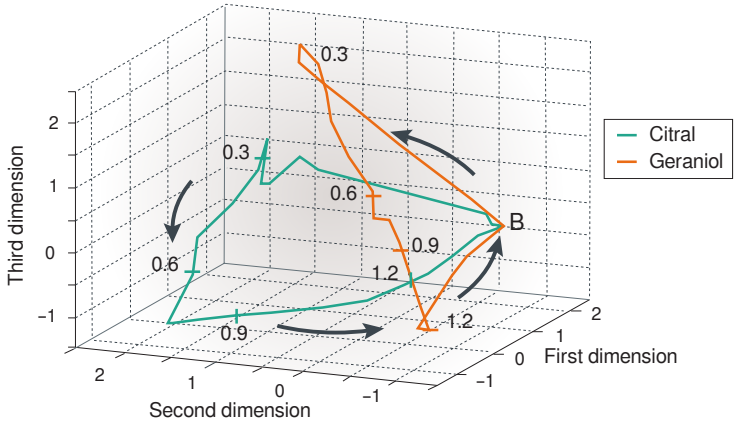

Figure 1.13: Recurrent neuronal networks project stimuli as a trajectory into a multidimensional space. The two trajectories represent the time activity of a population of neurons from the locust antennal lobe during presentations of two odours (citral and geraniol). Each odour produces a different trajectory, and thus different spatio-temporal patterns of activity. The numbers along the trajectory indicate time (seconds) since the stimulus presentation, and the point marked B indicates the resting state of the neuronal population. From Buonomano and Maass (2009), original results from Broome et al. (2006).

The reservoir computing framework uses a RNN as a reservoir that is not trained, but is read out by a simple classification feed-forward layer (Figure 1.12c). The only requirement for the reservoir is that the RNN has to fulfil the so called fading memory or echo-state property, which implies that the perturbations in the system eventually fade out (Verstraeten et al., 2007). In other words, the current state of the system depends on inputs from a finite time window in the past. Mathematically, this is ensured if the spectral radius (the largest eigenvalue) of the adjacency matrix of

the network is smaller than 1. As shown by Boedecker et al. (2012), the memory of the system is maximised when the spectral radius is just slightly smaller than one and the reservoir is close to criticality. Finally, a purely feed-forward structure is found downstream of the RNN and is trained to extract the relevant information from the transient multidimensional dynamics of the reservoir. The advantage of the reservoir computing paradigm is that it describes the possibility of using a RNN without the need of adapting the weights of the internal connections of the network. Instead, only the links towards and between the output layers need to be trained to learn a new task (Buonomano and Maass, 2009). Thus, the approach based on reservoir computing simplifies notably the training of the RNN and makes it more meaningful as a biological model.

2 *Dissection of a stochastic memory circuit*

Cells need to continuously monitor the environment for changes and coordinate the appropriate responses when necessary. Nutrient availability is a particularly relevant aspect of the environment that challenges the decision-making mechanisms of cells. Indeed, the fitness of microbes is largely determined by their ability to integrate information about the availability of multiple nutrients and to articulate an adequate and coordinated response (Cai and Tu, 2012; Broach, 2012; Chubukov et al., 2014). A paramount example is the catabolite repression mechanisms that allow cells to prioritize carbon sources depending on their energy content, the metabolic cost of building the machinery required to process them and their relative abundances (Chubukov et al., 2014).

A well-studied yet complex case of catabolism regulation is the GAL network of *Saccharomyces cerevisiae*. This network consists of a small set of proteins that import and metabolise galactose. Additionally, a subset of the GAL network regulates very efficiently the expression of most of its elements in response to galactose concentration in an all-or-none manner, varying up to 1000-fold the expression level of some of its genes

(Lohr et al., 1995). See page Figure 1.6 and 12 in Section 1.2.1 for a detailed description of the GAL network.

The GAL regulatory network contains five interlocked regulatory feedback loops, among which the *GAL1* and *GAL3* positive feedbacks strongly influence its dynamics, leading to bistability (Venturelli et al., 2012). As a consequence of bistability the system has hysteresis, which is a form of memory: within the bistability region the state of the system depends on its history (Stockwell et al., 2015). Indeed, for a range of galactose concentrations where the system is bistable the level of expression of the GAL genes does not depend only on the current inputs but also on the initial conditions (Venturelli et al., 2012).

Within the bistable regime, although the cell is committed to one of the two states (namely, high or low expression) depending on its history, transitions can still occur, and their likelihood be modulated depending on the external cues (Acar et al., 2005; Ramsey et al., 2006). Stochastic fluctuations in the level of expression of the GAL genes can force the system to transition from one metastable state to the other. The likelihood of these memory-loss events depends on the metastable state the system is at and its robustness to molecular noise, which in turn depends on the availability of galactose (Acar et al., 2005). Thus, even under hysteresis, the cues from the environment and the interior of the cell can be used to adapt the cellular state by modulating the frequency of stochastic transitions.

Here we aim to develop a method to characterize the noise-induced decision-making process underlying galactose utilization, with the goal of predicting the effect of genetic variability on this phenotype. We intend to eventually study how genetic variants affect the mechanisms involved, and predict the outcome of combinations of mutations. This chapter details a first approach of a project aimed at understanding epistasis —

i.e. non-linear interaction of combination of mutations in the phenotype (Phillips, 2008)—, from a dynamical and mechanistic point of view. The work presented in this chapter has been done in collaboration with Aaron New and Ben Lehner, at the Centre for Genomic Regulation. Unless otherwise specified, all unpublished experiments mentioned in this chapter have been carried out by Aaron New.

2.1 Population dynamics

We used a modified YPS128 strain with a yellow fluorescent protein (YFP) reporter under the control of the *GAL1* promoter to monitor the activation of the GAL network in individual cells. In all cases cultures were maintained at low cellular density to ensure that the concentration of available nutrients did not change significantly during the course of the experiment. Flow cytometry experiments confirmed that for intermediate concentrations of galactose and glucose the levels of expression of the *GAL1* promoter produced a bimodal distribution (Figure 2.1, green line). This means that cells are either expressing the *GAL1* gene at its maximum rate (ON) or repressing it (OFF), but that there is not a continuum of expression levels.

Bimodality might be caused by bistable, excitable or even oscillatory dynamics. Nevertheless, it is well known that the GAL network displays a bistable behaviour (Acar et al., 2005). The difference is that in this case the transitions between the two states observed in the population are driven by noise. If bimodality were caused by excitable or oscillatory dynamics, either one or the two transitions between states, respectively, would be deterministic.

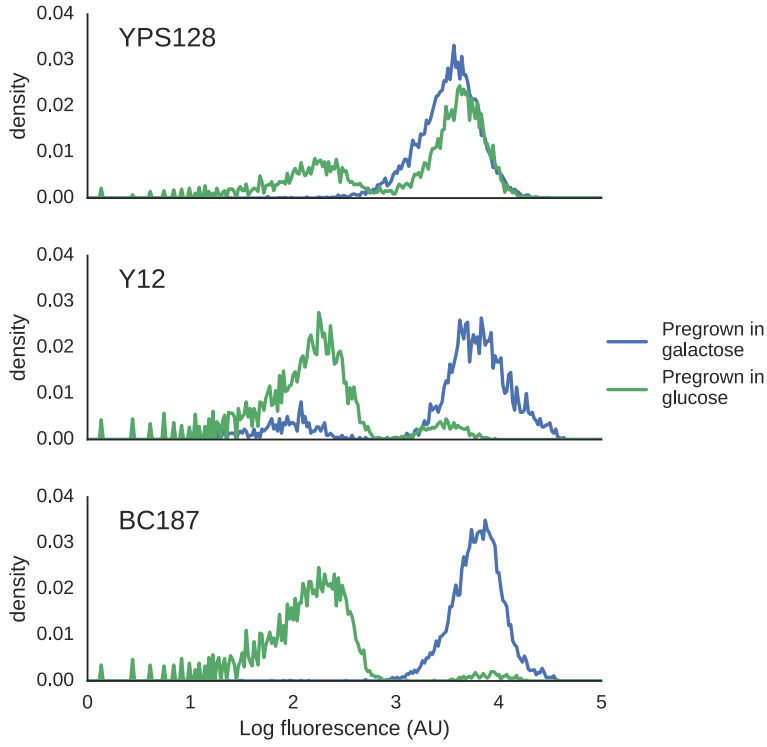

Figure 2.1: The distribution of expression of Gal1p is bimodal and sensitive to previous growth conditions. Densities of fluorescence intensity in cultures of *S. cerevisiae* after growing for 24 hours in a mixture of 2.5% galactose and 0.5% glucose are shown. The blue (green) line denotes a population that was precultured in glucose (galactose) only. The GAL genes were activated (repressed) at the beginning of the experiment, and produced an unimodal distribution of fluorescence centred at low (high) values. Fluorescence is produced by a copy of the *yfp* gene under the control of the *GAL1* promoter, and measured by flow cytometry. Different strains where tested, namely YPS128, Y12 and BC187.

Interestingly, the distribution of cells that activate the GAL network in the same intermediate medium varies among strains (Figure 2.1). The reason is that the genetic differences between those strains affect how robust to noise the two stable states of the GAL network are, and how this robustness changes depending on the galactose concentration. For example, strain BC187 shows practically no transitions between states, as the cells that start in the ON state remain there, and only a very small fraction of the population that started in the OFF state activate the GAL genes. On the other hand, in the case of Y12 transitions in both directions are slightly more common. Finally, in the case of YPS128 transitions from the OFF to the ON state seem to be much more common, as a population starting from the OFF state mostly transitioned to the ON state, while no net transitions to the OFF state was observed in the population starting from the ON state.

The large number of interlocked regulatory interactions that form the GAL regulatory network makes it difficult to deduce what are the mechanistic differences between each strain. Likewise, it is not trivial to anticipate how changes in each one of the interlocked regulatory interactions will affect the sensitivity to noise, let alone to predict how multiple mutations or variants will combine.

To study how mutations affect the probability of noise-induced transitions, first we need to be able to measure the frequency of those transitions. Using flow cytometry we can monitor how the distribution of *GAL1* expression in a population of cells changes over time. The advantage of this population data is that it makes it easier to capture rare events. On the other hand, however, it is not straightforward to measure the frequency of activation or deactivation of the GAL network from these observations. To tackle this limitation, we developed a simple model that

accounts for the population dynamics. Then, we used experimental data to fit the parameters of the model, and to obtain from there an estimation of the frequency of the GAL-state switching events.

2.1.1 Population model

To characterize the dynamics of the two subpopulations we defined a minimal model where cells can divide, die or switch their state. Let us define n_T as the total number of cells and n_H and n_L as the number of cells with high and low levels of Gal1p, respectively, so that $n_T = n_H + n_L$. The dynamics of n_H and n_L are governed by

$$\frac{dn_H}{dt} = \mu_H n_H + k_{L\to H} n_L - k_{H\to L} n_H - \gamma_H n_H \tag{2.1}$$

$$\frac{dn_L}{dt} = \mu_L n_L + k_{H\to L} n_H - k_{L\to H} n_L - \gamma_L n_L \tag{2.2}$$

where $\mu_{H,L}$ are the respective growth rates, $\gamma_{H,L}$ are the respective death rates, and $k_{L\to H}$ and $k_{H\to L}$ are switching rates from low to high levels (activation) of expression of Gal1p and vice versa (deactivation), respectively.

In our experiments the cell density is maintained at low levels via consecutive dilutions. This ensures that the availability of nutrients does not change significantly during the course of the experiment, and that the cells stay at exponential growth phase. The same conditions are considered in the model, which implies that neither the growth rate nor the death rate change over time. Since it is very hard to distinguish these two parameters using population data, we consider a net growth rate α defined as $\alpha_{H,L} = \mu_{H,L} - \gamma_{H,L}$.

Although equations Eqs. (2.1) and (2.2) describe total number of cells, we are interested in the relative size of both populations. This will allow us to further simplify the system to work with a single equation. Let us define the proportion that each subpopulation represents over the total number of cells as $f_H = n_H/n_T$ and $f_L = n_L/n_T$, so that $f_H + f_L = 1$. We can then describe the variation of f_H as

$$\begin{aligned}\frac{df_H}{dt} &= \frac{1}{n_T}\frac{dn_H}{dt} - \frac{f_H}{n_T}\frac{dn_T}{dt} \\ &= (\alpha_H - k_{H\to L})f_H + k_{L\to H}f_L - \alpha_H f_H^2 - \alpha_L f_H f_L\end{aligned} \tag{2.3}$$

and similarly for f_L. However, since $f_H + f_L = 1$ the two fractions are redundant and the dynamics of the system can be described with respect to only one of them:

$$\frac{df_H}{dt} = k_{L\to H} + (\alpha_H - k_{H\to L} - k_{L\to H} - \alpha_L)f_H + (\alpha_L - \alpha_H)f_H^2 \tag{2.4}$$

2.1.2 Fitting the model parameters with experimental data

To obtain an estimation of the switching rates, we fitted the parameters of Eq. (2.4) to our experimental data. To do so, the fraction of cells with high expression of GAL proteins was monitored at different time-points during 32 hours in cultures growing in a range of galactose concentrations. This dynamic data was used to fit the population model parameters using a basin-hopping optimization algorithm (Wales and Doye, 1997; Rossi and Ferrando, 2009; Wales, 2010; Olson et al., 2012; Rondina and Da Silva, 2013). Basin-hopping is an optimization algorithm that combines the Metropolis algorithm approach to global search with gradient-based

local search, significantly improving its efficiency for continuous objective functions.

As Figure 2.1 shows, for strain YPS128, there is no net transition of cells to the repressed (OFF) state from a population starting at the active (ON) state. At the same time, most of the population starting at the OFF state becomes active during the same period of time. This indicates that, for this strain, transitions from the ON to OFF state are significantly improbable in comparison with the opposite transitions. Thus, we will assume that the transition rate from the ON to the OFF state $k_{H\to L}$ is negligible.

Besides, expressing the GAL proteins imposes a significant metabolic burden that slows down the cellular growth. Indeed, when the system is active, the GAL proteins account for up to 4% of the protein expression of the cell (Lohr et al., 1995). Not surprisingly, we observed that the growth rate of cells expressing the GAL proteins was on the order of 5% slower than the one of cells repressing it. As the expression of GAL proteins increases with the concentration of galactose in the medium, so does the growth burden.

We used this knowledge to reduce the degrees of freedom when fitting the model parameters with experimental data. By expecting a relationship between the growth rates and the concentration of galactose, we can fit the few parameters that define that relationship, instead of two distinct growth rates for each galactose concentration. Specifically, we assume that the growth rate α_L of cells repressing the GAL proteins is not affected by the galactose concentration. On the other hand, we consider that cells expressing the GAL proteins have a metabolic burden ψ_H that reduces

their growth rate α_H with increasing galactose concentrations.

$$\alpha_L([gal]) = \alpha^* \tag{2.5}$$

$$\alpha_H([gal]) = \alpha^*(1 - \psi_H([gal])) \tag{2.6}$$

where α^* is the growth rate when the GAL system is repressed and ψ_H is the reduction of growth rate due to the metabolic cost of expressing the GAL proteins. The dependence of ψ_H on the concentration of galactose is considered to follow a Michaelis-Menten equation with a basal term:

$$\psi_H([gal]) = \psi_{basal} + \frac{\psi_{max}[gal]}{K_{sat} + [gal]} \tag{2.7}$$

The values of ψ_{basal}, ψ_{max} and K_{sat} will be fitted with experimental data. Thus, the assumption made by using Eq. (2.7) does not impose strong constraints, as it can take the form of a linear or saturating dependence of ψ_H on galactose concentration, or even a constant value, depending on the three free parameters.

Fitting results

To assess the precision of the parameter values obtained during the fitting we analysed the best candidate solutions obtained by the basin-hopping algorithm. Specifically, after 5000 iterations of the basin-hopping algorithm, we selected all those parameter sets that adjusted the experimental data with an error within 5% of the lowest error. As Figure 2.2 shows, there is wide range of possible values for the growth parameters that can fit the experimental data with a low error. Consequently, it is not possible to obtain a good estimate of the real value of those parameters with high certainty using the data available.

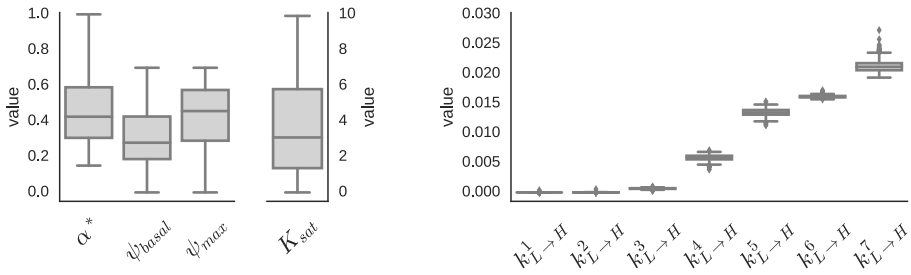

Figure 2.2: Variability of the parameter values of the best candidate solutions tried during the fitting process. The parameters that affect the growth rate (left) are much less sensitive than the switching rates (right). The seven $k_{L\to H}$ parameters correspond to the activation rates under each galactose concentration tested, namely 0%, 0.08%, 0.16%, 0.31%, 0.63%, 1.25% and 2.5%. The parameter values of the best candidate solutions tried by the basin-hopping algorithm are summarised here. The best candidate solutions are defined as those that give a fitting error within 5% of the best candidate solution. Three scaled vertical axis are used to ensure that all distributions are visible.

In any case, we are interested specifically in the switching rates and, conveniently, the fitting error is much more sensitive to the value of those parameters (see right panel of Figure 2.2). This makes it possible to obtain a better estimate of the real switching rates. Finally, to determine the uncertainty of the fitted values introduced by biological and technical errors, we repeated the fitting bootstrapping the experimental measures. Figure 2.3 shows the standard deviation of the $k_{L\to H}$ activation rates fitted from 1000 resampled datasets.

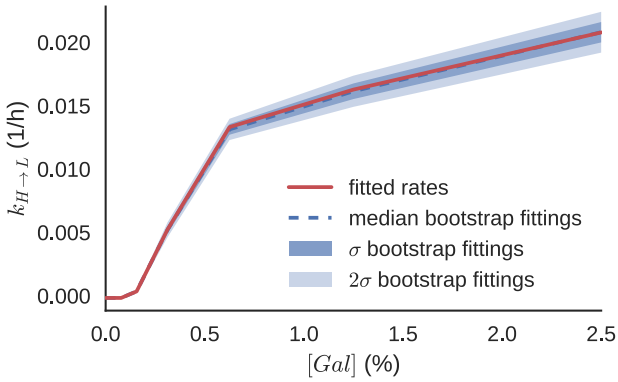

Figure 2.3: Bootstraping the original data and producing new fits suggests low sample bias. The red line marks the activation rates $k_{L\to H}$ of the best parameter set found using the whole dataset. The blue line marks the median of the activation rates found bootstrapping the original data; the two blue shadows denote one and two standard deviations away from the median.

2.2 Simplified model of the GAL network

2.2.1 Model description

Our first tentative approach to describe the molecular interactions of the GAL regulatory network was to use the mathematical model proposed by Acar et al. (2005). The model assumes that Gal4p is saturating the Upstream Activation Sequences (UAS) that control the expression of the GAL genes (see the end of Section 1.2.1 and Figure 1.6). Furthermore, it also assumes that Gal80p and Gal4p only bind when Gal4p is attached to the UAS, but not when it is in its free form. Consequently, the dynamics of Gal4p can be omitted, as the amount of Gal80p blocking the UAS is what determines the expression of the GAL genes.

Additionally, all those processes with a time-scale faster than transcription are considered at quasi-equilibrium. That is, as their dynamics are much faster than transcription, we assume that they only deviate from equilibrium by an infinitesimal amount. Consequently, we can treat them *as if they were* at equilibrium. Specifically, this simplification affects the following reactions:

$$[gal][Gal3p] \overset{\alpha}{\rightleftharpoons} [gal \cdot Gal3p] \tag{2.8}$$

$$[gal \cdot Gal3p][Gal80p^C] \overset{\beta}{\rightleftharpoons} [gal \cdot Gal3p \cdot Gal80p^C] \tag{2.9}$$

$$[Gal80p^C] \overset{\gamma}{\rightleftharpoons} [Gal80p^N] \tag{2.10}$$

$$[UAS_{free}][Gal80p^N] \overset{\delta}{\rightleftharpoons} [UAS \cdot Gal80p^N] \tag{2.11}$$

where $[Gal80p^C]$ and $[Gal80p^N]$ are the concentrations of Gal80p in the cytosol and nucleus, respectively, $[UAS_{free}]$ is the concentration of Upstream Activation Sequences not blocked by Gal80p, and $x_1 \cdot x_2$ refers to a molecular complex formed by the reversible binding of x_1 and x_2.

The quasi-equilibrium assumption allows us to use the following equilibrium constants

$$\alpha = \frac{[gal][Gal3p]}{[gal \cdot Gal3p]} \tag{2.12}$$

$$\beta = \frac{[gal \cdot Gal3p][Gal80p^C]}{[gal \cdot Gal3p \cdot Gal80p^C]} \tag{2.13}$$

$$\gamma = \frac{[Gal80p^C]}{[Gal80p^N]} \tag{2.14}$$

$$\delta = \frac{[UAS_{free}][Gal80p^N]}{[UAS \cdot Gal80p^N]} \tag{2.15}$$

and the following mass-balance equations

$$\begin{aligned}[Gal80p]_{Total} =& [Gal80p^C] + [Gal80p^N] + [UAS \cdot Gal80p^N] \\ &+ [gal \cdot Gal3p \cdot Gal80p^C]\end{aligned} \tag{2.16}$$

$$[Gal3p]_{Total} = [Gal3p] + [gal \cdot Gal3p] + [gal \cdot Gal3p \cdot Gal80p^C] \tag{2.17}$$

$$[UAS]_{Total} = [UAS_{free}] + [UAS \cdot Gal80p^N] \tag{2.18}$$

Using both the equilibrium constant and the mass-balance equations above it is possible to calculate $[UAS_{free}]$ for a given total concentration of Gal80p, Gal3p, galactose and UAS with the following equation:

$$[UAS_{free}] = \frac{[UAS]_{Total}}{1 + \tilde{x}/\delta\gamma} \tag{2.19}$$

where $\tilde{x}$ is the solution of the following equation

$$Ax + \frac{Bx}{C + x} + \frac{Dx}{E + x} - F = 0 \tag{2.20}$$

where,

$$x \equiv [Gal80p^C] \tag{2.21}$$

$$A \equiv \frac{\gamma + 1}{\gamma} \tag{2.22}$$

$$B \equiv [Gal3p]_{Total} \tag{2.23}$$

$$C \equiv \beta \left(1 + \frac{\alpha}{[gal]}\right) \tag{2.24}$$

$$D \equiv [UAS]_{Total} \tag{2.25}$$

$$E \equiv \delta\gamma \tag{2.26}$$

$$F \equiv \frac{[Gal80p]_{Total}}{F_0} \tag{2.27}$$

Acar et al. (2005) use the model above to describe the behaviour of a genetically modified strain where *GAL80* is expressed constitutively. In other words, the synthesis of Gal80p is not regulated by Gal4p and Gal80p itself. This modification effectively interrupts the regulatory negative feedback loop mediated by *GAL80*. This modified strain motivates them to consider the total concentration of Gal80p ($[Gal80p]_{Total}$) as constant. This additional simplification makes it possible to describe the behaviour of the GAL network modelling solely the dynamics of Gal3p:

$$\frac{d[Gal3p]_{Total}}{dt} = k[UAS_{free}] - \Gamma[Gal3p]_{Total} \tag{2.28}$$

The *S. cerevisiae* strain that we study, on the other hand, has the *GAL80* negative feedback loop intact. Nevertheless, in the current section we will adopt this same simplification. We argue that the regulatory positive feedback loops, and specially the one mediated by *GAL3*, are the ones that define the bistable behaviour of the GAL system. Furthermore, reducing the system to a single variable model greatly reduces the computational cost to calculate the switching rates. All in all, this mathematical model reduces the complexity of the regulatory network significantly, at risk of losing some relevant elements, but trying to capture the essential ones to explain its bistable nature in a tractable manner (indeed, as any mathematical or conceptual model ultimately tries to do).

2.2.2 *Computing the switching rates*

The activation and deactivation events of the GAL network are, in fact, escape processes from metastable states. Thus, we can use reaction-rate

theory to compute the respective rates. Specifically, as we are modelling a single-dimensional metastable system, we can use Kramers theory to compute the escape rate (Hänggi et al., 1990).

First of all, let $x(t) \equiv [Gal3p]_{Total}(t)$ be the reaction coordinate that describes the dynamics of the escape process. From Eq. (2.28), we can define the one-dimensional potential landscape as

$$U = -\int_0^{[Gal3p]_{Total}} \left(k[UAS_{free}] - \Gamma[Gal3p]_{Total}\right) d[Gal3p]_{Total} \quad (2.29)$$

The two minima x_L and x_H of the potential $U(x)$ correspond to the two metastable states for *low* and *high* expression of the GAL genes, respectively, and the local maximum X_b separating these states corresponds to the boundary or transition state (Figure 2.4).

The basic escape model introduced by Kramers (1940) consists of a classical particle of mass M moving in the one-dimensional double-well potential $U(x)$. According to the model, all the remaining degrees of freedom of the real system not captured by $x(t)$ constitute a heat bath at temperature T. The effects of the heat bath on the reacting particle are described by a fluctuating force $\xi(t)$ and by a linear damping force $-M\gamma\dot{x}$, where γ is a constant damping rate. Considering these forces, the motion of the particle is then described with the Langevin equation

$$M\ddot{x} = -\frac{dU(x)}{dx} - \gamma M\dot{x} + \xi(t) \quad (2.30)$$

where $\dot{x}$ and $\ddot{x}$ are the first and second derivatives of x with respect to time t, and $\xi(t)$ is a white Gaussian noise with zero mean and denotes a force

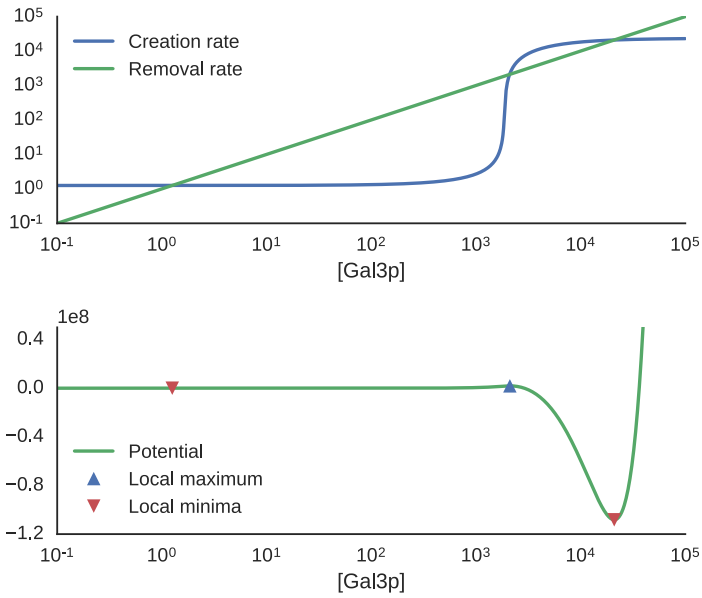

Figure 2.4: Creation and destruction rates and energy potential of the *GAL3* feedback loop. Top, creation (blue) and destruction (green) rates of Gal3p proteins as a function of the concentration of Gal3p. Bottom, energy potential of the GAL regulatory network as a function of the concentration of Gal3p. Local minima (red triangles) correspond to the two metastable states of the system: low and high expression of Gal1p. The local maximum (blue triangle) corresponds to the transition state between the two metastable states, defining the energy barrier that separates them.

that obeys the fluctuation-dissipation theorem (Kubo, 1966),

$$\langle \xi(t) \rangle = 0 \tag{2.31}$$

$$\langle \xi(t)\xi(s) \rangle = 2M\gamma k_B \delta(t-s) \tag{2.32}$$

where k_B is the Boltzman constant. In other words, the escape process is considered to be determined by Brownian motion dynamics driven by thermal forces, which are connected with the friction coefficient γ.

We can then define the angular frequencies ω_L, ω_H for the two metastable states and ω_b for the unstable fixed point that separates them as

$$\omega_L^2 \equiv M^{-1}U''(X_L) \tag{2.33}$$

$$\omega_H^2 \equiv M^{-1}U''(X_H) \tag{2.34}$$

$$\omega_b^2 \equiv M^{-1}|U''(X_b)| \tag{2.35}$$

Additionally, if we can consider that the Brownian particle is in a regime with strong friction γ:

$$M\ddot{x} = 0; \quad \dot{x} = -\frac{U'(x)}{\gamma M} + \frac{\xi(t)}{\gamma M} \tag{2.36}$$

whose deterministic part we can compare with Eq. (2.28), taking into account Eq. (2.29). This analysis makes it possible to compute the rate at which the system (in our case, the GAL regulatory network) crosses the transition state x_b from one basin of attraction towards the other. Or equivalently, we can compute at what rate the GAL regulatory network escapes from one metastable state towards the other. It can be calculated using the following expressions:

$$k_{L\to H} = \frac{\omega_L\omega_b}{2\pi}e^{-\beta E_{Lb}} \tag{2.37}$$

$$k_{H\to L} = \frac{\omega_H\omega_b}{2\pi}e^{-\beta E_{Hb}} \tag{2.38}$$

where $E_{Yb} = U(x_b) - U(x_Y)$ for $Y \in \{H, L\}$, $\beta = (k_B T)^{-1}$, k_B is the Boltzman constant and T is the temperature parameter, which effectively determines the intensity of noise. A detailed discussion of how Eqs. (2.37) and (2.38) are obtained can be found at Hänggi et al. (1990).

2.2.3 *Interaction of genetic mutations*

Since the experiments that motivated the parameter set used by Acar et al. (2005) were done with a strain different to ours, it was expected that we had to adjust some of the parameters to our case when using our experimental data. Additionally, the level of noise β of the Kramers rate equation that we introduced also needed to be adjusted. For that purpose we used a local minimisation gradient-descend algorithm. Trying to introduce as few changes as possible, we fitted the noise level β together with the equilibrium constant α of the Gal3p and galactose binding (Eq. (2.12)), as it was enough to provide a reasonable fitting of the experimental activation rates. For the rest of parameters we used the values provided by Acar et al. (2005). The fitted values of both parameters are $\beta = 1.42 \cdot 10^{-6} J^{-1}$ and $\alpha = 4.67 \cdot 10^{-1}\%$. Figure 2.5 shows the accurate matching between experimental and simulated activation rates obtained.

With a model that quantitatively reproduces the dependence of memory stability as a response to external galactose concentrations, we are able to explore how genetic alterations affect this response. Specifically, it is straightforward to explore how the dependence of the activation rate on the galactose concentration varies with some of the parameters (Figure 2.6a). Likewise it is possible to explore what is the effect of combinations of changes in two parameters (Figure 2.6b). Since all parameters have a physiological interpretation, a change in the value of any of them can be related to some possible mutations with certain effects. For exam-

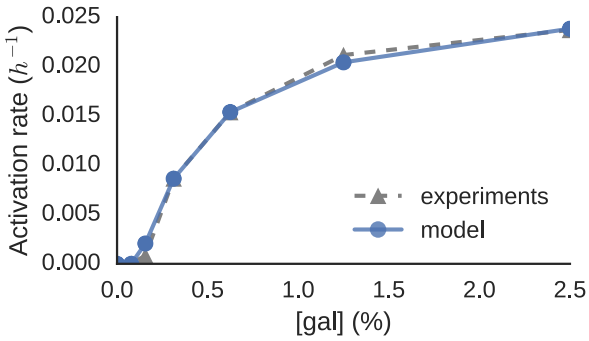

Figure 2.5: Experimental and simulated activation rates with respect to galactose concentration.

ple, the most likely justifications for change in k will be either variations in the promoter or ribosome binding site sequences of the *GAL3* gene, or a change in the structure of Gal4p that affects its ability to initiate transcription. In the same way, it might be possible to obtain information about the mutations of a strain by exploring the parameter changes needed to reproduce its behaviour.

In that sense, however, this specific model might not be able to differentiate between distinct mutations with similar but not identical effects. While the model greatly simplifies the GAL regulatory network, which makes possible to compute the activation rate in a computationally very efficient manner, this simplification may also compromise its ability to capture the effect of specific genetic variants. The reduced set of phenomenological parameters limits the precision of the model to distinguish between mechanistic differences. Furthermore, although the *GAL3* feedback loop is the main cause of the bistability (Acar et al., 2005), which is the most salient property of the system, the model does not explicitly consider some modules of the regulatory network and the effect that

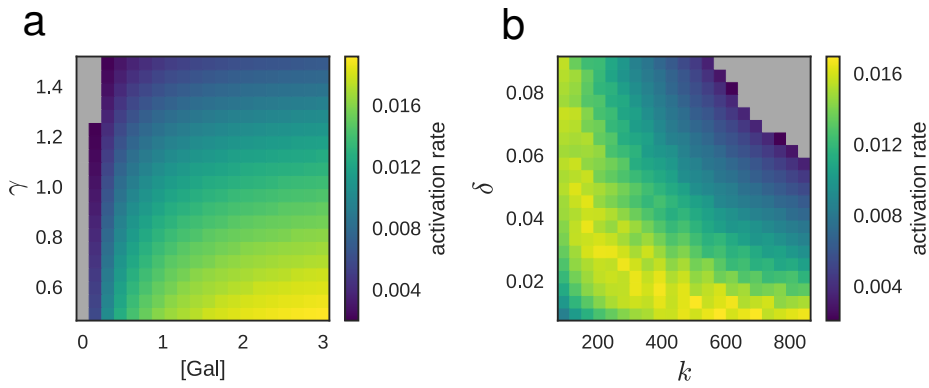

Figure 2.6: Systematic exploration of the activation rates with respect to changes in galactose concentration and other physiological parameters. a, Dependence of the activation rate on the combination of galactose concentration and one of the parameters determined by the genetics of the strain, in this case the equilibrium constant γ for Gal80p translocation to the nucleus (Eq. (2.14)). b, Dependence of the activation rate on the combination of two model parameters determined by the genetics of the strain, in this case the production rate k of Gal3p (Eq. (2.28)) and the equilibrium constant δ for Gal80p binding to Gal4p in the UAS (Eq. (2.15)). In this case the galactose concentration is set to $[gal] = 1\%$.

their dynamics might have on the system behaviour (Johnston, 1987; Bhat and Hopper, 1992; Ramsey et al., 2006; Hawkins and Smolke, 2006; Venturelli et al., 2012). For example, *GAL80* forms part of a negative feedback loop that can reduce the noise of the system (Acar et al., 2005; Dublanche et al., 2006). Also, Gal1p is known to play a regulatory role similar to that of Gal3p (Zacharioudakis et al., 2007; Abramczyk et al., 2012) and neither is considered in the model.

All in all, depending on the specific mutations that we intend to study this model might constitute a very practical instrument to efficiently explore huge regions of the parameter space, or a blunt unprecise tool that

cannot grasp the subtleties of the data. Still, its advantages clearly make it worth considering it. We now consider a more detailed model.

2.3 Mechanistically detailed model of the GAL network

To overcome the limitations of the previous mathematical model we implemented a new approach to characterise, in a more mechanistically detailed manner, how the concentration of galactose modulates the loss of memory in the GAL network. To do so, we switch to a more comprehensive mathematical model introduced by Venturelli et al. (2012). This model explicitly takes into account both the negative and positive feedbacks mediated by *GAL80* and *GAL3*, respectively, as well as the positive feedback loop mediated by *GAL1*.

2.3.1 Deterministic model

The model from Venturelli et al. (2012) not only considers the dynamics of Gal3p, but also explicitly includes those of Gal1p, Gal4p, Gal80p and the complexes that Gal80p forms with each one of the other three proteins. Furthermore, it also makes a different set of assumptions and simplifications. Namely:

- Most noteworthy, based on previous studies (Zacharioudakis et al., 2007; Abramczyk et al., 2012), Gal1p is considered to play an equivalent role to Gal3p in sequestering Gal80p.

- Nuclear and cytoplasmic partitioning of the GAL proteins was not included (Wightman et al., 2008; Peng and Hopper, 2000; Egriboz et al., 2011).

- No explicit distinction is made between the forms of Gal1p, and Gal3p, that are bound and unbound to galactose, as both of them are known to induce GAL gene expression (Bhat and Hopper, 1992).

On the other hand, a few additional simplifications are shared with the model of Acar et al. (2005) discussed earlier in this chapter. First, the intracellular concentration of galactose is considered to remain constant, and the role of Gal2p is excluded from the model. And secondly, the dimerisation of Gal4p and Gal80p are not taken into account.

The dynamics of the GAL regulatory network are described with the following set of equations:

$$\frac{d[G1]}{dt} = \epsilon\alpha_{gal} + \alpha_{G1}\left(\frac{[G4]^{n_1}}{K_{G1}^{n_1} + [G4]^{n_1}}\right) - k_{f81}[G1][G80] + k_{r81}[C81] - \gamma_{G1}[G1] \tag{2.39}$$

$$\frac{d[G3]}{dt} = \alpha_{gal} + \alpha_{G3}\left(\frac{[G4]^{n_3}}{K_{G3}^{n_3} + [G4]^{n_3}}\right) - k_{f83}[G3][G80] + k_{r83}[C83] - \gamma_{G3}[G3] \tag{2.40}$$

$$\frac{d[G4]}{dt} = \alpha_{G4} - k_{f84}[G4][G80] + k_{r84}[C84] - \gamma_{G4}[G4] \tag{2.41}$$

$$\frac{d[G8]}{dt} = \alpha_{oG80} + \alpha_{G80}\left(\frac{[G4]^{n_{80}}}{K_{G80}^{n_{80}} + [G4]^{n_{80}}}\right) - k_{f81}[G1][G80] + k_{r81}[C81] - k_{f83}[G3][G80] + k_{r83}[C83] - k_{f84}[G4][G80] + k_{r84}[C84] - \gamma_{G80}[G80] \tag{2.42}$$

$$\frac{d[C81]}{dt} = k_{f81}[G1][G80] - k_{r81}[C81] - \gamma_{C81}[C81] \tag{2.43}$$

$$\frac{d[C83]}{dt} = k_{f83}[G3][G80] - k_{r83}[C83] - \gamma_{C83}[C83] \tag{2.44}$$

$$\frac{d[C84]}{dt} = k_{f84}[G4][G80] - k_{r84}[C84] - \gamma_{C84}[C84] \tag{2.45}$$

where $[G1]$, $[G3]$, $[G4]$ and $[G80]$ are the concentrations of Gal1p, Gal3p, Gal4p and Gal80p proteins, and $[C81]$, $[C83]$ and $[C84]$ are the concentrations of the complexes formed by Gal80p with either Gal1p, Gal3p or Gal4p, respectively. The different terms account for the basal production of Gal4p and Gal80p (α_{G4} and α_{oG80}), Gal4p-activated production of Gal1p, Gal3p and Gal80p ($\alpha_{Gi}\frac{[G4]^{n_i}}{K_{Gi}^{n_i}+[G4]^{n_i}}$ for $i \in \{1,3,80\}$), the complexation of Gal80p with each of the other three proteins ($k_{f8i}[Gi][G80]$ for $i \in \{1,3,4\}$) and the unbinding of these complexes ($k_{r8i}[C8i]$ for $i \in \{1,3,4\}$). Furthermore, all proteins and complexes experience a linear decay ($\gamma_x[x]$ for $x \in \{G1, G3, G4, G80, C81, C83, C84\}$), mainly due to dilution as a consequence of cell growth. Finally, the model approximates the increased affinity of galactose-bound Gal1p and Gal3p for Gal80p as an increase of the effective amount of Gal1p and Gal3p. This is the interpretation of the terms $\epsilon\alpha_{gal}$ and α_{gal}. These are zeroth-order approximations to the first-order reactions of activation of Gal1p and Gal3p in constant galactose concentration. α_{gal} will be, then, a proxy parameter to indirectly model the effects of different galactose concentrations in the system.

As shown in Figure 2.7, this model exhibits bistability for a wide region of α_{gal}, flanked by two saddle-node bifurcations. It is noteworthy that in the bifurcation diagrams of most of the variables, the basin of attraction of the lower stable branch is smaller than the one of the upper

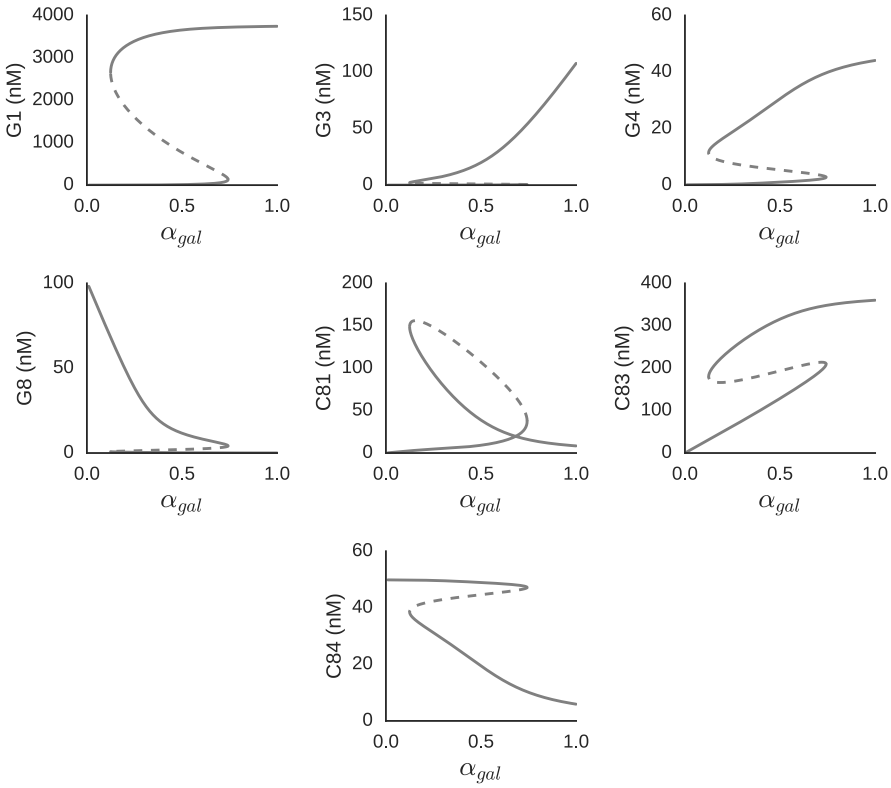

Figure 2.7: Bifurcation diagram of the GAL regulatory network with respect to α_{gal}. The fixed points of each of the seven variables of the model are plotted. These plots are projections of an eight-dimensional system (including the free parameter α_{gal}). Solid lines denote stable branches, while dashed lines denote unstable ones. The system exhibits a bistability region flanked by two saddle-node bifurcations at $\alpha_{gal} = 1.25 \cdot 10^{-1}$ and $\alpha_{gal} = 7.43 \cdot 10^{-1}$ at which one of the two branches collapse with the unstable branch.

one. The high dimensionality of the system, though, makes it very difficult to draw any conclusion from a collection of two-dimensional bifurcation diagrams. In any case, the apparently different size of the basins of attractions of the two metastable states, together with the major impact of noise in the protein copy number at low concentrations, is compatible with a much higher frequency of transitions from the OFF to the ON states than the other way around.

2.3.2 Computing the switching rate with stochastic simulations

For this model we also need a way to accurately compute the rate of noise-driven transitions from the low to the high-expression metastable states. Since this is a multidimensional system, i.e. its state is defined by more than one variable, the approach described in Section 2.2.2 to compute the switching rate between the two metastable states is no longer valid.

Instead, we will simulate enough exact stochastic trajectories and compute the switching rates from them. To do so, we will use the Stochastic Simulation Algorithm (SSA) also known as Gillespie algorithm (Gillespie, 2007). In the first place we need to reformulate the model in terms of individual reactions. Then, we need to define the probability that each one of the reaction occurs somewhere inside the volume V in the next infinitesimal time interval. In other words, the propensity of each individual reaction needs to be expressed using extensive quantities: in terms of the total amount of reactants instead of their concentration.

As an example, to describe the dynamics of Gal1p, instead of relaying on its concentration $[G1]$, we must do it in terms of the total number of

molecules $N_{G1} = [G1]V$, where V is the constant total volume:

$$\begin{aligned}\frac{d[G1]}{dt} &= \frac{1}{V}\frac{dN_{G1}}{dt} \\ \frac{dN_{G1}}{dt} &= \epsilon\alpha_{gal}V + \alpha_{G1}V\left(\frac{N_{G4}^{n_1}}{(K_{G1}V)^{n_1} + N_{G4}^{n_1}}\right) - \frac{kf81}{V}N_{G1}N_{G80} \\ &\quad + k_{r81}N_{C81} - \gamma_{G1}N_{G1}\end{aligned} \tag{2.46}$$

Then, we need to explicitly define each one of the individual reactions and their propensities:

$$\begin{aligned}
&\alpha_{gal}\epsilon V && \varnothing \longrightarrow G1 \\
&\alpha_{G1}V\frac{G4^{n_1}}{(K_{G1}V)^{n_1} + G4^{n_1}} && G4 \longrightarrow G1 + G4 \\
&\frac{k_{f81}}{V}G80G1 && G80 + G1 \longrightarrow C81 \\
&k_{r81}C81 && C81 \longrightarrow G80 + G1 \\
&\gamma_{G1}G1 && G1 \longrightarrow \varnothing
\end{aligned}$$

where, for the sake of simplicity, we express the number of molecules as $G1 \equiv N_{G1}$, and equivalently for the other species. While the SSA, in principle, can only consider elementary reactions, we have included a Hill equation in a propensity expression, summarizing a multi-step reaction. It has already been shown that this is a valid approach for Michaelis-Menten processes, and that similar derivations could be provided also for other complex reaction mechanisms known from deterministic kinetics (Gend et al., 2001; Rao and Arkin, 2003; Sanft et al., 2011), at least as long as it is not an instance of self-regulatory dynamics (Bundschuh et al., 2003).

The rest of the reactions and their propensities can be obtained in a similar way:

$$
\begin{array}{ll}
\alpha_{gal}V & \varnothing \longrightarrow G3 \\
\alpha_{G4}V & \varnothing \longrightarrow G4 \\
\alpha_{oG80}V & \varnothing \longrightarrow G80 \\
\alpha_{G3}V\dfrac{G4^{n_3}}{(K_{G3}V)^{n_3}+G4^{n_3}} & G4 \longrightarrow G3+G4 \\
\alpha_{G80}V\dfrac{G4^{n_80}}{(K_{G80}V)^{n_80}+G4^{n_80}} & G4 \longrightarrow G80+G4 \\
\dfrac{k_{f83}}{V}G80G3 & G80+G3 \longrightarrow C83 \\
\dfrac{k_{f84}}{V}G80G4 & G80+G4 \longrightarrow C84 \\
k_{r83}C83 & C83 \longrightarrow G80+G3 \\
k_{r84}C84 & C84 \longrightarrow G80+G4 \\
\gamma_{G3}G3 & G3 \longrightarrow \varnothing \\
\gamma_{G4}G4 & G4 \longrightarrow \varnothing \\
\gamma_{G80}G80 & G80 \longrightarrow \varnothing \\
\gamma_{C81}C81 & C81 \longrightarrow \varnothing \\
\gamma_{C83}C83 & C83 \longrightarrow \varnothing \\
\gamma_{C84}C84 & C84 \longrightarrow \varnothing
\end{array}
$$

Once we know the propensity of each reaction we could define the Chemical Master Equation of the system, which describes the probability that each species has a specified number of molecules at a given time. With seven species, though, this equation becomes intractable. Instead,

it is also possible to describe the joint probability that, given the current state of the system, the next reaction to occur will be of a certain type and that will occur at a certain time. This is the approach the SSA exploits to produce statistically correct trajectories of the system (Figure 2.8).

Finally, by sampling enough of these possible trajectories, the rate of escape of the metastable states of the system could be estimated.

2.3.3 Missing information

The conversion of a deterministic model of the dynamics of concentrations of GAL proteins to our stochastic model has a shortcoming: the volume factor V is not defined. The *real* volume of *S. cerevisiae* cells is 40-60 femtoliters (Bryan et al., 2010). On the other hand, the protein concentrations predicted by the model from Venturelli et al. (2012), while realistically plausible, are arbitrary and could be rescaled; only the ratio

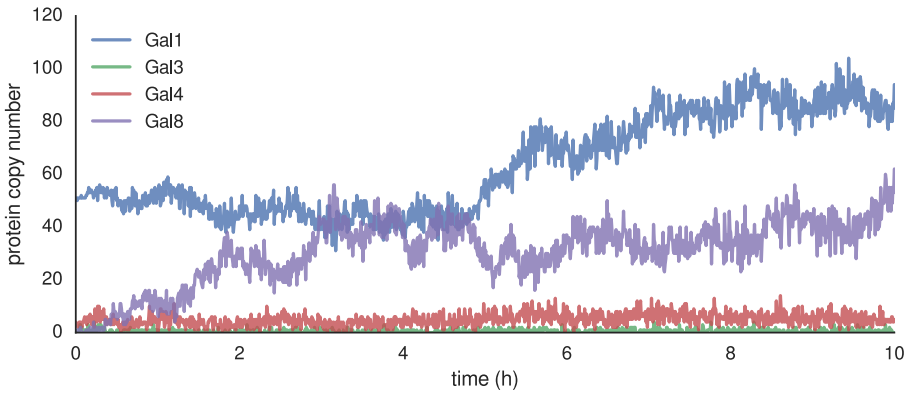

Figure 2.8: Stochastic simulation of the time evolution of the protein copy number of the GAL regulatory proteins. Only the populations of proteins in their free form are shown: the amount of complexes formed either by Gal80p and Gal1p, Gal80p and Gal3p or Gal80p and Gal4p are not included in the figure.

between them is relevant. Thus, V must not be understood as the *real* volume, but as the scaling factor that relates the protein concentrations predicted by the deterministic model with the real protein copy number.

The value of V will have a crucial effect on the level of noise in the dynamics of the system. Since here stochasticity is a consequence of the finite character of the number of molecules in the system, the smaller V is, the smaller the number of molecules and the higher the noise will be. It is thus necessary to establish the value of V in order to compute the reaction propensities. To do so, we need to estimate the relationship between the copy number of GAL proteins in yeast cells and the prediction of the deterministic model under comparable conditions.

Besides, a limitation of the model taken from Venturelli et al. (2012) is that it does not explicitly incorporate the galactose concentration. Instead, they use the rate of activation of Gal1p and Gal3p (α_{gal}) as a proxy parameter to represent the effects of different concentrations of galactose. They do not establish, however, what is the mapping between the values of α_{gal} and the actual galactose concentrations. Furthermore, the growth medium used by Venturelli et al. (2012) contains raffinose as alternative carbon source to galactose while ours has glucose, and glucose is known to affect significantly the GAL regulatory network (Johnston et al., 1994).

Indeed, comparing the dependence of the activation rate on α_{gal} for an arbitray volume V with the experiments, it becomes evident that the dependence of α_{gal} on the concentration of galactose is not trivial. As an example, Figure 2.9a shows the experimentally fitted rates together with the rates measured in stochastic simulations, assuming an arbitrary linear dependence of α_{gal} on galactose, $\alpha_{gal}([gal]) = 3[gal]$. Even though in both cases the rates monotonically increase with galactose, it is clear that the activation rate of Gal1p and Gal3p, α_{gal}, must be a nonlinear function

of galactose concentration. It is thus necessary to define this mapping in order to compare the simulation results with the experiments.

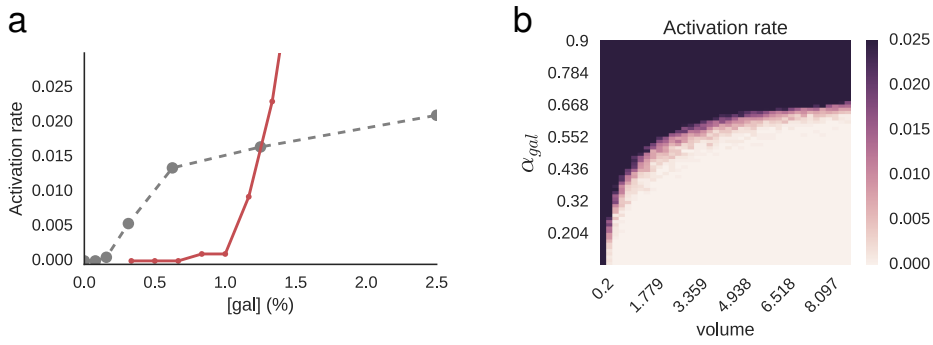

Figure 2.9: Dependence of the activation rate on V and α_{gal} with respect to the experimental observations. a, Activation rate with respect to galactose concentration as measured in the experiments (grey) and in the stochastic simulations assuming $\alpha_{gal}([gal]) = 3[gal]$ and $V = 1$ (red). b, Activation rate with respect to α_{gal} and volume V. To emphasise the comparison with the experimental rates the colormap saturates at a maximum activation rate $k_{OFF\to ON} = 0.25$, the largest experimental rates measured, although the simulated values become larger. In any case, for all the range of V explored the dependence of the rate with respect to α_{gal} in $k_{OFF\to ON} \in [0, 0.25]$ is defined by a concave function (as shown in a).

However, the effects of V and the $[gal]$ to α_{gal} mapping are highly interrelated (Figure 2.9b) and it is not easy to find one while missing the other. In the following sections we discuss how we approached the situation. First, we present our methodology to estimate the number of proteins in a yeast cell, which we need in order to find V. Then, we describe how we fitted both V and the $[gal]$ to α_{gal} mapping iteratively, using convergence as stopping criterion.

Protein copy number estimation

To obtain an estimate of the number of proteins of the GAL regulatory network we used RNA sequencing (RNA-seq) data obtained by our experimental collaborators. In short, an RNA-seq experiment consists in (i) extracting the mRNA from a biological sample, (ii) breaking it into fragments of a few hundred base pairs, and (iii) sequencing all or a subset of those fragments randomly, so that the probability of obtaining a read of a given sequence is proportional to its abundance (Wang et al., 2009). In our case we used an Illumina HiSeq sequencer at a 100x coverage (Sims et al., 2014). We employed random primers so that all transcripts are targeted. The samples were lysates of yeast cultures, having two biological replicates for four different growth conditions, namely:

- GLU: 14 hours in 0.5% glucose and 2.5% sorbitol
- GAL: 14 hours in 2.5% galactose and 0.5% sorbitol
- GLU+gal: 14 hours in 0.5% glucose and 2.5% sorbitol plus 14 hours in 0.5% glucose and 2.5% galactose
- GAL+glu: 14 hours in 2.5% galactose and 0.5% sorbitol plus 14 hours in 0.5% glucose and 2.5% galactose

RNA-seq data gives quantitative information about the relative abundances of the mRNAs in the cell. The number of reads belonging to the transcript of each known gene needs to be normalized by the length of the transcript and the total number of reads, obtaining the *fragments per kilobase per million* measure or FPKM (Mortazavi et al., 2008). Thus, FPKM express the relative abundance of that specific transcript among the total number of transcripts. The relevance of this measure is that the

ratio between the FPKM of two genes in a RNA-seq experiment should be proportional to the relative abundance of mRNA transcripts of these two genes.

Consequently, with this data it is possible to infer the abundance of GAL proteins if we know the copy number of some other reference proteins. To do this we need to assume that the post-transcriptional regulation of all those genes do not differ significantly. In other words, we need to assume that the ratio between mRNA abundance and protein copy number is approximately constant for the selected proteins. Then, we can estimate the abundance of GAL proteins knowing the abundance of a protein whose expression is not affected by glucose or galactose. The publicly available database YeastGFP lists the copy number measured in *S. cerevisiae* cells of more than 3800 proteins (Ghaemmaghami et al., 2003). Yet, it is necessary to define what proteins are adequate to be used as references.

Those genes that have a constant expression, and thus are good references, will show proportional changes in FPMK across samples. FPKM is a relative measure: a difference in FPKM of a given gene reflects that the fraction of the total mRNA being transcripts of that gene has changed. This can be caused by a change in the number of transcripts of that gene or by a change in the total number of transcripts of all genes. And certainly, GAL genes can significantly change the total number of transcripts: when active they can increase its expression more than 1000-fold and a few of them can produce up to 1% of the total mRNA of the cell (Lohr et al., 1995).

We assumed, on the other hand, that for the vast majority of genes the level of expression will not depend on galactose and glucose concentration. For these genes the differences in the total number of transcripts will

be the sole cause of fluctuations in FPKM, and will affect all of them in a similar manner. Thus, we expect to identify appropriate reference genes by finding a large set of nodes with a high correlation in their changes of expressions across samples.

Taking these considerations into account, we computed the Pearson correlation of the FPKM of genes across samples. Figure 2.10a shows that the distribution of correlation coefficients is enriched for positive values. This indicates that either most of the genes change their expression levels proportionally in response to galactose and glucose concentrations or that, as we assumed, their expression levels remain constant in a proportional manner (Figure 2.10b).

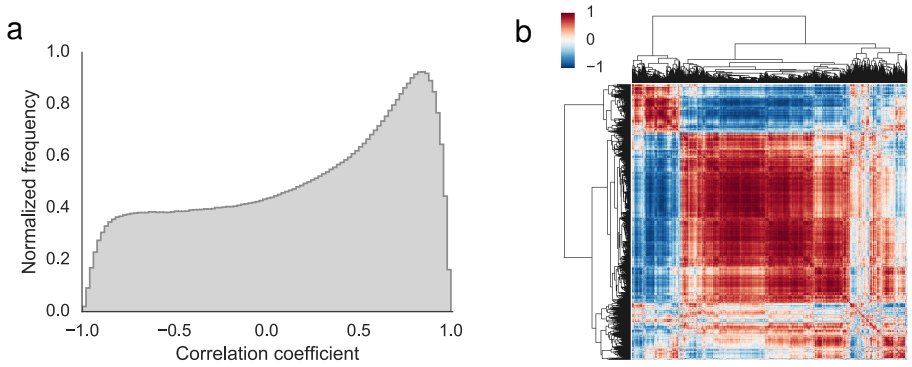

Figure 2.10: Analysis of the Pearson correlation coefficients of the RNA-seq gene expression data. a, Distribution of the Pearson correlation coefficients of the FPKM measured for each gene across conditions. The population of correlation coefficients is enriched for values larger than 0.5. b, Matrix of correlation coefficients after a hierarchical clustering of the rows and columns. Each row and column corresponds to a gene, and the order of genes in rows and columns is the same. There is a large group of genes with a high degree of correlation among them.

Then, we considered a graph defined by the adjacency matrix A:

$$a_{i,j} = \begin{cases} 1 & \text{if } \mathrm{corr}(FPKM_i, FPKM_j) > 0.98 \\ 0 & \text{otherwise} \end{cases} \tag{2.47}$$

where the 6364 nodes are genes and the edges indicate a high degree of correlation. Then, we extracted the largest connected component of the graph, composed by 2738 nodes (the second and third largest connected components had 447 and 12 nodes, respectively). Out of those, for 1793 of them there is copy number information available in YeastGFP. These 1793 genes, showing a high degree of correlation among them and representing above 40% of the network, were used as reference genes to estimate the abundance of GAL genes.

Figure 2.11 shows the distribution of estimates obtained using each one of the reference genes and each one of the biological replicates. We used the median of each distribution as the copy number estimate of the GAL proteins.

Iterative fitting by convergence criterion

To calculate the volume factor V we need to compare the actual protein copy number with concentration predicted by the deterministic model under equivalent conditions. We obtained protein copy number estimates for some experimental galactose concentrations using RNA-seq data. On the other hand, the concentrations predicted by the model for any α_{gal} value can be obtained by many numerical approaches: integration of the system dynamics, continuation of steady states or even finding the solution of the ODE system when all derivatives are equal to zero. Yet, to define what are equivalent conditions between both results, the relationship between

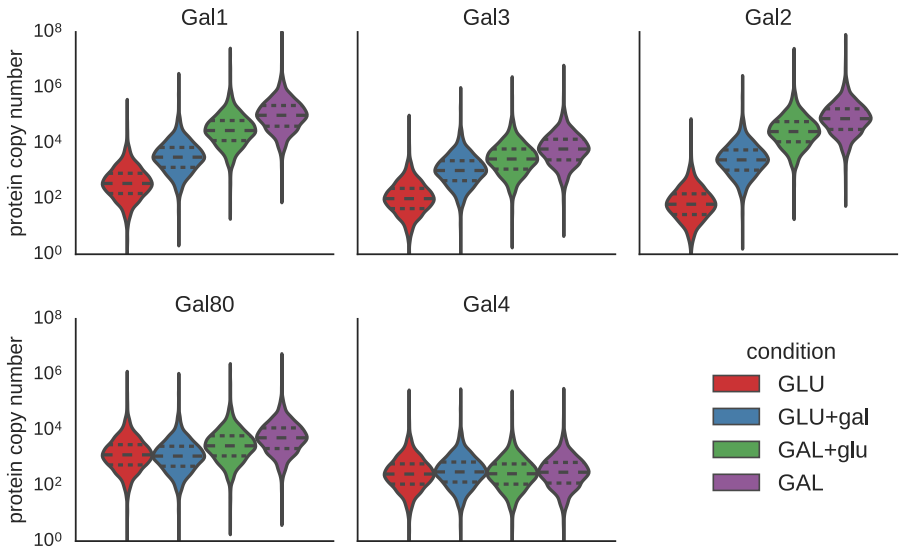

Figure 2.11: Distribution of predicted protein copy number for each GAL protein using a population of reference proteins. For each condition and GAL protein, the 1793 reference proteins are used to obtain as many estimates of protein copy number. The different conditions correspond to the growth conditions of the yeast culture used in the RNA-seq experiment. The tested growth conditions are glucose only (GLU), a mixture of glucose and galactose after a pregrowth in glucose (GLU+gal), a mixture of glucose and galactose after a pregrowth in galactose (GAL+glu) and galactose only (GAL). The horizontal lines inside each violin correspond to the first, second and third quartile.

the concentration of galactose and the value of α_{gal} is still missing. More precisely, it is necessary to define the α_{gal} values that correspond to the experimental conditions $[gal] = 0$ and $[gal] = 2.5\%$

On the other hand, to define the equivalence between $[gal]$ and α_{gal} values we used the stochastic model and the fitted experimental rates from Section 2.1.2. Specifically, we computed the switching rate for 60 evenly spaced values of α_{gal} in the range between $\alpha_{gal}([gal] = 0)$ and

$\alpha_{gal}([gal] = 2.5)$ (Figure 2.12a). To do so we simulated the system in its low-expression state for 6000 hours and measured the frequency of noise-induced activation events. This data was used to define a new mapping relating the experimental galactose concentrations to values of α_{gal} that would produce the same switching rate. Nevertheless, it is again necessary to set a value for V in order to compute the *in silico* switching rates.

To overcome the interdependence of both fitting strategies, we used an iterative approach and convergence as stopping criterion. First, an arbitrary mapping $\alpha_{gal} = [gal]$ was used to approximate the value of V. Then, this value of V was used to compute how the rate changes as a function of α_{gal} (Figure 2.12a), which served to define a new mapping from $[gal]$ to α_{gal} (Figure 2.12b). Finally, this process was repeated iteratively until convergence was reached.

2.3.4 *Interaction of genetic mutations*

With this model, once more, we are able to reproduce quantitatively how the galactose concentration modulates the frequency of stochastic transitions of the regulatory circuit of the GAL system. Again, we are in position to explore how specific mutations can affect the way the system reacts to galactose. In this case, however, we can fully describe the mechanistic effect of mutations in the circuit regulation. Additionally, with this new model it is possible to consider mutations that affect the dynamics of all the core proteins of the GAL regulatory circuit, and not only Gal3p. This can be crucial to explore the effects of the combination of mutations whose effects would be summarized by the same phenomenological parameter. As an example, Figure 2.13 shows the systematic exploration of changes in basal production rate of Gal80p α_{0G80} and the maximal upregulation of the production rate of Gal80p α_{G80}.

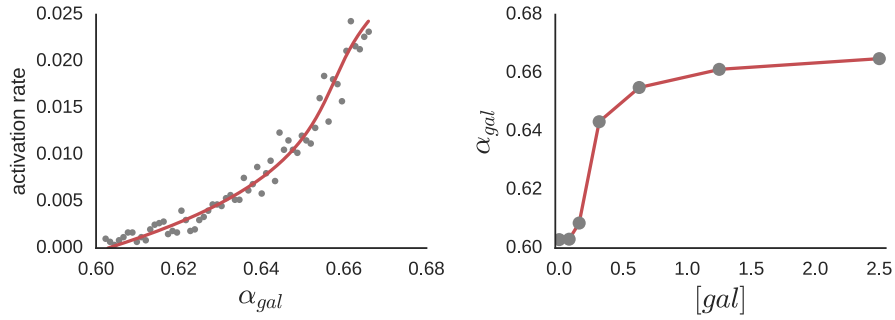

Figure 2.12: Results of the iterative fitting of volume and galactose to α_{gal} mapping. a, Activation rates measured in 60 simulations (grey) for a range of α_{gal} values with a given volume V and the interpolation with an spline (red). b, Estimated shape of the $\alpha_{gal}([gal])$ function, obtained by interpolating the value of α_{gal} that would produce the activation rate that was observed in the experiments under a given concentration of galactose $[gal]$. The grey dots correspond to the actual values of $[gal]$ for which there is experimental data available and the red line is a piecewise linear interpolation. Both a and b are produced with $V = 7.31$ fL, which corresponds to the final volume obtained after the iterative fitting process.

2.4 Discussion

In this chapter we have quantified the stability of memory in the GAL network. The GAL regulatory circuit regulates its activity in a switch-like behaviour responding to the galactose concentrations in the medium. For a range of galactose concentrations it displays hysteresis, so that the state of the system depends on its history. In this context, stochastic transitions can force the loss of this memory. We characterized the relationship between the external galactose concentration and the probability of such stochastic transition in our experimental conditions for the YPS128 strain of *Saccharomyces cerevisiae*. The results suggest that even when

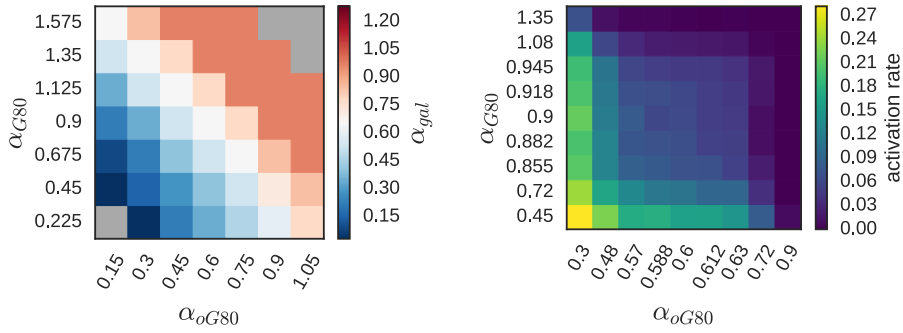

Figure 2.13: Systematic exploration of two parameter changes in the stochastic mathematical model. The combined effect of changes in two model parameters are studied, in this case the basal production rate of Gal80p α_{0G80} and the maximal upregulation of the production rate of Gal80p α_{G80}. The plots show the α_{gal} necessary to maintain the same activation rate of $0.015h^{-1}$ (left) and the dependence of the activation rate for $[gal] = 1\%$ (right) on the value of the two parameters analysed.

it is displaying hysteresis, the GAL regulatory circuit is incorporating information of the nutrients available to control the robustness to noise of its current state.

Furthermore, we developed and adjusted two models with the final goal of exploring how genetic variants affect the behaviour of this robust memory circuit. Far from being redundant, the two models are complementary as they have different sets of advantages and limitations. The first one, based on the work by Acar et al. (2005), is computationally cheap and uses a reduced set of phenomenological parameters to describe the memory circuit, which makes it an appropriate tool for coarse-grained exploratory studies. The second one, on the other hand, despite being computationally more expensive, captures more accurately the interactions in the GAL circuit and makes it possible to distinguish between related mechanistic traits.

Peng et al. (2015) studied the combinatorial effects of genetic variants in the activation of the GAL system. However, they focused their study in the effects of variability exclusively in the promoter regions of *GAL2*, *GAL3*, *GAL4* and *GAL80*. Additionally, their phenotype of interest was the general response curve of activation of the GAL network in response to galactose concentration. Instead, we want to characterise specifically the decision making process that takes place when the network is in its bistable regime. Furthermore, our aim is to describe how mutations, individual and combined, modify the dynamical features of the system that modulate its stability in response to external cues. The model dissection presented in this Chapter should be understood as a first step in this direction.

3 State-dependent computation in gene regulatory networks

The survival of any cell, either as an individual being or as part of a multicellular organism, depends on its capacity to respond to changes in the environment. From stress response cellular programs to embryogenesis driven by morphogens, and from the immune response to metabolism adaptation to varying energy sources, cells need to sense multiple signals in their surroundings, integrate them, and activate an adequate response. Orchestrating the best possible response with the right intensity is crucial, but so is doing it at the right moment and fast enough. The importance of timing and speed implies that cells able to anticipate changes in the environment have a critical advantage.

Although most changes in the environment are stochastic from the point of view of a cell, many others are predictable. In many cases, the likelihood of future events is encoded by the recent history of the environment. In these cases, the ability to take the temporal information into account gives a clear advantage. Periodic changes in the environment, for example, can be anticipated through molecular oscillators or cellular clocks, as seen in the way cyanobacteria anticipate daily light-dark cycles (Golden et al., 1997; Mori and Johnson, 2001). Another example is given

by groups of events that tend to occur together or in a specific order. This kind of association, for instance, allows the bacteria *Escherichia coli* to prepare for oxygen depletion when they sense an increase in temperature, an indication that they have been ingested by a mammal (Tagkopoulos et al., 2008). Similarly, enterobacteria anticipate sequential changes in sugars as they pass through the intestinal tract, and yeast expects a specific sequence of stresses during alcoholic fermentation (Mitchell et al., 2009). Another example is given by *Candida albicans* cells, a fungal pathogen, that upregulate oxidative stress resistance genes when they detect higher glucose levels, which indicate they have entered a host, preemptively reacting to the immune system of the host (Schild et al., 2007). Furthermore, experimental evolution studies have shown that predictive environmental sensing can evolve in relatively short periods of time in a laboratory setting (Dhar et al., 2013).

Beyond the ability to associate concurrent events, recent studies have shown that microbes have both short-term and long-term memory mechanisms that influence cellular decisions. The stress response of *Bacillus subtilis*, for example, depends not only on the condition in which it is currently growing, but also on past growth conditions (Wolf et al., 2008). However, the way this record of previous history –i.e. memory– is integrated and stored in cells is not yet fully understood. Knowledge of the conceptual limits of this cellular memory is also scarce. Consequently, because memory is a key limitation to recognizing temporally varying signals, the prediction capabilities of cells are still to be delimited.

While some of the simplest prediction mechanisms involve only a handful of elements –e.g. molecular oscillators–, adaptability of cells relies on a complex network of interactions between molecules (Mattick, 2001; Stelling et al., 2002). Here we hypothesize that the structure of

this network establishes how memory is encoded. We aim to determine how these cellular regulatory networks integrate complex inputs, and especially how they process complex temporal information. After analysing various gene regulatory networks, we propose that they encode temporal information in a state-dependent manner: the complex transient dynamics of the system encodes the recent history. Specifically, we postulate that the gene regulatory networks work as a reservoir computing system.

Reservoir computing is a functional neural-network organization that can process temporal information while featuring a very efficient learning process. The key characteristic of reservoir computing —also known as *liquid state machines* in neuroscience (Maass et al., 2002) and *echo state networks* in machine learning (Jaeger, 2001)— is that it separates memory encoding and prediction in different network substructures (Figure 3.1). First, a substructure with recurrent connections is needed to encode history: the reservoir. It is well known that recurrences —i.e. a cyclic paths— allow the network to retain information for a certain time, giving memory to the system. Technically, the reservoir projects the stimuli non-linearly into a high-dimensional space (Buonomano and Maass, 2009; Maass et al., 2002; Jaeger, 2001). The second substructure, the readout, is a feed-forward architecture —i.e. directed acyclic graph— placed downstream of the reservoir. The readout layer uses the history record encoded in the state of the reservoir to make a prediction or a classification. Again, technically it separates the trajectories in the high-dimensional space. Feed-forward structures, lacking cyclic paths, are much easier to train —i.e. to adapt the strength of the interactions between its elements so that they produce the expected dynamics. Thus, this separation of roles allows to focus the training process solely on the readout, giving reservoir computing the computational power of a recurrent

network and the ease of training of a feed-forward one (Buonomano and Maass, 2009). Furthermore, by adding independent readouts the system it can gain the ability to perform additional tasks without interfering with the existing ones (Lukoševičius and Jaeger, 2009).

In this study we propose that gene regulatory networks can follow a reservoir computing organization. To do so, we first analyse the topology of the gene regulatory networks of five distant organisms to assess if they are compatible with the reservoir computing paradigm. Furthermore, we inquire how efficient these networks are at encoding recent history, and whether this capability can be attributed to the reservoir-like structures found. We then investigate if the information arriving through different biological stress signalling pathways can be stored in the reservoirs. Finally, to answer whether the training of the readout could occur in a biological context we show that it can be done using evolutionary algorithms.

3.1 Reservoir computing structure in gene regulatory networks

We analysed the gene regulatory networks from five distant organisms as representative cellular regulatory networks: *Bacillus subtilis*, *Escherichia coli*, *Saccharomyces cerevisiae*, *Drosophila melanogaster* and *Homo sapiens*. The regulatory interaction data was extracted from five different publicly available databases and publications. Data for *B. subtilis* was obtained from DBTBS (Sierro et al., 2008). Data for *E. coli* was extracted from EcoCyc (Keseler et al., 2011), including the sigma factors as transcription factors. Data for *S. cerevisiae* was obtained from YEASTRACT (Teixeira et al., 2014). The gene regulatory network for *D. melanogaster* was obtained from the modENCODE initiative (Roy

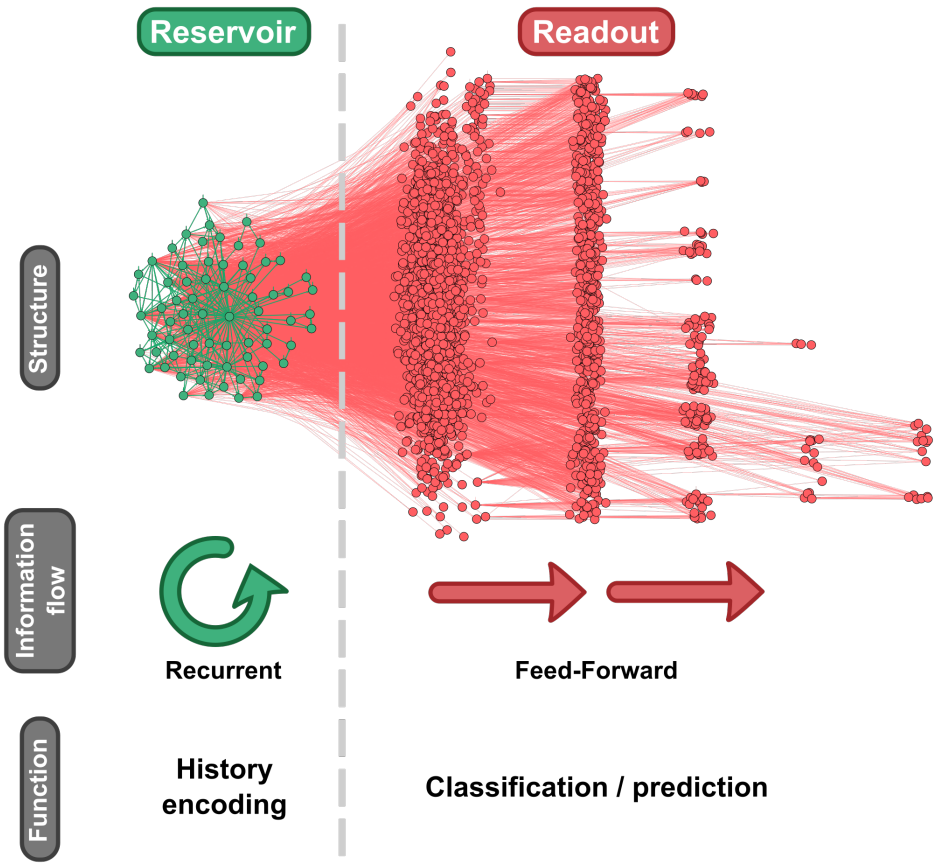

Figure 3.1: Structural and functional organization of reservoir computing. The reservoir (green) is a subgraph with cyclic paths that allow it to maintain a record of the recent history in its dynamics. The readout (red) is a directed acyclic subgraph that reads the information encoded in the reservoir state to perform a given task. The network shown is the reservoir and readout from the *Escherichia coli* network (see Section 3.1). The nodes in the readout are grouped by the length of the longest path from the reservoir to them.

et al., 2010). Finally, data for *H. sapiens* was extracted from the ENCODE project (Gerstein et al., 2012).

We restricted ourselves to transcriptional networks because in other types of regulatory networks —e.g. protein-protein interaction networks— the directionality of the interactions, and thus of the flow of information, is not so well documented. Table 3.1 shows some structural descriptors of these gene regulatory networks. Additionally, the degree distribution of all the networks shows that they have a non-trivial structure, resembling in most of the cases a power-law distribution (Figure 3.2). In the case of *Saccharomyces cerevisiae*, the bump observed in the degree distribution plot (Figure 3.2c) is likely an artefact: since most of the data in this database comes from compiling a large number of low throughput studies, nodes with lower degree can be expected to be under-represented, as studies tend to focus on genes involved in more regulatory interactions.

Each network has a single main recurrent structure. Despite the complexity and large size of the networks, the subgraphs containing recurrent connections are the ones considered to be relevant for the computational capabilities of a reservoir (Rodan and Tino, 2011). The networks were simplified by removing all the strictly feed-forward nodes, obtaining a minimal recursive subgraph, i.e. a subgraph containing only the nodes and edges that form recursive structures and the nodes that interconnect this structures. To do so we *pruned* the networks by iteratively removing any node that had either in-degree or out-degree equal to zero, until no more nodes could be removed. The resulting subgraph is what will be referred from now on as *core* or *reservoir*. As the reader may have noticed, the recurrent structures obtained are not necessarily strongly connected components. Rather, they comprise all the strongly connected compo-

	Nodes	Edges	Self loops	Mean degree
Whole graph				
B. subtilis	886	1358	49	3.06
E. coli	3236	8366	126	5.17
S. cerevisiae	6725	201972	197	60.06
D. melanogaster	9432	231174	0	49.01
H. sapiens	16354	163271	28	19.96
Recurrent core				
B. subtilis	13	30	7	4.61
E. coli	70	317	55	9.05
S. cerevisiae	289	9046	195	62.60
D. melanogaster	486	23470	0	96.58
H. sapiens	207	1434	26	13.85

Table 3.1: General properties of the gene regulatory networks studied here and their recurrent cores.

	Number of CC	Size 1st CC	Size 2nd CC	Number of isolated nodes
B. subtilis	18	886	5	3
E. coli	1201	3236	6	1199
S. cerevisiae	1	6725	0	0
D. melanogaster	4	9432	4	0
H. sapiens	2	16354	2	0

Table 3.2: Connected components in the gene regulatory networks. Number of connected components, size of the two largest connected components, and number of connected components that consist in a single node. In all the gene regulatory networks analysed there is one big connected component and the second largest connected component, if present, has only a few nodes. Even though the number of single unconnected genes in the *E. coli* network is large, it is still smaller than the number of nodes in the largest connected component.

nents and any path between them. Table 3.1 shows some descriptors of the recurrent cores.

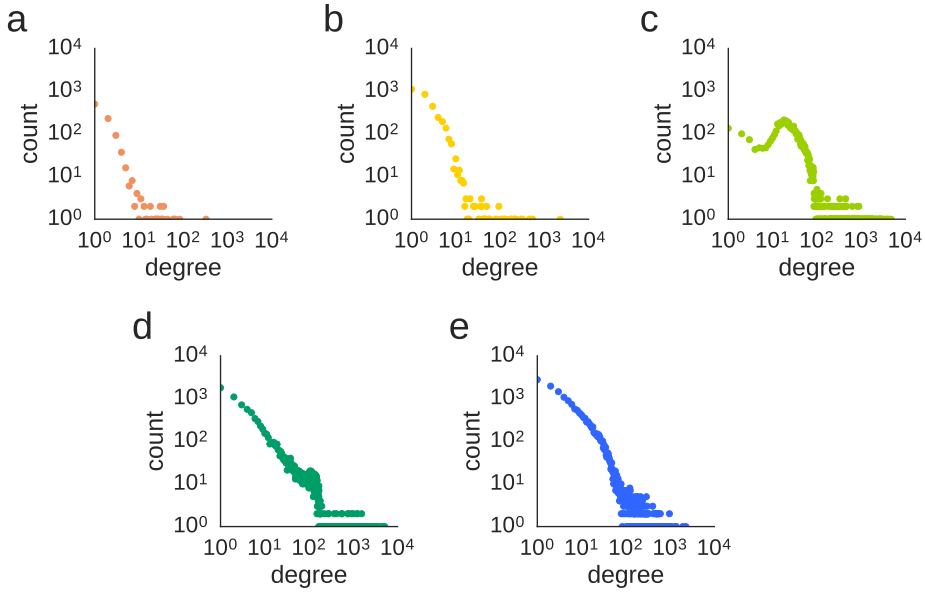

Figure 3.2: Degree distribution of the gene regulatory networks. Degree distribution of the gene regulatory networks studied here, on a log-log scale: *B. subtilis* (A), *E. coli* (B), *S. cerevisiae* (C), *D. melanogaster* (D) and *H. sapiens* (E).

Despite the small size of the core subgraphs relative to the whole network, their location is central. Table 3.3 groups the nodes in each network depending on whether they are part of the reservoir or are placed downstream of it. Additionally, a distinction is made between the downstream nodes that regulate other nodes and those that have out-degree equal to zero, which are named *terminal* nodes. As can be observed, the vast majority of nodes are placed downstream of the recurrent core, forming what we called the *readout*. This is further emphasised in Figure 3.3, where the relative size of each group of nodes is compared.

The obvious implication of the central location of the core is that the dynamics of most of the network is affected by it. It is worth to note again that by definition there are no recurrences outside the reservoir, and thus all these readout nodes cannot affect back the reservoir. Furthermore, Table 3.3 and Figure 3.3 also shows that a very large proportion of the readout nodes are terminal —i.e. nodes with no output connections. That limits the potential complexity that the readout topology can have, and thus the ability to process information, giving an even more central role to the recurrent core or reservoir. A more detailed characterization of the nodes using the bow-tie (Broder et al., 2000) classification is discussed in Appendix A.

	Reservoir	**Readout**	**Terminal**	**Other**
B. subtilis	13	537	500	336
E. coli	70	3133	3025	33
S. cerevisiae	289	6436	6419	0
D. melanogaster	486	8795	8721	151
H. sapiens	207	13497	13449	2650

Table 3.3: Number of nodes by category in the largest connected component of each network. Reservoir nodes are those that form the recurrent core of the network after pruning it; readout nodes are those that can be reached from the reservoir following the directed edges; and Terminal are the nodes from the readout that do not affect any other node.

3.2 Dynamical encoding of information in the biological reservoirs

Next, we inquired if the topologies of these recurrent cores are able to encode temporal information in their dynamics. To do so we confronted

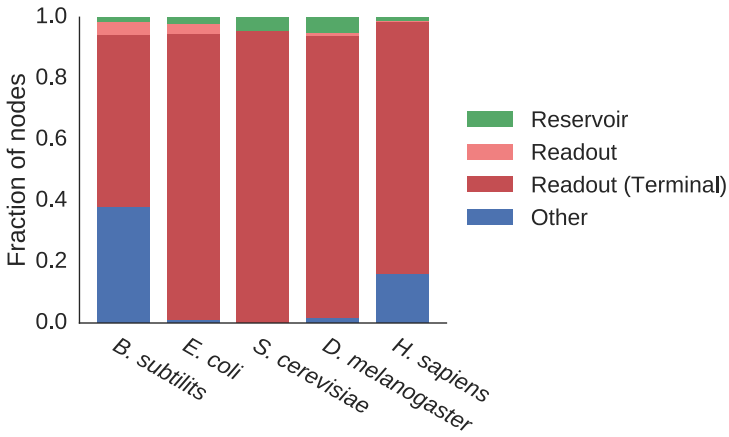

Figure 3.3: Relative size of functional groups of a reservoir for each network. The fraction of the total number of nodes that belong to the each of functional groups of a reservoir are shown. Reservoir nodes are the ones left over after the pruning of the network. The nodes placed downstream of the reservoir are assigned to the readout structure, distinguishing between the terminal ones, that have zero out-degree, and the rest. Finally, all nodes that do not fall in any of the previous groups are counted as *others*.

them to the 10th order NARMA task, a memory demanding benchmark commonly used in the context of neural networks. To test if the dynamics of the network cores can represent the recent history, the network was simulated with simplified dynamics and a time-varying random input was applied to it. Then, an *ad hoc* readout node was trained to compute the output of the 10th order NARMA system using uniquely the instantaneous state of the network (Figure 3.4). The challenge is that the 10th order NARMA system depends on the input and output values of the last 10 time steps. Thus, only if this information about the recent past is encoded in the reservoir state will the readout node be able to accurately model the NARMA system.

3.2.1 Reservoir dynamics

The dynamics of the gene regulatory network cores were simulated using a discrete time updating rule defined as

$$x_{i,t+1} = \tanh(v_i z_t + \sum_{j=1}^{n} w_{ji} x_{j,t}) \tag{3.1}$$

where z_t is the time varying input signal that reaches the system; $x_{i,t}$ is the state of the ith node of the reservoir at time t; n is the number of nodes in the reservoir; W is the weighted adjacency matrix of the reservoir, so that w_{ji} is the weight of the link from the jth to the ith node; and V is the input weight vector so that v_i is the weight of the link from the input to the ith node (Figure 3.4). The values of the vector V are randomly chosen to be either -0.05 or 0.05. At the same time, the values of the matrix W are real random numbers drawn from a uniform distribution between -1 and 1 if the link exists, and 0 otherwise. Additionally, the W matrix was normalized to have a spectral radius of 0.9 to ensure the echo state property, which means that the effect of initial conditions should vanish as time passes (Jaeger, 2001; Lukoševičius and Jaeger, 2009).

The dynamics of the reservoir is then fed into an *ad hoc* readout node. Except were otherwise stated we used as readouts ridge regression nodes. A ridge regression readout computes a weighted sum of the state of the nodes it receives information from (Figure 3.4):

$$\tilde{Y} = W^{out} X \tag{3.2}$$

where X is a matrix of the $x_{i,t}$ states of the ith node at time t, W^{out} is a vector of the w_i^{out} weights given to the ith node by the readout, and $\tilde{Y}$ is a vector with all predicted outputs over time.

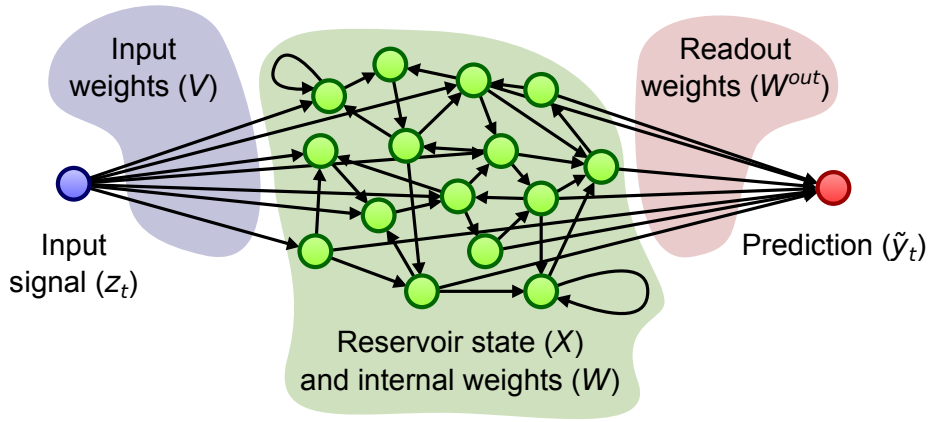

Figure 3.4: Setup to test the memory of a network A reservoir is built with a connectivity matrix W that defines the topology of a given network. A signal z_t arrives to the nodes of the reservoir with different strengths defined by the input weight vector V. Then, one or more readout nodes compute a weighted sum of the state of the reservoir X. The weight vector W^{out} is tuned so that the output y_t of the readout approximates a target output signal.

Ridge regression is equivalent to a linear regression but instead of fitting the regression coefficients with least squares it uses ridge regression

$$W^{out} = Y_{target} X^T (X X^T + \gamma^2 I)^{-1} \tag{3.3}$$

where X^T is the transpose of X, Y is a matrix with all expected outputs over time, I is the identity matrix and γ is a regularization parameter.

Ridge regression is a variation of the least squares method that penalizes the size of the regression coefficients. In doing so it introduces a certain bias, but on the other hand it also reduces the variance of the estimate. This allows estimating the parameters of a linear regression when the predictor variables are strongly correlated, making it a common readout choice in the context of reservoir computing (Wyffels et al., 2008).

3.2.2 NARMA task

The NARMA task has been widely used as a memory test in the context of neural networks since (Jaeger, 2002a). It consists in training a network to model the output of the 10th order Nonlinear Auto-Regressive Moving Average (NARMA) system, introduced by Atiya and Parlos (2000). This is a discrete time system where the input values $s(t)$ are drawn from the uniform distribution $\mathcal{U}(0, 0.5)$ and the output $y(t)$ is defined by

$$y(t+1) = 0.3y(t) + 0.05y(t)\sum_{i=0}^{9} y(t-i) + 1.5s(t-9)s(t) + 0.1 \quad (3.4)$$

Thus, output at time t depends on both input at time t and previous input and output history.

In our case, we simulated a network with teach one of the five topologies described above and a single input node feeding the $s(t)$ series in the system. Then, a ridge regression readout node was trained to model the dynamics of $y(t)$. For each realization a NARMA series of 10000 steps was generated, using 9000 of them for the training phase of the ridge regression, and the remaining 1000 steps to test its performance.

Figure 3.5 shows a representative time trace of the NARMA input signal, the system output, and a prediction obtained with each of the biological networks. It can be observed that the precision of the prediction improves with larger cores (core sizes are shown in Table 3.3).

The performance of the reservoirs during the NARMA task was quantitatively evaluated using the normalized root mean squared error measure, defined as

$$NRMSE = \sqrt{\frac{\langle(\tilde{y}(t) - y(t))^2\rangle_t}{\langle(y(t) - \langle y(t)\rangle_t)^2\rangle_t}} \quad (3.5)$$

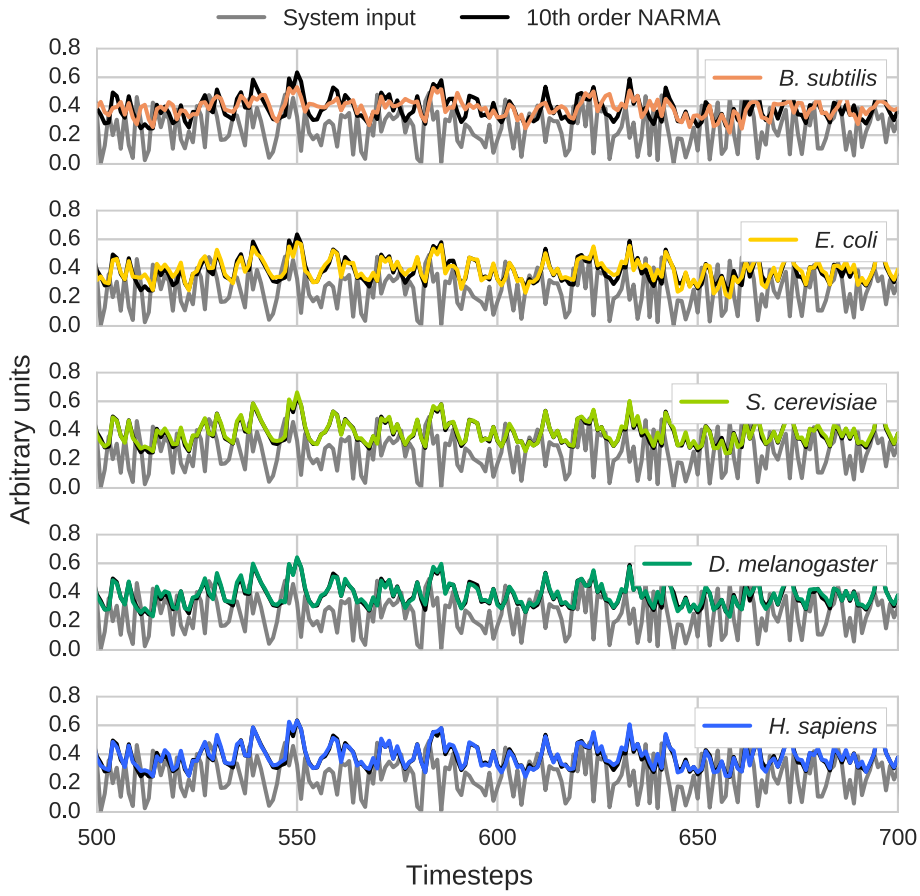

Figure 3.5: Representative time series of the test phase of a 10th order NARMA prediction task. For each biological network studied, the topology of its core was used to build a reservoir, and a readout node was trained to predict the output of the 10th order NARMA system using the state of this reservoir. The different lines correspond to the random input of the system, the actual output of the NARMA function and, for each case, the output predicted by the readout node.

where $\tilde{y}(t)$ is the output predicted by the readout, $y(t)$ is the output of the actual NARMA system and $\langle\cdot\rangle_t$ indicates the mean over time. Finally, the following topologies were used as controls:

- Echo State Network – fixed mean degree (kESN): Erdös-Rényi random network with the same mean degree ($2 \times n_{edges}/n_{nodes}$) as the original network topology.

- Echo State Network – fixed fraction of links (fESN): Erdös-Rényi random network with the same density —i.e. fraction of existing links over all possible ones— (n_{edges}/n_{nodes}^2) as the original network topology.

- Simple Cycle Reservoir (SCR): a directed circular graph, being the simplest topology to work as computational reservoir according to Rodan and Tino (2011).

Note that for control networks with the same number of nodes as the problem topology, kESN is equal to pESN. This is not the case, however, when the number of nodes changes.

With these measures and controls, we observed that the biological cores performed in the NARMA test as well as the *de facto* standard topologies (random networks) in the reservoir computing literature. Figure 3.6 shows the median NRMSE achieved by reservoirs with the topology of the biological cores and control topologies within a range of sizes. As can be observed, the biological cores always performed as well as the random fESN and kESN control networks of the same size. Results also suggest that the different performance of each GRN is related to their size. In this regard it is worth noting that despite the fact that the number of edges in the control networks scales linearly with the size for kESN and

quadratically for fESN, they show similar performance for all the range of sizes. Thus, this discards any major effect of the number of edges in the reservoir performance in these conditions, at least within those wide ranges. Besides, while all biological and control networks with sizes under 40 nodes performed similarly worse, for a larger number of nodes the recurrent but structurally constrained SCR is outperformed by the less structured fESN and kESN and by the biological cores. In fact, differences between SCR and the ESN variants increase with size within the interval analysed.

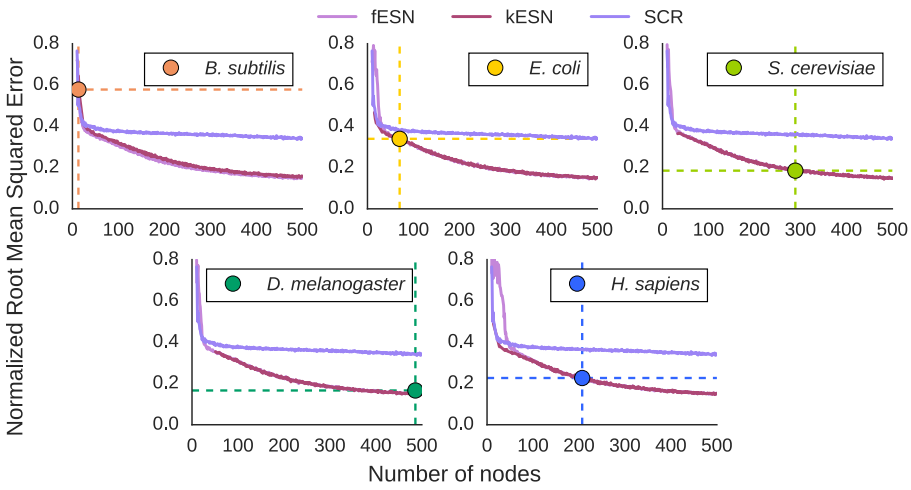

Figure 3.6: Predicting performance of reservoirs with topologies from the biological cores compared to control topologies. Reservoir performance is evaluated with the normalized root mean squared error (NRSME) of their predictions, where lower NRMSE values are better. The value represented for each biological network topology corresponds to the median NRMSE value for 10000 trials (with edge weights and data series randomization). The values plotted for each control network (fESN, kESN and SCR) correspond to the median value of 100 trials for each network size from 10 to 500 nodes. In each case, fESN and kESN are produced taking the fraction of links and the mean degree, respectively, from the biological core they are compared to.

These results quantify the performance of each network in a test that requires a certain memory. And this is the reason why the performance saturates for increasing reservoir sizes. They do not shed any light, however, on *how much* memory a network can encode in its dynamics as long as it is *enough* to fulfil the task.

3.2.3 Critical memory capacity

Aiming to quantify the amount of temporal information that a system can store, we computed the critical memory capacity of the networks. We defined the critical memory capacity as a variation of the short-term memory capacity (Jaeger, 2002b; Boedecker et al., 2012). We simulated the problem network with a single input node feeding an signal $u(t)$ drawn from a random uniform distribution $\mathcal{U}(-1, 1)$. Then, a ridge regression node was trained to obtain an output $\tilde{y}(t)$ that models a delayed version of the input signal $u(t-k)$.

From here, the k-delay memory capacity (MC_k) is defined as

$$MC_k = \frac{cov^2(u(t-k), \tilde{y}(t))}{\sigma^2(u(t-k)) \cdot \sigma^2(\tilde{y}(t))} \tag{3.6}$$

The short-term memory capacity is typically defined as $MC = \sum_{k=1}^{\infty} MC_k$, where the infinite summation is approximated to a long enough finite one (Jaeger, 2002b; Boedecker et al., 2012). The limitation, though, is that the time series needs to be orders of magnitude longer than the size of the network to ensure that MC_k tends to 0 as k increases. Otherwise, MC_k will never reach 0 and MC will never converge. Since we deal with fairly large networks, computing the short-term memory capacity with a reasonable precision was not

feasible. As an alternative measure we defined the critical memory capacity k^* as the maximum delay k that fulfils $MC_k > 0.5$.

For each gene regulatory network, we applied this test to three subnetworks: the main connected component, the reservoir or recurrent core, and the readout —i.e. all the nodes that can be reached from the reservoir. The results shown in Figure 3.7 confirm that the reservoirs are responsible for the ability of the gene regulatory networks to encode history. Accordingly, the critical memory capacity of the main connected component closely matches that of the reservoir. Contrarily, the readouts have much lower memory capacities that are mainly determined by length of the longest path: the *Bacillus subtilis* readout has 4 steps, the *Escherichia coli* has 5, the *Saccharomyces cerevisiae* has 2, the *Drosophila melanogaster* has 3 and the *Homo sapiens* has 5 (Figure 3.8). These results confirm that the recurrent cores are responsible for most of the capacity of the gene regulatory networks to dynamically store temporal information.

3.3 Encoding biologically relevant inputs

Having seen that the networks studied can encode the recent history the cell in their dynamics, we aimed to determine if these results would hold for realistic biological inputs. To test this we worked specifically with the *E. coli* regulatory network, in order to keep a balance between tractability and computational capabilities of the network. We analysed which of the 70 genes present in the *E. coli* reservoir were known to be affected by signalling pathways that react to different types of stress. Biological stresses of seven different classes were considered, namely: antibiotics, anaerobiosis, osmotic stress, oxidative stress, starvation, changes in temperature and changes in pH. Using the annotations of the Ecocyc database (Keseler

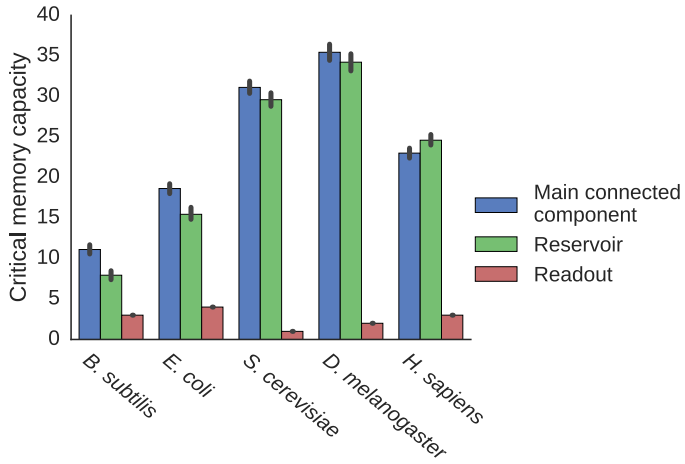

Figure 3.7: Critical memory capacity of different structures of the GRN. Critical memory capacity of the reservoir (green) and readout (red) structures and of the largest connected component, which comprises the other two, (blue) for each of the gene regulatory networks analysed. Median values are shown (n=30). The error bars indicate the 98% CI computed by bootstrapping.

et al., 2011), we manually set a confidence score to each possible stress-gene interaction. This score indicates the level of evidence supporting that such node may receive a signal in the presence of that stress, or in other words, that the product of a given gene is affected by a signalling pathway in response to a given stress. Both transcriptional and post-transcriptional regulations were considered to determine if a signalling pathway could reach a given gene as long as these regulations were not already included in the network core structure.

Using this information we subjected again the *E. coli* reservoir to a variation of the NARMA test. This time the input signal arrived solely to nodes targeted by signalling pathways that react to different stresses that *E. coli* may encounter. An input weight vector V was constructed for each stress class and confidence threshold. Given a confidence threshold,

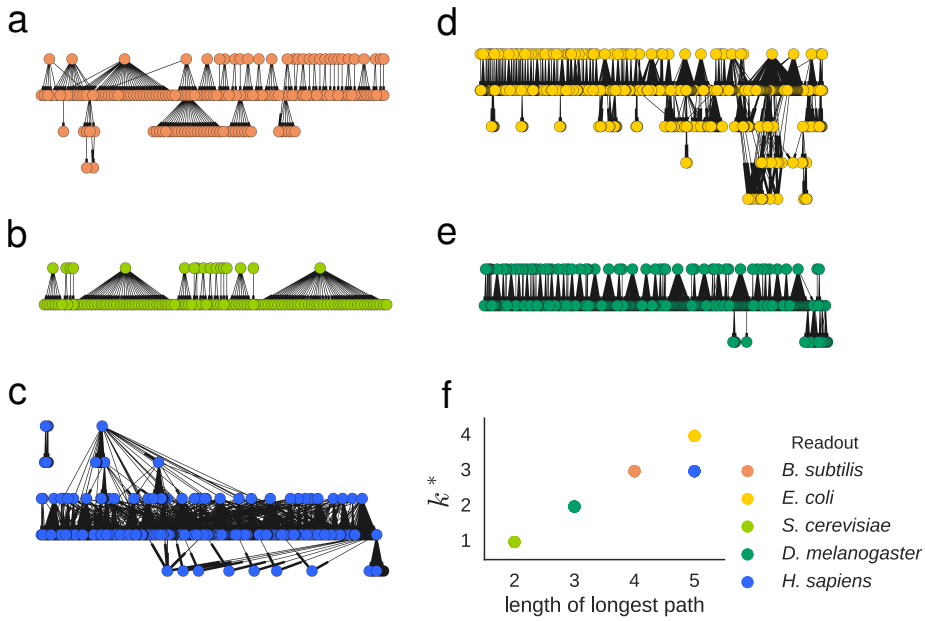

Figure 3.8: Topology and critical memory capacity of the readout structures of the gene regulatory networks. Hierarchical representation of the readout of the gene regulatory network of *Bacillus subtilis* (A), *Escherichia coli* (B), *Saccharomyces cerevisiae* (C), *Drosophila melanogaster* (D) and *Homo sapiens* (E). (F) shows the relation between the length of the longest path in each of the readouts and their critical memory capacity k^*, which measures the number of timesteps in the past that can be remembered with a certain precision in the system dynamics.

all interactions with a lower score would be set to zero, while the the remaining ones would be randomly set as before. Additionally, the sign of each entry was set to be positive (negative) if the interaction was known to produce an activation (repression) of the gene, and it was randomly set otherwise.

We found that the information arriving to the network from the different stress signalling pathways can be encoded efficiently in the reservoir

	Evidence threshold				
Stress type	**1**	**2**	**3**	**4**	**5**
Antibiotics	16	10	6	3	1
Anaerobiosis	6	6	5	3	2
Osmotic stress	13	13	7	6	3
Oxidative stress	17	17	13	9	5
Starvation	30	29	25	18	10
Changes in temperature	8	8	7	6	5
Changes in pH	15	14	10	7	1
Any stress	58	57	52	43	29

Table 3.4: Number of nodes affected by stress type for each evidence threshold.

dynamics. Table 3.4 lists the number of nodes that are considered to receive information from every stress with a confidence score equal or larger than various thresholds. Figure 3.9 shows the NRMSE obtained applying the input signal to those nodes affected by each stress type with each confidence threshold. As a control, the figure also shows the NRSME obtained applying the input signal to random sets of nodes. The most obvious conclusion from these results is that the precision of the system increases monotonically and saturates with respect to the number of nodes that receive the input signal. Besides, it is noteworthy that although the different stress signalling pathways affect different sets of nodes, their ability to introduce information in the system is comparable. This highlights the fact that memory is encoded in the network in a delocalised manner, without depending on specialized circuits or structures. This independence of the specific entry point of the information in the system confers robustness to failure of some of the nodes. While some of the

input streams may get compromised, it is unlikely that a large group of them would fail simultaneously.

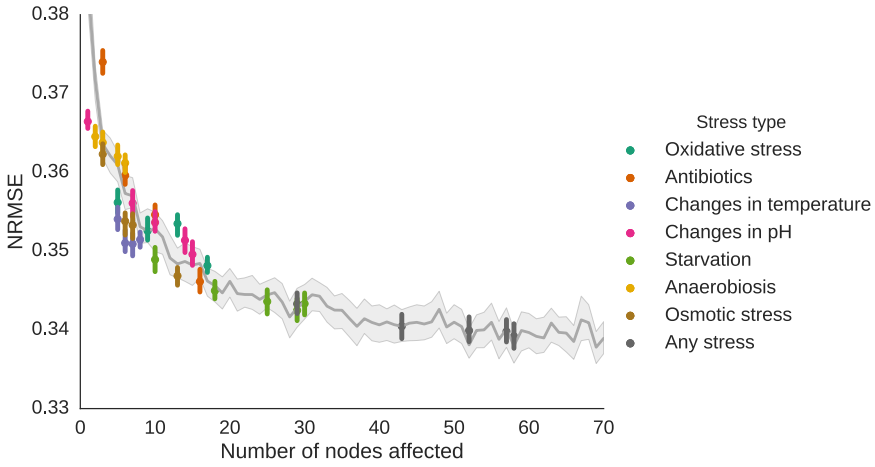

Figure 3.9: Efficiency of signalling pathways feeding an input signal to the reservoir. Median NRMSE obtained in the NARMA test using as input nodes those genes affected by signalling pathways that react to different stress types. The grey line marks the NRMSE obtained when applying the input to a random set of nodes of a given size. The median values are shown of 1000 replicates for each random input size and 2000 for each biological input set. Error bars and shaded area indicate the 98% CI computed by bootstrapping.

3.4 Evolvability of a functional readout

Finally, we tested whether biological processes can shape a readout structure that can use the temporal information encoded in the dynamics of a reservoir. Specifically, we examined if the interaction weights towards the readout can evolve under conditions where the information about recent events gives a evolutionary advantage. For that purpose, we simulated

an evolutionary process using an evolutionary algorithm. These are optimization algorithms inspired by the biological evolution. They consist in, starting from a first generation of random candidate solutions, iteratively creating a new generation by duplicating and introducing variations to the best performing solutions in the current generation.

Specifically, we used the the Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) algorithm (Hansen and Ostermeier, 2001; Gagn, 2012). The CMA-ES algorithm learns the covariance matrix of mutations in successful individuals so that beneficial mutations are sampled more often. This approach significantly reduces the computational cost while preserving the biological relevance.

The rest of the test included a setup similar to the one used in the NARMA task: a single input node feeding an $s(t)$ signal drawn from a random uniform distribution $\mathcal{U}(0, 0.5)$ into the *E. coli* reservoir. Besides, the input vector V was defined so that the signal would only reach genes known to be affected by at least one stress type with a confidence score of 3 or more.

A population of readout nodes was initiated, represented by a collection of weight vectors W^{out}. These readouts were let to evolve and compete among them so that the ones that modelled better a NARMA system had higher fitness thereby passing to the next generation with higher probability. This added a selective pressure for the individual readouts to use the temporal information encoded in the reservoir dynamics towards predicting the output of a 10th order NARMA system. Furthermore, for each generation in the evolutionary process a new realization of the $s(t)$ input signal was generated, recomputing the reservoir dynamics and the expected output. On the other hand, the internal weights of the reservoir W and the input weights V were kept constant for the whole experiment.

Figure 3.10 features a representative instance of such evolutionary processes. It shows that evolution can, indeed, tune a readout structure to read the temporal information stored in a reservoir. The performance of the evolutionary readouts improve through generations, asymptotically approaching the performance of the ridge regression node. In our simulations the performance of both types of readouts were practically indistinguishable after 2000 generations.

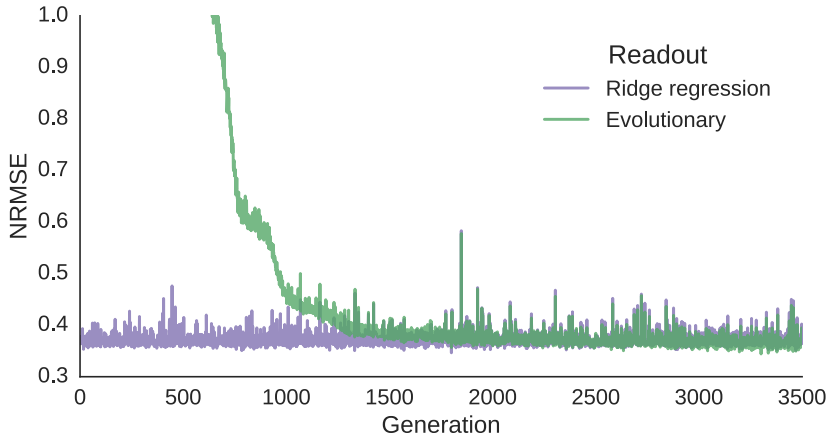

Figure 3.10: Training a readout through an evolutionary process. NRMSE during the evolutionary training process of a readout for the *E. coli* reservoir. The weights of the reservoir node were trained using the CMA-ES (Hansen and Ostermeier, 2001) algorithm to model the 10th order NARMA system. Conceptually, a population of 200 candidate solutions is let to evolve under selective advantage for those with a lower NRMSE. The green line shows the NRMSE of the centroid of the best solutions in each generation. The purple line shows the NRMSE obtained in the same situation by a readout trained using ridge regression.

3.5 Discussion

In the present chapter we propose a new paradigm to understand how cellular regulatory networks can store and process temporal information. Specifically, we suggest that these networks can function as reservoir computing systems. A division of labour allows to separate the processes of memory encoding and decision-making in two distinct structures of the network. The first structure, the reservoir, has recurrences —i.e. cyclic paths— that give it a *fading memory* property so that it can efficiently encode recent history. The latter region, the readout, has a feed-forward or acyclic structure and uses the information it gets from the reservoir to make a classification or prediction. This separation of roles allows the system to process temporal information while still being very efficient at learning new tasks (Buonomano and Maass, 2009).

The results of analysing gene regulatory networks of five distant organisms show that this is well-grounded hypothesis. In the first place, the topology of all networks matches the structural characteristics of a reservoir computing system. Furthermore, we prove that these loosely defined reservoir structures are able to encode in their dynamics an amount of temporal information that is non-trivial given their sizes. As a matter of fact, for all networks the reservoir is from one to two orders of magnitude smaller than the readout, and yet its critical memory capacity is 2 to 29 times higher. Moreover, in the case of *E. coli*, biological signals relevant for the cell such as physiological stresses arrive at the reservoir at different locations with similar performance, pointing that the ability to encode temporal information is distributed in the reservoir, which confers robustness to failure of input streams to the reservoir. Finally, we prove that evolution can produce readout structures able to decode the state of

a reservoir, provided temporal information processing gives a selective advantage.

Gene regulatory networks have been modelled as neural networks before (Alon, 2007a). However, the processing capabilities of those systems have been mostly studied with a reductionist approach, in terms of network motifs (Shen-Orr et al., 2002) and specific dynamics of small circuits (Cotterell and Sharpe, 2010). Similarly, the learning process of the network has been addressed using models of associative learning, where the direct association of two stimuli is learned (McGregor et al., 2012; Sorek et al., 2013). Here we apply a non-conventional computation framework to understand how cells react to their environment. This implies, in the first place, that the integration of information is distributed across the network in larger and more diffuse structures with well-defined functional roles. And, in the second place, that more complex associations can be learned.

A similar connection between gene regulatory networks and reservoir computing systems was hinted by Jones et al. (2007). However, that study does not provide any clear evidence to support it. Firstly, they use as putative reservoir a gene regulatory network that does not include any recurrences other than self-regulations of some nodes that are barely interconnected. But most importantly, they test their system with a task that does not require memory.

Even though it is clear that cells benefit from anticipating the environment, we know of no example yet of a cellular system that processes complex temporal information in nature. The most studied types of anticipation, involving periodic (Golden et al., 1997; Mori and Johnson, 2001) and sequential (Tagkopoulos et al., 2008; Mitchell et al., 2009) events are mechanistically fairly simple. However, *ad hoc* experiments have shown

that *Physarum polycephalum*, also known as slime mould, and *Plasmodium cudatum* are able to learn very efficiently new temporal structures. Slime mould, in particular, can anticipate a shock after experiencing a single series of three ten-minute low-temperature shocks at one-hour intervals. Moreover, if the organism experiences a new shock several hours later, it pre-emptively reacts to the two missing following shocks (Saigusa et al., 2008). Similarly, *P. cudatum* is able to learn that an electric shock follows an innocuous vibratory or luminous stimulus (Hennessey et al., 1979; Armus et al., 2006). All these results hint at capabilities to learn temporal structures larger than what can be easily explained with current models.

In our study, the dynamics of the networks have been largely simplified with a formalism used for neural networks. The real dynamics, with nonlinear interactions and different time scales for each gene, would add more complexity to the network behaviour and increase the memory of the system (Büsing et al., 2010; Dambre et al., 2012; Tanaka et al., 2016). Furthermore, the interaction of layers of regulatory networks with different time scales —e.g. transcriptional, protein protein interaction or metabolic networks— also could increase the memory capacity of the system (Dambre et al., 2012; Gallicchio and Micheli, 2016). Far from invalidating our results, those simplifications made our tests more stringent. Similarly, the NARMA task is known to be highly demanding for the systems tested, as it requires a significant level of precision in the results. It is easy to consider that life does not need to be as precise.

We propose that cells can process temporal information and anticipate their environment by using their regulatory networks as computational reservoirs. To that end, here we explored the potential of transcriptional networks to encode the recent history of cells, but other regulatory

networks such as protein-protein interaction or metabolic networks may play a similar role. The combination of the different timescales (minutes or seconds) and learning mechanisms (evolution, chromatin regulation for transcriptional reservoirs or expression regulation for post-transcriptional reservoirs) could give rise to much richer behaviours.

4 Collective oscillations in cellular communities

4.1 Cooperation versus competition

Cooperation and competition are complex social interactions that can have critical roles in biological communities. Cooperative behaviour often increases the overall fitness of the population through processes such as division of labour and production of common goods (Branda et al., 2001; Eldar, 2011; Gregor et al., 2010; Wingreen and Levin, 2006). At the same time, individuals in a community compete with each other for limited resources, such as nutrients (Hibbing et al., 2010; Oliveira et al., 2014). But in any case, any collective behaviour vastly increases the complexity of the interactions of individual organisms with their environment.

Here, in collaboration with the experimental group of Gürol Süel at University of California San Diego, we investigated bacterial biofilms (Davies, 2003; Donlan and Costerton, 2002; Vlamakis et al., 2008; Yildiz and Visick, 2009) as model systems of self-organized cellular populations prone to both cooperation and competition.

Biofilms typically form under environmental stress conditions, such as nutrient limitation (Berk et al., 2012; Costerton et al., 1999; Hall-

Stoodley et al., 2004). As these bacterial communities grow larger, the supply of nutrients to interior cells becomes limited due to an increase in nutrient consumption associated with the growth of cells in the biofilm periphery. Severe nutrient limitation for interior cells is detrimental to the colony, since the sheltered interior cells are critical to the survival of the biofilm community in the event of an external challenge. This defines a fundamental conflict between the opposing demands for biofilm growth and viability of protected (interior) cells (Figure 4.1). The identification of possible mechanisms that ensure the viability of the protected interior cells is fundamental to understanding biofilm development (Asally et al., 2012; Wilking et al., 2013).

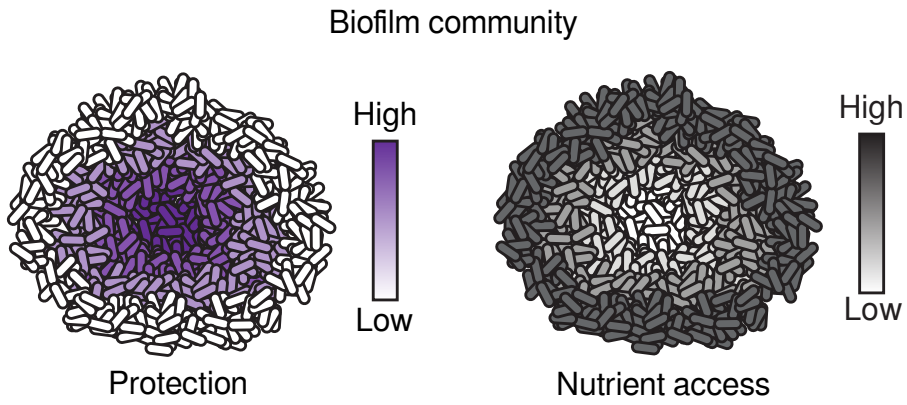

Figure 4.1: Biofilms must balance nutrient access and protection. Biofilms must reconcile opposing demands for protection from external challenges (gradient indicated in purple) and access to nutrients (gradient indicated in grey).

4.1.1 Growth oscillations

To investigate directly how *Bacillus subtilis* biofilms continue expanding while sustaining interior cells, we converted the potentially complex

three-dimensional problem to a simpler two-dimensional scenario using microfluidics. To that end, our experimental collaborators used growth chambers that are unconventionally large in the lateral x-y dimensions (3×3 mm), while confining biofilm thickness (z dimension) to only a few micrometres (Figure 4.2a). Therefore, biofilm expansion in this device is predominantly limited to two dimensions, creating a 'pancake-like' configuration. In fact, biofilms often form in confined aqueous environments and thus this microfluidic chamber may better mimic those growth conditions (Berk et al., 2012; Costerton et al., 1999; Hall-Stoodley et al., 2004). This experimental set-up is thus ideal to interrogate how biofilms can reconcile the opposing benefits of growth and protection during biofilm development.

Unexpectedly, we observed oscillations in biofilm expansion despite constant media flow within the microfluidic device (Figure 4.2b, c, Figure 4.3). Specifically, biofilms exhibit periodic reduction in colony expansion that is self-sustained and can last for more than a day (Figure 4.2d). The period of oscillations had a mean of 2.5 ± 0.8 h (standard deviation (s.d.), n = 63 colonies), which is less than the duration of the average cell replication time of 3.4 ± 0.2 h (s.d., n = 21 cell cycles) under this growth condition (Figure 4.2e). Moreover, oscillations only arise when the biofilm exceeds a certain colony size (Figure 4.4). In particular, quantitative measurements obtained from 53 individual biofilms indicate that oscillations emerge in colonies that exceed an average diameter of 580 ± 85 μm (s.d., n = 53 colonies), which corresponds to approximately one million cells (Figure 4.2f). Together, these data show that oscillations arise during biofilm formation and are self-sustained.

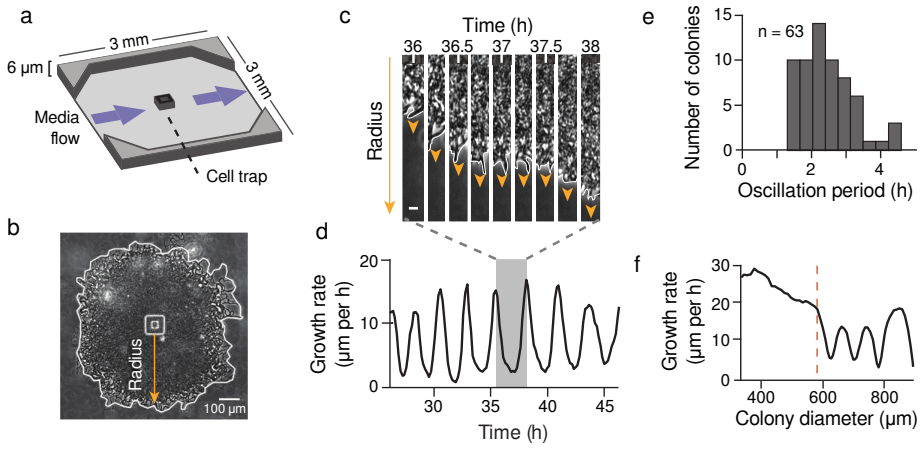

Figure 4.2: Biofilms grown in microfluidic devices show oscillations in colony expansion. a, Schematic of the microfluidic device used throughout this study. Direction of media flow is indicated by the blue arrows. b, Phase contrast image of a biofilm growing in the microfluidic device. The yellow arrow indicates the region of interest in panel c. c, Film strip of a radius of the biofilm over time shows a pause in colony expansion. This film strip represents one cycle of biofilm oscillations, indicated by the shaded region in panel d. Scale bar, 5 mm. The arrowheads indicate direction of biofilm growth. d, Growth rate over time shows persistent oscillations in colony expansion. e, Histogram of the average period of oscillations for each colony ($n = 63$ colonies, $mean = 2.5h$, $s.d. = 0.8h$). The cell replication time is approximately $3.4h$ under these conditions.f, Growth rate as a function of colony diameter (which increases in time) shows that early colony growth does not exhibit oscillations. The orange line indicates the diameter (~600 mm) at which this colony initiates oscillations. From Liu et al. (2015).

4.1.2 Metabolic regulation of biofilm growth

Given that biofilms typically form under nutrient-limited conditions and bacterial growth is generally controlled by metabolism, we hypothesized that metabolic limitation has a key role in the observed periodic halting of biofilm expansion. In particular, after determining that carbon source lim-

Figure 4.3: Stable oscillations in biofilm growth (Video). Oscillations in biofilm growth. To better visualize the growth oscillation, the periphery region the biofilm is shown. Also shown is the quantification of growth rate over time. Growth rate is defined as the speed at which the biofilm edge expands along the radial direction.
https://youtu.be/zHIjJRLGfFE

Figure 4.4: Onset of oscillations in biofilm growth (Video). Oscillations in biofilm growth. The associated time trace shows the growth rate as a function of colony diameter, where oscillatory growth only emerges when the biofilm exceeds a minimum size.
https://youtu.be/lA6zlskphmU

itation did not have an essential role in the oscillations by discarting any effect of their availability on the oscillation growth, we focused on nitrogen limitation. The standard biofilm growth media (MSgg, minimal salts glutamat glycerol) used to study *B. subtilis* biofilm development contains glutamate as the only nitrogen source (Branda et al., 2001). In most organisms including *B. subtilis*, glutamate is combined with ammonium by glutamine synthetase (GS) to produce glutamine, which is essential for biomass production and growth (Figure 4.5a) (Gunka and Commichau, 2012). Cells can obtain the necessary ammonium from glutamate through the enzymatic activity of glutamate dehydrogenase (GDH), expressed by the *rocG* and *gudB* genes in the undomesticated *B. subtilis* used in this study (Figure 4.5a) (Stannek et al., 2015; Belitsky and Sonenshein, 1998; Zeigler et al., 2008). To determine whether biofilms experience glutamine limitation, we measured expression of *nasA*, one of several genes activated in response to a lack of glutamine (Nakano et al., 1995). Results

show that biofilms indeed experience glutamine limitation during growth. Specifically, supplementation of growth media directly with glutamine reduced *nasA* promoter expression, but did not affect expression of a constitutive promoter, confirming glutamine limitation within the biofilm (Figure 4.5b). More strikingly, addition of exogenous glutamine eliminated periodic halting of biofilm growth (Figure 4.5c). These findings suggest that glutamine limitation plays a critical part in the observed oscillations during biofilm expansion.

The synthesis of glutamine requires both glutamate and ammonium; therefore, we investigated which of these substrates could be responsible for the observed glutamine limitation. Glutamate is provided in the media and is thus readily available to cells in the periphery of the biofilm. However, consumption of glutamate by peripheral cells is likely to limit its availability to cells in the biofilm interior (Figure 4.5d). One may thus expect that oscillations in biofilm expansion could be due to periodic pausing of cell growth in the biofilm interior. Accordingly, we set out to establish whether interior or peripheral cells exhibited changes in growth. By tracking physical movement within the biofilm, we uncovered that only peripheral cells grow, and that oscillations in biofilm expansion therefore arise exclusively from periodic halting of peripheral cell growth (Figure 4.5e). This finding was further confirmed by single-cell resolution analysis that directly showed periodic reduction in the growth of peripheral cells. This surprising pausing of cell growth in the periphery, despite unrestricted access to glutamate, suggests that glutamate cannot be the limiting substrate for glutamine synthesis. Consistent with this expectation, biofilm oscillations were not quenched by supplementation of the media with glutamate (Figure 4.5f). In contrast, ammonium addition does suppress the oscillations (Figure 4.5g) Therefore, it is not glutamate

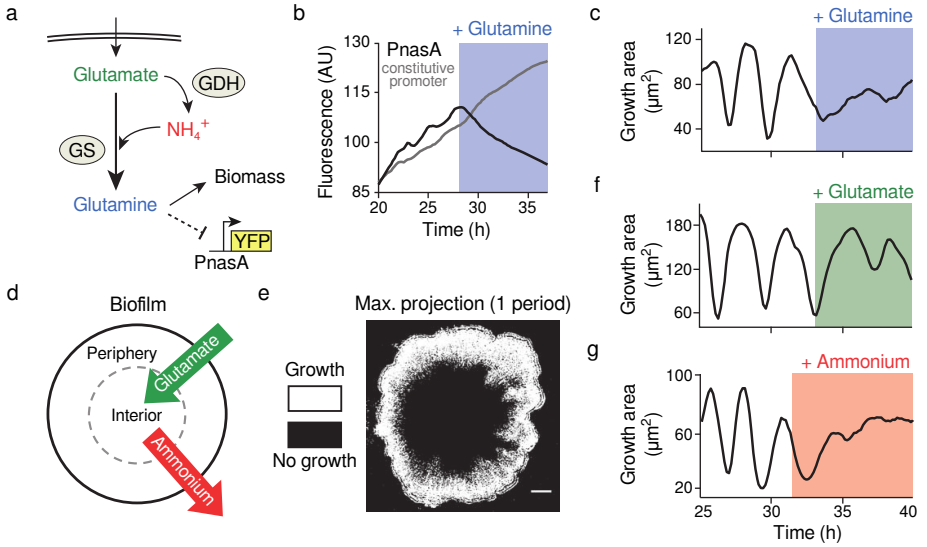

Figure 4.5: Biofilm growth depends specifically on extracellular ammonium availability. a, Colony growth in MSgg medium depends on the production of glutamine from externally supplied glutamate and self-produced or scavenged ammonium. Glutamine limitation was monitored using yellow fluorescent protein (YFP) expressed from the *nasA* promoter, which is activated upon glutamine limitation (Nakano et al., 1995). b, Addition of 1 mM glutamine (blue shading) represses expression from the P*nasA*-YFP reporter (black), but does not affect expression from a constitutive reporter (P*hyperspank*-CFP + 1 mM IPTG, grey). c, Growth area before and after addition of 1 mM glutamine to an oscillating colony. d, Of the two nutrients required for glutamine production, externally supplied glutamate (green) is most abundant in the biofilm periphery, while biofilm-produced ammonium (red) is most abundant in the biofilm interior. e, Maximum intensity projection over one period of a colony oscillation, made from a difference movie (Liu et al., 2015), which shows regions of growth (white) and no growth (black). Scale bar, 100 μm. f, Growth area of an oscillating colony before and after addition of 30 mM glutamate (green shading). g, Growth area of an oscillating colony before and after addition of 1 mM ammonium (red shading). From Liu et al. (2015).

but ammonium that appears to be the limiting substrate for glutamine synthesis in the biofilm periphery.

Because cells can self-produce ammonium from glutamate, we next sought to determine how peripheral cells could experience periodic ammonium limitation despite a constant supply of glutamate in the media. It is well known that ammonium production is a highly regulated process that is dependent on the metabolic state of the cell and the ambient level of ammonium in the environment (Kleiner, 1985). In particular, since ammonium is in equilibrium with ammonia vapour, which can freely cross the cell membrane and be lost to the extracellular media (Castorph and Kleiner, 1984), the production of ammonium is known as a 'futile cycle'. Cells therefore preferentially use extracellular (ambient) ammonium for growth, rather than producing their own (Boogerd et al., 2011; Jayakumar et al., 1986; Kim et al., 2012). Since peripheral cells are exposed to media flow, they are particularly susceptible to this futile cycle of ammonia loss. In this sense, as ammonium is not provided in the media, even if all cells produce ammonium, the biofilm interior will be the major source for ambient ammonium (Figure 4.5d). Consequently, the simplifying hypothesis is that growth of peripheral cells relies on ammonium produced within the biofilm. To test this conjecture, we supplemented the media with 1 mM ammonium, which eliminated the periodic halting in biofilm expansion (Figure 4.5g). When additional ammonium was suddenly removed from the media, growth in the biofilm periphery halted, as expected. These findings indicate that peripheral cells preferentially rely on extracellular ammonium produced within the biofilm for their growth.

4.2 *Mathematical model of metabolic codependence*

The results described above evoke the possibility that ammonium limitation for peripheral cells may arise due to glutamate limitation for interior cells. Specifically, persistent consumption of glutamate by peripheral cells can deprive the interior cells of the necessary glutamate for ammonium production. To explore this nontrivial hypothesis, we turned to mathematical modelling to develop a conceptual framework and generate experimentally testable predictions.

Our model describes separately the metabolic dynamics of interior and peripheral cells and the metabolite exchange between them, where the distinction of the two subpopulations depends on nutrient availability. The two populations are assumed to be located in a frame of reference that moves as the biofilm grows, so that they are always placed at a constant distance from the physical edge of the biofilm (Figure 4.6).

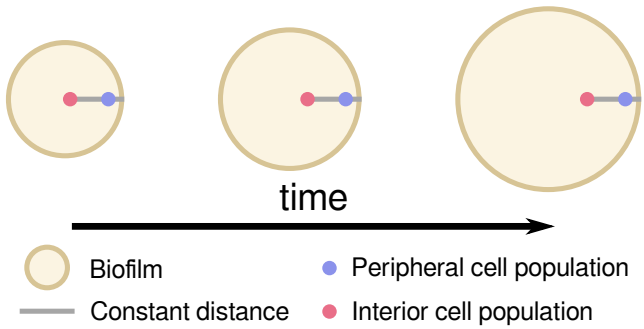

Figure 4.6: The model describes the dynamics of two cell populations in a biofilm, interior and peripheral. As the biofilm grows, there is a constant distance between the interior population and the biofilm edge.

The model considers two main assumptions regarding the interplay between the two regions (Figure 4.7):

1. Consumption of glutamate during growth of peripheral cells deprives interior cells of this nutrient and thus inhibits ammonium production in the biofilm interior. Specifically, activation of GDH is strongly reduced when the concentration of available glutamate is below a given threshold. This can be due to explicit regulatory interactions or simply as a consequence of the slowdown of cellular processes in the absence of nutrients.

2. The growth of peripheral cells depends predominantly on ammonium that is produced by metabolically stressed interior cells. As a simplification, we assume only the interior cells have active GDH.

Two additional assumptions are made regarding the effect of starvation in the cell activity:

3. Consumption of ammonium and glutamate depends on the metabolic activity of the cell. The higher the concentration of housekeeping proteins — a proxy for the metabolic state of the cell — the faster the consumption of nutrients.

4. The production of housekeeping proteins — and thus the metabolic activity of the cell — increases with the concentrations of available glutamate and ammonium.

With these postulates, the metabolic state of the biofilm is described by the following variables: 1) the concentrations of glutamate G_i in the interior and G_p in the periphery of the biofilm; 2) the concentration A of ammonium in the biofilm, which is assumed to be equal for the two populations due to its fast diffusion; 3) the concentration H_i of active glutamate

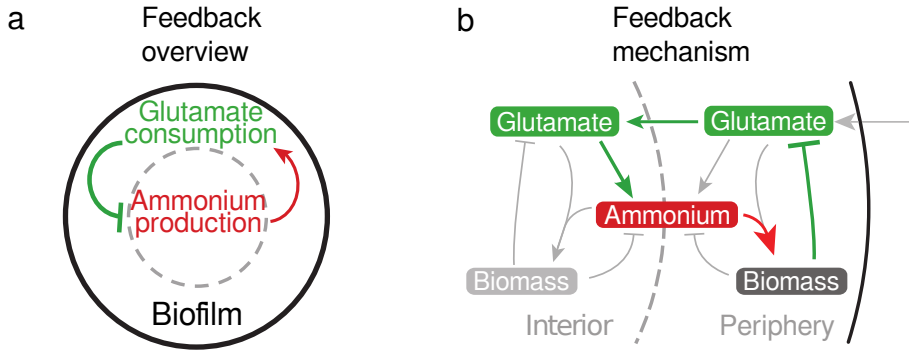

Figure 4.7: A metabolic negative feedback loop as the cause of the growth oscillations. The production of ammonium in the interior is limited by and at the same time triggers the consumption of glutamate in the periphery (green and red arrows, respectively), producing a delayed negative feedback loop. b, The excess glutamate not consumed by the biofilm periphery diffuses to the interior, where it can be converted into ammonium (green arrows). The ammonium in turn enhances growth in the periphery (red arrow) and consequently reduces the supply of glutamate to the interior.

dehydrogenase (GDH) in the interior cells; and 4) the concentrations of housekeeping proteins (such as ribosomal proteins) r_i in the interior and r_p in the periphery, which is assumed to determine the rate of biomass production in the areas of the biofilm.

The dynamics of these state variables are described by the following set of ordinary differential equations:

$$\frac{dA}{dt} = \alpha G_i H_i - \delta_A A(r_i + r_p) \tag{4.1}$$

$$\frac{dG_i}{dt} = D(G_p - G_i) - \alpha G_i H_i - \delta_G G_i r_i \tag{4.2}$$

$$\frac{dG_i}{dt} = D(G_i - G_p) + D_E(G_E - G_p) - \delta_G G_p r_p \tag{4.3}$$

$$\frac{dH_i}{dt} = \beta_H \frac{G_i^n}{K_H^n + G_i^n} - \gamma_H H_i \tag{4.4}$$

$$\frac{dr_i}{dt} = \beta_r A G_i - \gamma_r r_i \tag{4.5}$$

$$\frac{dr_p}{dt} = \beta_r A G_p - \gamma_r r_p \tag{4.6}$$

Where the terms in the equations are interpreted as follows:

$\alpha G_i H_i$: ammonium production from glutamate, catalysed by the enzyme GDH (Figure 4.5a)

$\delta_A A(r_i + r_p)$: ammonium consumption by interior and peripheral cells

$\delta_G G_i r_i$ and $\delta_G G_p r_p$: glutamate consumption by interior and peripheral cells, respectively

$D(G_p - G_i)$: glutamate diffusion between peripheral and interior regions

$D_E(G_E - G_p)$: glutamate diffusion between the environment and the periphery of the biofilm

$\beta_H \frac{G_i^n}{K_H^n + G_i^n}$: GDH activation in the interior cells

$\gamma_H H_i$: GDH deactivation in the interior cells

$\beta_r A G_i$ and $\beta_r A G_p$: production of housekeeping proteins in the interior and peripheral cells, respectively

$\gamma_r r_i$ and $\gamma_r r_p$: degradation of housekeeping proteins in interior and peripheral cells, respectively

In order to extract from the model the population expansion, which is the experimental observable, we consider that the dynamics of the cell

density ρ of the two populations are given by:

$$\frac{d\rho_{(i,p)}}{dt} = \eta r_{(i,p)}\rho_{(i,p)}\left(1 - \frac{\rho_{(i,p)}}{K(G_{(i,p)})}\right) - \lambda_{(i,p)}\rho_{(i,p)} \tag{4.7}$$

The first term in the right-hand side is a logistic-growth term, where the maximal growth rate is considered to be proportional to the concentrations of housekeeping proteins r_i and r_p. Additionally, we consider that the carrying capacity K depends on the concentration of glutamate:

$$K(G) = G^m/(K_k^m + G^m) \tag{4.8}$$

Thus $K(G)$ varies between 0 and 1 depending on whether glutamate concentration is below or above a given threshold, denoted as K_k. Note that the cell density $\rho_{(i,p)}$ defined here is relative to the carrying capacity, therefore, both K and $\rho_{(i,p)}$ are dimensionless.

The logistic-growth term in the density equation shown above describes the standard birth/death processes that occur in an unmoving bacterial population. In our system, however, the peripheral cells are always expanding into the open area outside of the biofilm. Therefore, a decay term has to be included to account for this fact, which is the motivation behind the second term, $-\lambda_{(i,p)}\rho_{(i,p)}$, on the right-hand side of the density equation. Thus we assume the parameter $\lambda_{(i,p)}$ to be non-zero only for the peripheral population. However, in Section 4.4.1 we use non-zero λ_i for the interior population in the case of chemical attack, where the peripheral population is eradicated and consequently the interior cells can expand.

Considering the above described dynamics for the cell densities, the population expansion rate is given by the logistic growth term, since this

is the only term related with actual growth of the population:

$$\mu_{(i,p)} = \eta r_{(i,p)} \rho_{(i,p)} \left(1 - \frac{\rho_{(i,p)}}{K(G_{(i,p)})} \right) \tag{4.9}$$

The dynamics of this model reproduce the periodic halting of the growth observed in the experiments, as shown in Figure 4.8. The value of each parameter of the model is detailed in Table 4.1. Furthermore, the parameter sensitivity analysis in Figure 4.9 shows that the period and amplitude of the oscillations are reasonably robust to parameter changes.

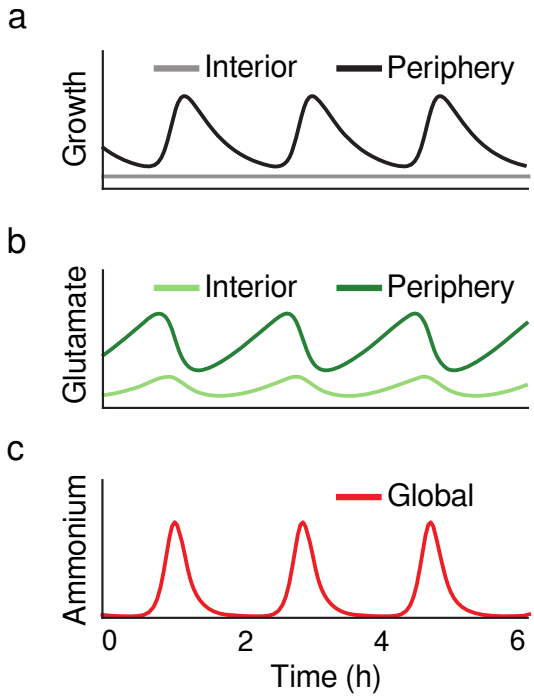

Figure 4.8: Metabolic oscillations produced by the matematical model. a, Biofilm growth over time. b, Glutamate concentration over time. c, Ammonium concentration over time.

Table 4.1: Parameters of the mathematical model of the biofilm growth oscillations.

	Description	Value	Units
α	Glutamate dehydrogenation coefficient	50	$\mu M^{-1}h^{-1}$
δ_A	Ammonium consumption coefficient	4	$\mu M^{-1}h^{-1}$
δ_G	Glutamate consumption rate	4	$\mu M^{-1}h^{-1}$
D	Glutamate diffusion constant within the biofilm.	0.4	h^{-1}
D_E	Glutamate diffusion constant between biofilm and exterior	0.6	h^{-1}
G_E	Glutamate concentration in the external medium	30	mM
β_H	Maximal activation rate of GDH	50	$\mu M^{-1}h^{-1}$
γ_H	Deactivation rate of GDH	7.5	h^{-1}
K_H	GDH activation threshold	7.2	mM
n	Hill coefficient for GDH activation	7	
β_r	Expression coefficient of ribosomal/housekeeping proteins	0.14	$mM^{-1}h^{-1}$
γ_r	Degradation rate of ribosomal/housekeeping proteins	2	h^{-1}
η	Population growth rate coefficient	100	mM^{-1}
K_K	Glutamate threshold for carrying capacity	0.85	mM
m	Hill coefficient for carrying capacity	12	

λ_i	Expansion rate of interior cells	0	h^{-1}
λ_p	Expansion rate of peripheral cells	0.032	h^{-1}
$[Gln]$	Concentration of glutamine in the medium	1	mM
K_α	Glutamine inhibition threshold on GDH activity	$5 \cdot 10^{-8}$	mM
K_δ	Glutamine inhibition threshold on GS activity	$5 \cdot 10^{-2}$	mM
α_0	Rate of ammonium entering the biofilm from the external medium	0.03	mMh^{-1}
β_0	Expression rate of GDH from the additional copy of the gene	$1.5 \cdot 10^{-6}$	mMh^{-1}
$\lambda_{H_2O_2}$	Death rate due to hydrogen peroxide	5	h^{-1}

4.3 Model perturbations

4.3.1 Addition of glutamine

The first experimental perturbation used to validate the model is the addition of Glutamine in the medium of an oscillating biofilm (Figure 4.5c). Glutamine is synthesized by glutamine synthetase (GS) in the cell, and it also regulates the activity of GS through a negative feedback (Kim et al., 2012). Therefore, external addition of glutamine reduces GS activity, and consequently lowers its consumption of ammonium and glutamate (used to synthesize glutamine). Additionally, we assumed that glutamine inhibits either directly or indirectly GDH activity, affecting the production

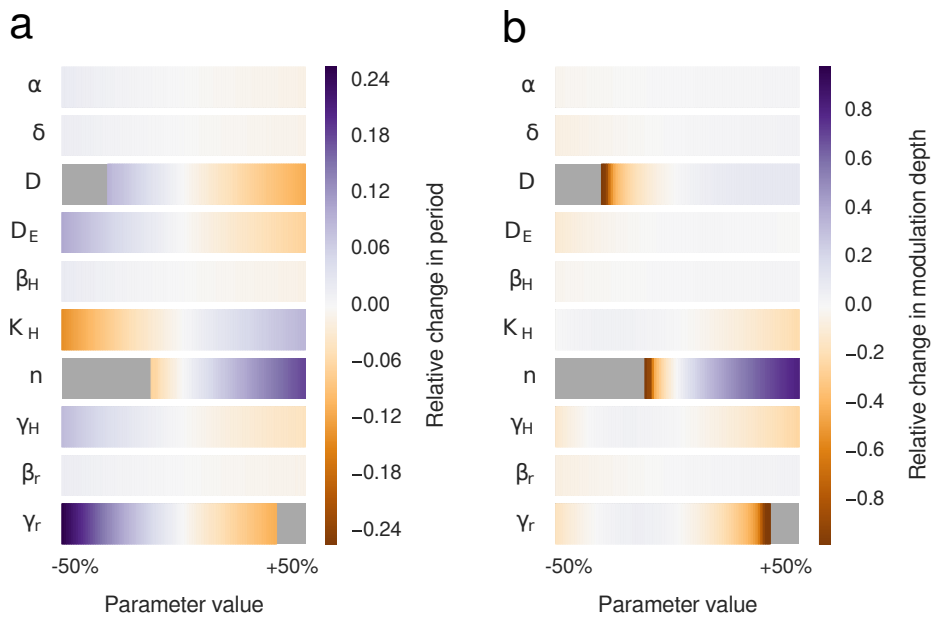

Figure 4.9: Systematic sensitivity analysis of the parameters of the model. Sensitivity analysis of oscillation period (a) and modulation depth (b) to changes in model parameters. Modulation depth is defined as the amplitude of the oscillations divided by the mean value. Grey colour denotes parameter regions where the system does not oscillate. In all the cases shown in this figure the transition towards non-oscillating states occurs trhough a supercritical Andronov-Hopf bifurcation.

of ammonium from glutamate. This is implemented in the model as non-competitive inhibition on the parameters α and δ. Specifically, the effective $\bar{\alpha}$ and $\bar{\delta}$ are given by:

$$\bar{\alpha} = \frac{\alpha}{\frac{[Gln]}{K_\alpha} + 1} \tag{4.10}$$

$$\bar{\delta}_{(A,G)} = \frac{\delta_{(A,G)}}{\frac{[Gln]}{K_\delta} + 1} \tag{4.11}$$

Figure 4.10a shows the model prediction: in agreement with the experimental observations, external addition of glutamine leads to the quenching of oscillation. A systematic analysis of the effect of glutamine addition to the dynamics of the system is shown in Figure 4.10b and c. The bifurcation diagram shows that as the external glutamine concentration increases the amplitude of the stable limit cycle shrinks until it collapses into the unstable fixed point giving birth to a stable branch. Specifically, the system undergoes a supercritical Andronov-Hopf bifurcation at external glutamine concentration $[Gln] = 3.20 \cdot 10^{-6}$ mM.

4.3.2 Addition of glutamate

Another experimental perturbation that the model needs to reproduce is the two-fold increase of the concentration of glutamate in the medium (Figure 4.5f). The concentration of glutamate in the external medium is explicitly defined in the model by the parameter G_E. Thus, supplementation of additional glutamate is straightforward in the model — it increases the value of G_E. Figure 4.11a shows the model prediction: consistent with the experimental observations, a moderate increase in external glutamate does not eliminate the oscillations. A systematic study also shows that further increasing glutamate can reduce the period of oscillation by more than two-fold, and sufficiently high concentrations leads to quenching of oscillations (Figure 4.11b and c). The bifurcation diagram shows that the system undergoes two Andronov-Hopf bifurcations at external concentrations of glutamate $G_E = 1.92 \cdot 10^1$ mM and $G_E = 7.31 \cdot 10^2$ mM, respectively.

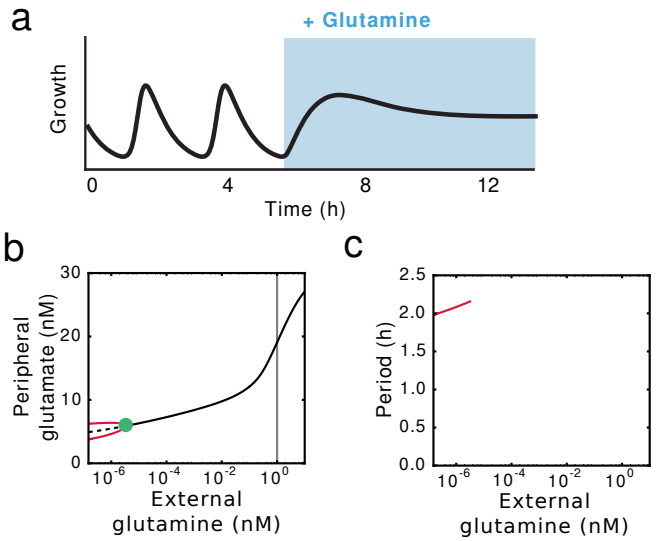

Figure 4.10: Modelling results for glutamine supplementation. a, Colony growth before and after glutamine addition (indicated by blue shading). b, Bifurcation diagrams showing systematic analysis on the effects of external glutamine. The solid black line denotes a stable fixed point. The dashed black line corresponds to an unstable fixed point. The red lines correspond to the extrema of oscillations in peripheral glutamate (stable limit cycle). The green dot marks the point where the system undergoes a supercritical Andronov-Hopf bifurcation. The vertical grey line highlights the state of the system in panel a. c, Effect of external glutamine on the period of the oscillations.

4.3.3 Addition of ammonium

Finally, the model also had to reproduce the effect of adding ammonium in the medium observed in the experiments (Figure 4.5g). The concentration of ammonium is explicitly represented in the model with the variable A. Thus, an addition of ammonium to the media can be represented as an additional creation term (α_0) so that equation 4.1 becomes:

$$\frac{dA}{dt} = \alpha G_i H_i - \delta_A A(r_i + r_p) + \boldsymbol{\alpha_0} \tag{4.12}$$

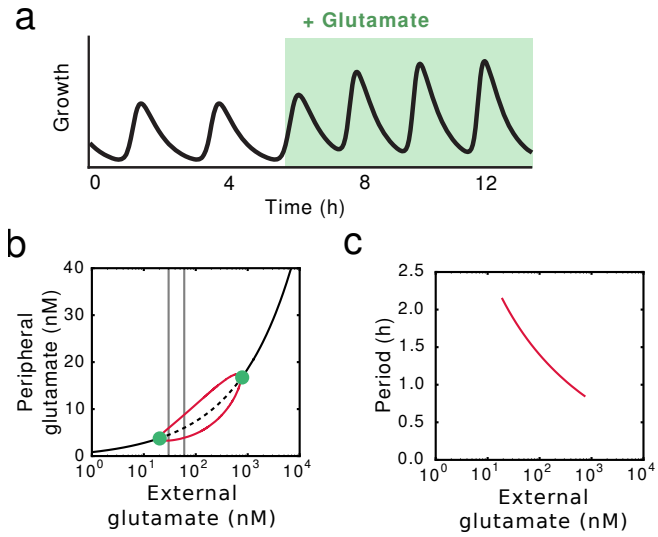

Figure 4.11: Modelling results for glutamate supplementation. a, Colony growth before and after glutamate addition (indicated by green shading). b, Bifurcation diagrams showing systematic analysis on the effects of external glutamate. The solid black line denotes stable fixed point. The dashed black line corresponds to an unstable fixed point. The red lines correspond to the extrema of oscillations in peripheral glutamate (stable limit cycle). The green dots mark the point where the system undergoes supercritical Andronov-Hopf bifurcations. The vertical grey lines highlight the states of the system in panel a. c, Effect of external glutamate on the period of the oscillations.

Figure 4.12a shows the model prediction: in agreement with the experiments, externally adding ammonium quenches oscillation. We also systematically explored the effect of different ammonium concentrations through a bifurcation diagram of the system with respect to α_0 (Figure 4.12b and c). As in the case of glutamine, the bifurcation diagram shows that increasing ammonium concentrations in the medium shrink the amplitude of the stable limit cycle until it collapses into the unstable fixed point, giving birth to a stable branch. Again, the system un-

dergoes a supercritical Andronov-Hopf bifurcation, this time at $\alpha_0 = 6.30 \cdot 10^{-4}\,\mathrm{mM\,h^{-1}}$.

In contrast to the effect of increasing glutamate, the model predicts that the period of oscillations does not change substantially as ammonium varies.

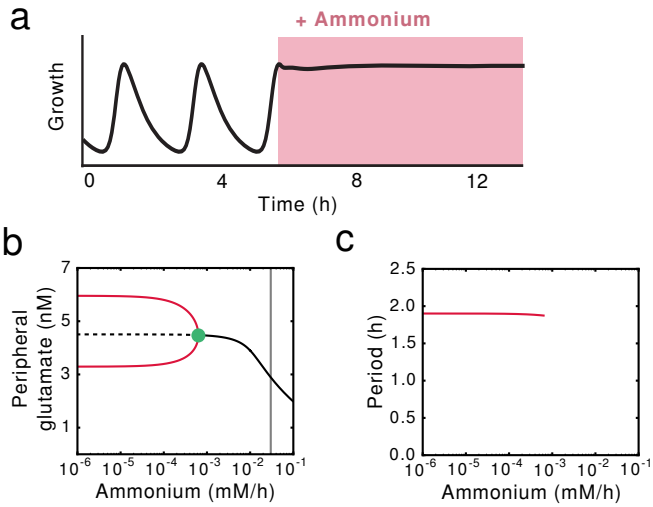

Figure 4.12: Modelling results for ammonium supplementation. a, Colony growth before and after ammonium addition (indicated by red shading). b, Bifurcation diagrams showing systematic analysis on the effects of external ammonium. The solid black line denotes stable fixed point. The dashed black line corresponds to an unstable fixed point. The red lines correspond to the extrema of oscillations in peripheral glutamate (stable limit cycle). The green dot marks the point where the system undergoes a supercritical Andronov-Hopf bifurcation. The vertical grey line highlights the state of the system in panel a. c, systematic analysis on the effects of external ammonium on the period of the oscillations.

4.4 Model predictions

A model based on the simplifying assumptions described at the beginning of the previous section (Figure 4.7) generates oscillations consistent with our experimental observations (Figure 4.8) and reproduces the effects of supplementing the media with glutamine (Figure 4.10), glutamate (Figure 4.11) and ammonium (Figure 4.12). Therefore, this simple model shows that periodic halting in biofilm growth can result from metabolic co-dependence between cells in the biofilm periphery and interior that is driven by glutamate consumption and ammonium production, respectively. We now turn to specific predictions of the model that shed light on the mechanism of the observed oscillations.

4.4.1 Growth in the interior after an external attack

Bacteria growing inside biofilms have an increased resistance to antimicrobial agents (Høiby et al., 2010). It has been shown that this can be explained in part by a limited diffusion of these compounds within the biofilm (Mah and O'Toole, 2001; Stewart, 2002). We resolved to test if, considering such external attack, our model can produce any testable prediction.

To do so, we use the model to explore the predicted outcome of adding an antimicrobial agent such as hydrogen peroxide (H_2O_2) to the medium. Hydrogen peroxide is a strong oxidizer that can kill the cells on the periphery of the biofilm. Dead cells in the biofilm will still affect glutamate diffusion but will be metabolically inactive. Thus, the killing is implemented in the model by removing the production term of housekeeping proteins in the peripheral cell population. Additionally, a new negative term in the cellular density equation is introduced to account for cell

death. To that end, the differential equations for r_p and ρ_p are modified as shown below:

$$\frac{dr_p}{dt} = -\gamma_r r_p \tag{4.13}$$

$$\frac{d\rho_p}{dt} = \eta r_p \rho_p \left(1 - \frac{\rho_p}{K(G_p)}\right) - \boldsymbol{\lambda_{H_2O_2}\rho_p} - \lambda_p \rho_p \tag{4.14}$$

This new term is also added to the equation for the rate population expansion:

$$\mu_p = \eta r_p \rho_p \left(1 - \frac{\rho_p}{K(G_p)}\right) - \boldsymbol{\lambda_{H_2O_2}\rho_p} \tag{4.15}$$

The metabolic co-dependence between interior and peripheral cells gives rise to the surprising prediction that external attack could promote growth within the biofilm. Specifically, killing of peripheral cells will eliminate their glutamate consumption, which will increase glutamate availability in the biofilm and thereby promote growth of interior cells (Figure 4.13a). To test this hypothesis, we measured cell death and growth within oscillating biofilms (Figure 4.13c and d, top). When we exposed the biofilm to media containing hydrogen peroxide (H_2O_2), we observed increased cell death predominantly in the biofilm periphery (Figure 4.13c, bottom). As predicted, death of peripheral cells led to growth of interior cells (Figure 4.13d, bottom). Figure 4.13b shows the model prediction and the experimental results on the average growth rate and death in interior and peripheral populations after the addition of hydrogen peroxide, for both wild type and GDH overexpressing biofilms. To verify that this response is not uniquely triggered by H_2O_2, we exposed biofilms to the antibiotic chloramphenicol and again observed growth of interior cells (Liu et al., 2015). These findings further support our hypothesis that

glutamate consumption by peripheral cells limits its availability in the biofilm.

4.4.2 GDH overexpression

Our model also assumes that glutamate starvation of the biofilm interior reduces the production of ammonium that can support peripheral cell growth. This assumption provokes the question as to why peripheral cells do not simply overcome their dependence on extracellular ammonium by increasing intracellular production (Commichau et al., 2008; Detsch and Stülke, 2003). To address this question, we considered an inducible copy of the GDH gene *rocG* as a testable way to decouple the production of ammonium from the state of the biofilm interior (Figure 4.14a).

In the model, the overexpression of GDH is implemented by an additional creation term β_0 in the equation governing its dynamics (H_i). Furthermore, since the overexpression is applied throughout the entire biofilm, we now also need to include active GDH in the peripheral cells (H_p), and consequently the production of ammonium from those cells. To that end, the differential equations for A, G_p and H_i are modified as shown below, and an equation for GDH in the peripheral cell population (H_p) is also added:

$$\frac{dA}{dt} = \alpha G_i H_i + \boldsymbol{\alpha G_p H_p} - \delta_A A(r_i + r_p) \tag{4.16}$$

$$\frac{dG_p}{dt} = D(G_i - G_p) + D_E(G_E - G_p) \boldsymbol{- \alpha G_p H_p} - \delta_G G_p r_p \tag{4.17}$$

$$\frac{dH_i}{dt} = \boldsymbol{\beta_0} + \beta_H \frac{G_i^n}{K_H^n + G_i^n} - \gamma_H H_i \tag{4.18}$$

$$\boldsymbol{\frac{dH_p}{dt} = \beta_0 - \gamma_H H_p} \tag{4.19}$$

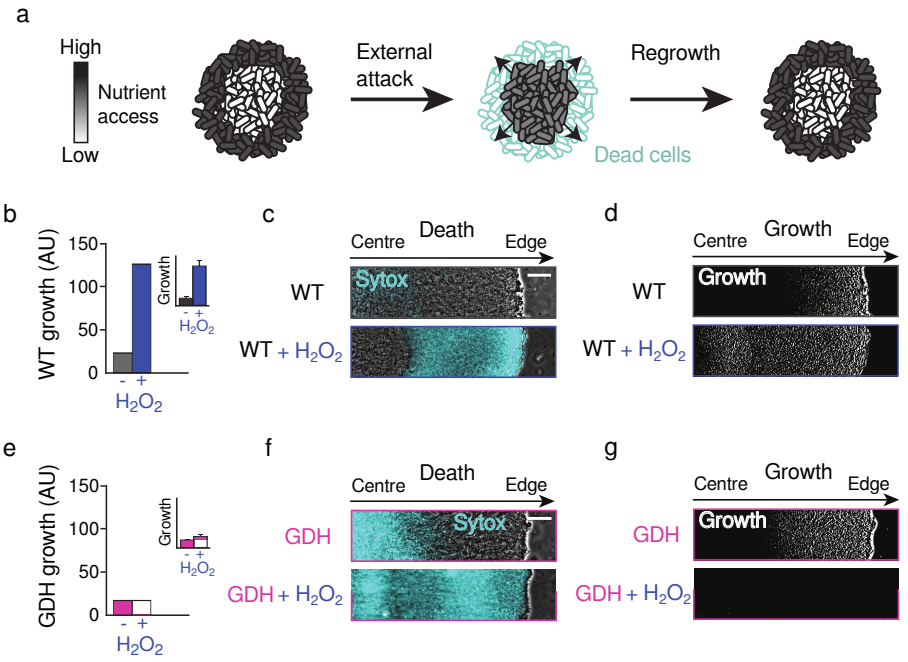

Figure 4.13: External and internal attacks on biofilms. a, Visual representation of the predicted outcome of an external attack on biofilm growth. b, Model prediction of total biofilm growth rate in wild-type strain upon challenge with H_2O_2. Experimental verification is shown as an inset, where error bars represent standard deviations ($n = 4$ colonies). c, Phase contrast merged with cell death marker (cyan, 1 μM Sytox green) images of a wild-type (WT) biofilm region show cell death with and without challenge by 2% (v/v) H_2O_2. Scale bar, 50 μm. d, In the same biofilm, difference images (white regions indicate cell growth) show wild-type growth with and without challenge by H_2O_2. e, Model prediction of total biofilm growth rate in GDH overexpression strain upon challenge with H_2O_2. Experimental verification is shown as an inset, where error bars represent standard deviations ($n = 3$ colonies). f, Phase contrast merged with cell death marker (cyan, 1 μM Sytox green) images of a colony overexpressing GDH with and without challenge by H_2O_2. g, In the same biofilm, difference images show cell growth during GDH overexpression alone, and with challenge by H_2O_2. Adapted from Liu et al. (2015).

Figure 4.14b shows the model prediction: overexpressing GDH leads to quenching of oscillation. A systematic analysis on different levels of overexpression is shown in the bifurcation diagram in the Figure 4.14d and e. Similarly than with the addition of external glutamine and ammonium, the overexpression of GDH takes the system through a supercritical Andronov-Hopf bifurcation. The stable limit cycle collapses with the unstable fixed branch and generates a stable fixed branch at $\beta_0 = 1.55 \cdot 10^{-8}\,\mathrm{mM\,h^{-1}}$. Additionally, the modelling results predict that continuous growth of the cells in the periphery deprives the interior population from nutrients, which causes and increased amount of cell death (Figure 4.13e, inset).

To experimentally validate this hypothesis, we constructed a strain that contains an inducible copy of the GDH gene *rocG* (Figure 4.14a). We confirmed that GDH overexpression was not toxic to individual cells and did not affect their growth rate (Liu et al., 2015). In contrast, the induction of GDH expression in the biofilm quenched growth oscillations (Figure 4.14b and c) and resulted in high levels of cell death in the colony interior (Figure 4.13f, top). This result explains why peripheral cells do not appear to utilize the simple strategy of overcoming their dependence on extracellular ammonium: such a strategy would result in the continuous growth of peripheral cells, starving and ultimately causing the death of sheltered interior cells within the biofilm. Periodic halting of peripheral cell growth due to extracellular ammonium limitation thus promotes the overall viability of the biofilm

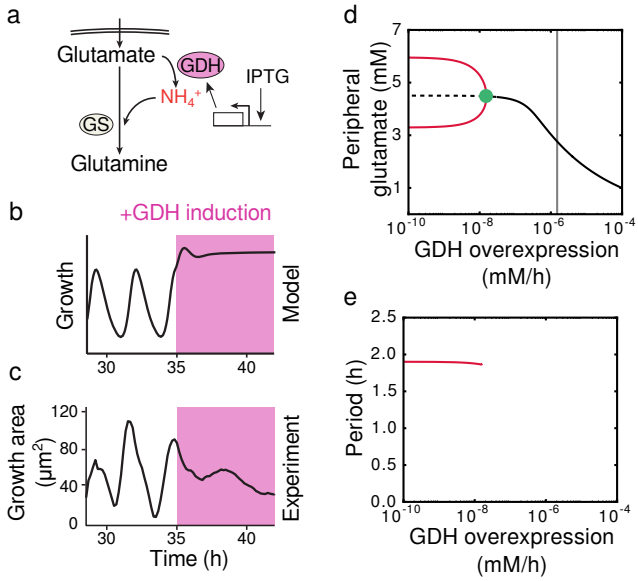

Figure 4.14: Glutamate dehydrogenase overexpression breaks the metabolic coupling. a, Overexpression of glutamate dehydrogenase (GDH, pink) promotes more production of ammonium from glutamate. b, modelling and c, experimental results of GDH overexpression (induced with 1 mM IPTG, indicated by pink shading). d, Bifurcation diagram showing systematic analysis on the effects of induction of GDH overexpression. The solid black line denotes stable fixed point. The dashed black line corresponds to an unstable fixed point. The red lines correspond to the extrema of oscillations in peripheral glutamate (stable limit cycle). The green dot marks the point where the system undergoes a supercritical Andronov-Hopf bifurcation. The vertical grey line highlights the state of the system in panel b. e, Effect of induction of GDH overexpression on the period of the oscillations.

4.4.3 Biofilm elimination: divide et impera

The ability of the biofilm to regenerate itself in the event of an external attack suggested that killing the biofilm interior first would be a more effective strategy for biofilm elimination.

From a mathematical point of view, this double perturbation was modelled by combining the modifications of the equations enumerated in the two previous sections. The only additional change is that when H_2O_2 is applied to a biofilm, if present, the production term for GDH β_0 in the periphery population is removed so that equation for H_p becomes:

$$\frac{dH_p}{dt} = -\gamma_H H_p \tag{4.20}$$

Accordingly, we exposed the GDH overexpression strain to hydrogen peroxide and again measured growth and death. As described above, GDH induction causes death of interior cells. Exposing the GDH overexpression strain to hydrogen peroxide resulted in more effective global killing throughout the biofilm (Figure 4.13f, g, bottom). While in the wild-type biofilm, interior cells begin to grow in response to an external attack, metabolic independence between interior and peripheral cells in the GDH strain interferes with this defence mechanism (Figure 4.13e, inset). This outcome is also consistent with modelling predictions (Figure 4.13e). Oscillations in biofilm growth that are driven by metabolic co-dependence thus promote the resilience of the biofilm community by sustaining the viability of the sheltered interior cells that are most likely to survive in the event of an environmental stress.

4.5 Oscillation onset and system size

Our experimental observations reveal two non-trivial features. First, the oscillations emerge at critical biofilm size (Figure 4.15a). Second, the period of the oscillations increases with time (Figure 4.15b). We now ask wheter these two features can be reproduced in our mathematical model.

To do so, we now explore in our model the effect of size of the biofilm. Specifically, we consider the effect of the change in relative sizes between the two populations as the biofilm grows.

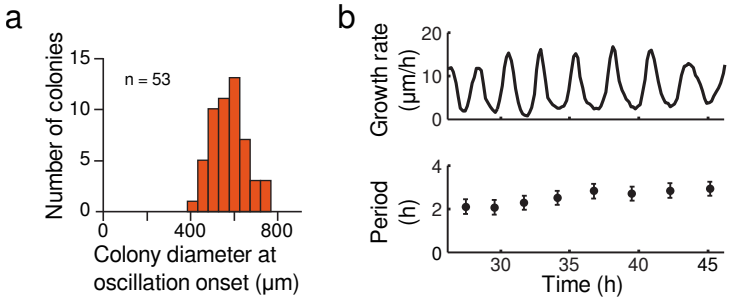

Figure 4.15: Oscillations start when the biofilm is large enough and become slower over time. a, Histogram of the diameter at which a colony begins to oscillate ($n = 53$ colonies, $mean = 576\mu m$, $s.d. = 85\mu m$). b, Top: growth rate of an oscillating colony. Bottom: period of each oscillation cycle, measured peak to peak. The error bars ($\pm$20 min) are determined by the imaging frequency (1 frame per 10 min). The period slightly increases over time. From Liu et al. (2015).

4.5.1 Relative size of interior and periphery

When cells grow and divide, the biofilm expands. We consider, however, that while this happens the cellular density remains approximately constant at any fixed distance to the edge of the biofilm, as shown by experimental data. Furthermore, we also consider that the cellular population is saturated from a given distance of the biofilm edge and inwards.

Let us approximate this situation by defining two discrete subpopulations occupying concentric areas in a circular biofilm, namely interior and periphery. By definition, the number of cells in the whole biofilm (N_T) is

a simple addition of the cells in interior (N_i) and periphery (N_p):

$$N_T = N_i + N_p \tag{4.21}$$

Specifically, let the interior section of the biofilm be a circle of radius R_i and the periphery a ring around the latter of width R_p constant in time (Figure 4.16a). In other words, the boundary between the two regions is found at a constant distance R_p from the biofilm edge. Then, the areas a_i and a_p occupied by each population can be easily calculated as

$$a_i = \pi R_i^2 \tag{4.22}$$

$$a_p = \pi(R_i + R_p)^2 - \pi R_i^2 \tag{4.23}$$

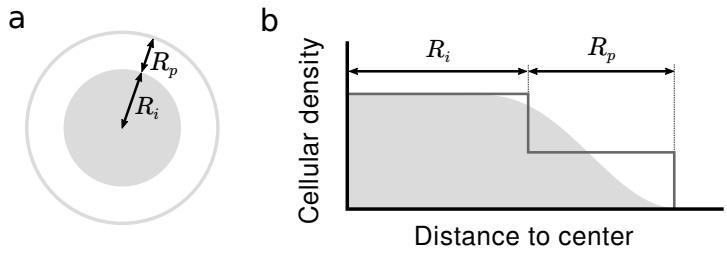

Figure 4.16: Modelling the biofilm as two concentric circular regions of constant density. a, Two distinct areas of the biofilm are considered: a central circle with radius R_i surrounded by a ring of width R_p. b, The cellular density profile is approximated with a step function. The width of the periphery ring R_p is such that the average density is close to saturation in the central circle and approximately half way to saturation in the periphery ring.

Furthermore, we can simplify the gradient of cellular density, static with respect of the distance to the biofilm edge, as a step function defined by the average density of each subpopulation and the distance from boundary between them to the biofilm edge R_p. The distance R_p is such

that cellular density of the interior subpopulation (ρ_i) is close to saturation while cellular density of the periphery subpopulation (ρ_p) is half-way to saturation (Figure 4.16b).

The dynamics of each subpopulation would be governed by a logistic growth if there were no transference of cells between them. This is not the case, however. The growth of the cells in the periphery pushes the biofilm edge forward, which, by definition moves the boundary between the two subpopulation and leaves behind some cells. These cells, let too far inward, will become part of the interior subpopulation. In other words, there is a net transference of cells from the periphery to the interior. This transference of cells ($\phi_{p\to i}$) depends in a non-trivial manner on the growth of both populations and on the radius of the interior circle (R_i). Taking this transference of cells into account, the dynamics of both subpopulations would be defined by:

$$\frac{dN_i}{dt} = r_i N_i \left(1 - \frac{N_i}{a_i K(G_i)}\right) + \phi_{p\to i} \tag{4.24}$$

$$\frac{dN_p}{dt} = r_p N_p \left(1 - \frac{N_p}{a_p K(G_p)}\right) - \phi_{p\to i} \tag{4.25}$$

where $K(G)$ is the carrying capacity (in concentration terms) and depends on the glutamate concentration as

$$K(G) = \frac{G^m}{K_k^m + G^m} \tag{4.26}$$

We can circumvent the calculation of $\phi_{p\to i}$ by taking together Eqs. (4.21), (4.24) and (4.25). In this way we can write the dynamics of

total number of cells as

$$\frac{dN_T}{dt} = r_i N_i \left(1 - \frac{N_i}{a_i K(G_i)}\right) + r_p N_p \left(1 - \frac{N_p}{a_p K(G_p)}\right) \tag{4.27}$$

where we now need to define N_i and N_p as functions of N_T:

$$N_i = \rho_i a_i = \rho_i \pi R_i^2 \tag{4.28}$$

$$N_p = \rho_p a_p = \rho_p [\pi (R_i + R_p)^2 - \pi R_i^2] = \rho_p \pi [2R_p R_i + R_p^2] \tag{4.29}$$

To do so, we need to find how R_i depends on N_T

$$\begin{aligned} N_T &= \rho_i a_i + \rho_p a_p = \rho_i \pi R_i^2 + \rho_p \pi [2R_p R_i + R_p^2] \\ 0 &= \rho_i R_i^2 + 2\rho_p R_p R_i + \rho_p R_p^2 - \frac{N_T}{\pi} \end{aligned}$$

which is a second order equation that can be easily solved

$$R_i = \frac{-2\rho_p R_p \pm \sqrt{4(\rho_p R_p)^2 - 4\rho_i \left(\rho_p R_p^2 - \frac{N_T}{\pi}\right)}}{2\rho_i}$$

But as negative values of R_i are not physically reasonable, we are left with the solution:

$$R_i(N_T) = \frac{-2\rho_p R_p + \sqrt{4(\rho_p R_p)^2 - 4\rho_i \left(\rho_p R_p^2 - \frac{N_T}{\pi}\right)}}{2\rho_i} \tag{4.30}$$

which finally allows us to describe the growth of the system with

$$\frac{dN_T}{dt} = r_i N_i(N_T)\left(1 - \frac{N_i(N_T)}{a_i K(G_i)}\right) + r_p N_p(N_T)\left(1 - \frac{N_p(N_T)}{a_p K(G_p)}\right) \tag{4.31}$$

$$N_i(N_T) = \rho_i \pi R_i(N_T)^2 \tag{4.32}$$

$$N_p(N_T) = \rho_p \pi [2R_p R_i(N_T) + R_p^2] \tag{4.33}$$

$$R_i(N_T) = \frac{-2\rho_p R_p + \sqrt{4(\rho_p R_p)^2 - 4\rho_i\left(\rho_p R_p^2 - \frac{N_T}{\pi}\right)}}{2\rho_i} \tag{4.34}$$

4.5.2 Death in the interior

In some experiments we expect to see cell death. In that case, as the densities of both interior and periphery are constant, the model described here would predict that the biofilm would shrink. To prevent this unrealistic result we need to explicitly consider the fact that dead cells do not disappear, and that they may still occupy a physical space. Based on this, we will assume that dead cells still count towards maintaining a constant density.

Let us define N_d as the number of dead cells in the interior of the biofilm, whose rate of change will be given by

$$\frac{dN_d}{dt} = \Theta\left(r_i N_i\left(1 - \frac{N_i}{a_i K(G_i)}\right)\right) \tag{4.35}$$

$$\Theta(x) = \begin{cases} -x & \text{if } x < 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.36}$$

so that any decrease in the number of cells in the interior is considered as death. Considering that the constant cell density in the interior takes into

account both cells that are death and alive, Eq. (4.28) would become

$$N_i + N_d = \rho_i a_i = \rho_i \pi R_i^2 \tag{4.37}$$

and Eq. (4.30) would become

$$R_i(N_T) = \frac{-2\rho_p R_p + \sqrt{4(\rho_p R_p)^2 - 4\rho_i \left(\rho_p R_p^2 - \frac{N_T + N_d}{\pi}\right)}}{2\rho_i} \tag{4.38}$$

so that, finally, the growth of the system would be described by

$$\frac{dN_T}{dt} = r_i N_i(N_T)\left(1 - \frac{N_i(N_T) + N_d}{a_i K(G_i)}\right) + r_p N_p(N_T)\left(1 - \frac{N_p(N_T)}{a_p K(G_p)}\right) \tag{4.39}$$

$$\frac{dN_d}{dt} = \Theta\left(r_i N_i\left(1 - \frac{N_i}{a_i K(G_i)}\right)\right) \tag{4.40}$$

$$\Theta(x) = \begin{cases} -x & if x < 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.41}$$

$$N_i(N_T) = \rho_i \pi R_i(N_T)^2 - N_d \tag{4.42}$$

$$N_p(N_T) = \rho_p \pi [2R_p R_i(N_T) + R_p^2] \tag{4.43}$$

$$R_i(N_T) = \frac{-2\rho_p R_p + \sqrt{4(\rho_p R_p)^2 - 4\rho_i \left(\rho_p R_p^2 - \frac{N_T + N_d}{\pi}\right)}}{2\rho_i} \tag{4.44}$$

4.5.3 Effect of the subpopulation size on the oscillator

All variables in the metabolic oscillator represent intensive values (they do not depend on the total volume or area). Additionally, all of them except A represent concentrations within one of the two compartments of

the biofilm (namely interior and periphery). Thus, none of them will be affected by changes in the relative sizes of the two compartments. Interestingly, the equation describing the evolution of A accounts for reactions that occur in both compartments: production and consumption in the interior and consumption in the periphery. The relative importance of these reactions will be given by the relative size of the two subpopulations.

If we define f_i as the fraction of cells found in the interior of the biofilm, $f_i = \frac{N_i}{N_T}$, Eq. (4.1) becomes

$$\frac{dA}{dt} = \alpha f_i G_i H_i - \delta A(f_i r_i + (1 - f_i) r_p) \tag{4.45}$$

This model reproduces the size-dependent onset of the oscillations. Specifically, oscillations start when the cells that are far enough from the biofilm edge to be considered interior cells represent at least a given fraction of the total (Figure 4.17a).

The model enters the oscillatory regime through a supercritical Andronov-Hopf bifurcation, and thus it predicts that the stable limit cycle appears with zero amplitude and grows in a continuous manner as f_i increases (Figure 4.17b). Apparently, this would contradict the experimental results where the oscillations appear with a non-null amplitude (Figure 4.2f). The bifurcation diagram, however, shows that the amplitude of the limit cycle increases rapidly within a small range of values of f_i.

Furthermore, the temporal dynamics of the system reveal that in the simulation the system exhibited detectable oscillations much after it had crossed the Andronov-Hopf bifurcation. This implies that the time scale of the transient to escape the unstable fixed point is slow with respect to the dynamics of the fraction of interior cells f_i (Figure 4.17a and b).

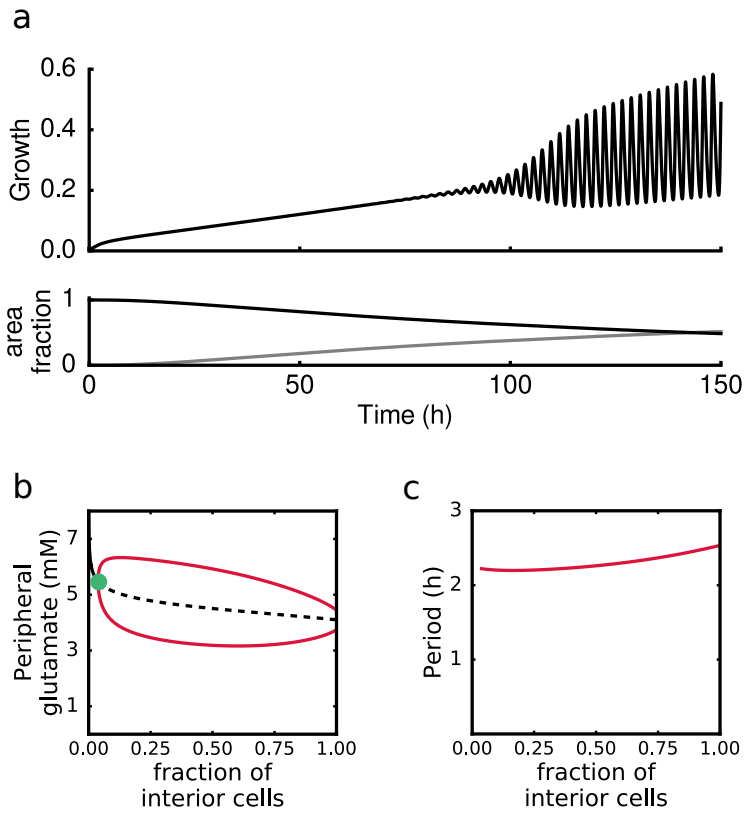

Figure 4.17: Effect of biofilm size on growth oscillations. a, Colony growth in time as the biofilm becomes larger and the interior subpopulation grows with respect to the periphery (top). Below, fraction of the total number of cells alive found in the interior (grey) and in the periphery (black). b, Bifurcation diagrams showing systematic analysis on the effects of the fraction of interior cells f_i. The solid black line denotes stable fixed point. The dashed black line corresponds to an unstable fixed point. The red lines correspond to the extrema of oscillations in peripheral glutamate (stable limit cycle). The green dot marks the point where the system undergoes a supercritical Andronov-Hopf bifurcation. c, Effect of the fraction of interior cells on the period of the oscillations.

We argue that small perturbations present during the experiment, while clearly insufficient to cause the oscillations, might have a role causing a faster and more abrupt transition of the system towards the limit cycle attractor once it has already crossed the Andronov-Hopf bifurcation. This would also help to explain the sudden appearance of oscillations with a non-null amplitude.

Finally, the model also accounts for the observed slight increase of the oscillation period by considering an increase in the ratio of interior to peripheral cells over time (Figures 4.15a and 4.17c).

4.6 Discussion

The data presented here reveal that intracellular metabolic activity within biofilms is organized in space and time, giving rise to co-dependence between interior and peripheral cells. Even though bacteria are single-celled organisms, the metabolic dynamics of individual cells can thus be regulated in the context of the community. This metabolic co-dependence can, in turn, give rise to collective oscillations that emerge during biofilm formation and promote the resilience of biofilms against chemical attack (Figure 4.18). The community-level oscillations also support the ability of biofilms to reach large sizes, while retaining a viable population of interior cells. Specifically, periodic halting of peripheral cell growth prevents complete starvation and death of the interior cells. This overcomes the colony size limitation for a viable biofilm interior that would otherwise be imposed by nutrient consumption in the biofilm periphery. Metabolic co-dependence in biofilms therefore offers a powerful mechanism to coordinate colonies of milions of organisms, an elegant solution that resolves

the social conflict between cooperation (protection) and competition (starvation) through oscillations.

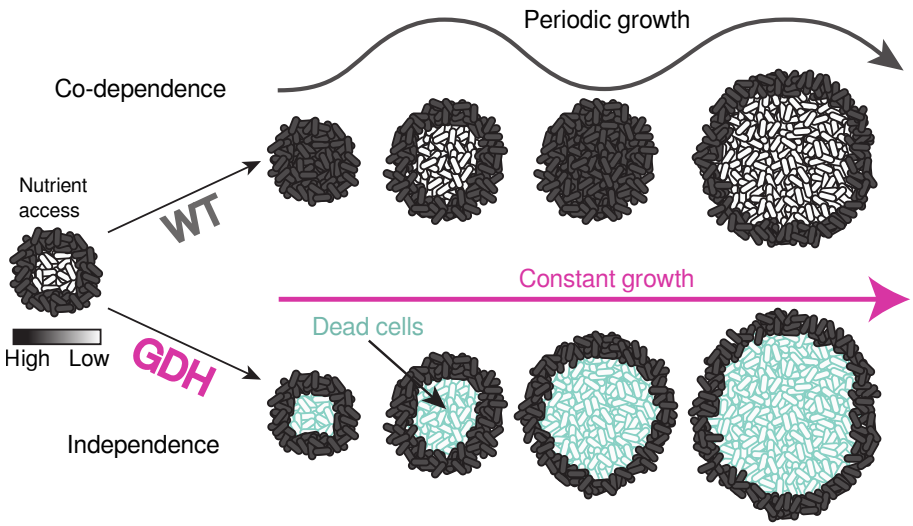

Figure 4.18: Metabolic coupling ensures the viability of the whole colony. Co-dependence between interior and peripheral cells exhibited in a wild-type strain results in a growth strategy that sustains the viability of interior cells, while independence enforced by a GDH overexpression strain results in starvation of interior cells and reduced resilience to external attack. From Liu et al. (2015).

The discovery of biofilm oscillations presented here also raises new questions. While cellular processes such as swarming or expression of extracellular matrix components are not required for the observed biofilm oscillations (Liu et al., 2015), it will be interesting to pursue whether such cellular processes are influenced by oscillatory dynamics (Anyan et al., 2014). Another question worth pursuing is whether metabolic co-dependence can also arise in other biofilm-forming species. Perhaps other metabolic branches where metabolites can be shared among cells could also give rise to oscillations in biofilm growth. It will be interesting to

pursue these questions in future studies to obtain a better understanding of biofilm development.

Our observations also suggest future strategies to cope with the intriguing resilience of biofilms in the face of environmental stresses, such as antibiotic exposure. In particular, our findings show that straightforward application of stress (such as H_2O_2 or chloramphenicol) to the biofilm counter-intuitively promotes growth, effectively rejuvenating the biofilm. Death of the colony periphery relieves the repression on the growth of interior cells, allowing them to regenerate a new biofilm periphery and interior. In contrast, manipulation of the metabolic co-dependence may yield a more effective approach to control biofilm formation. Specifically, promoting continuous growth of peripheral cells can starve the biofilm interior, leaving behind the exposed peripheral cells that can more easily be targeted by external killing factors. Therefore, the metabolically driven collective oscillations in biofilm expansion described here not only reveal fundamental insights into the principles that govern formation of multicellular communities, but also suggest new strategies for manipulating the growth of biofilms.

5 *Conclusions*

In this thesis we have analysed the dynamical and computational capabilities of a variety of regulatory mechanisms of cells and cellular communities. To do so so we have explored a wide range of dynamical behaviours and regulatory scales. Starting by studying the bistable behaviour of a small, yet rich, genetic circuit formed by a small set of genes, we have then described how gene regulatory networks of thousands of genes can encode in a decentralised manner the recent history of the cell in their transient dynamics. Finally we have characterised the collective metabolic oscillations in communities of millions of cells. All in all, this varied, albeit not exhaustive, repertoire is a reflection of the wide diversity of dynamical properties and regulatory scales that biological systems exhibit.

Specifically, in Chapter 2 we analysed the dynamic response of the GAL regulatory circuit of *S. cerevisiae*. First, we characterised the bistable dynamic response of this circuit to galactose. We quantified, using population data, how the concentration of galactose affects the rate of stochastic activation of the circuit, and thus the memory stability of this bistable system. Furthermore, we developed and adjusted two models of increasing complexity that allow the exploration of how

genetic variants affect the behaviour of this robust memory circuit. In the process, we inferred the abundance of the different components of the GAL regulatory circuit, and the mapping between external galactose level and the activation rate of Gal1p and Gal3p as transcription activators. This work supposes a first step towards characterising the epistatic interactions that affect the dynamics of a regulatory system.

In Chapter 3 we used the reservoir-computing paradigm to explore the computational capabilities of the transcriptional regulatory networks of five organisms. This paradigm, developed in the fields of machine learning and computational neuroscience, defines a system composed by a network of interacting elements that is able to process complex temporal information, and at the same time can be trained very efficiently. Specifically, such system is composed by the reservoir, which is a subnetwork with recurrences —i.e. cyclic paths— that encodes temporal information in its complex dynamics, and the readout, and a simple feed-forward structure that can extract the relevant information from the transient multidimensional dynamics of the reservoir. We propose a mechanism by which the integration of information in regulatory networks happens in its dynamics at the scale of the whole system, not by specialised architectures but in a decentralised manner.

We found that the gene regulatory networks of five distant organisms, namely *E. coli*, *B. subtilis*, *S. cerevisiae*, *D. melanogaster*, and *H. sapiens* share a similar organisation, where a group of genes with recurrent regulatory interactions is upstream of and controls most of the network. We also showed that this group of nodes with recurrent connections, which we call core or reservoir, defines a network that can encode temporal information through its dynamics in a state-dependent manner.

Our study shows that these core networks are as efficient in encoding temporal information as random networks, which are the *de facto* gold standard in reservoir computing. Furthermore, in the case of *E. coli*, we showed that input signals arriving at the reservoir through biologically relevant stress pathways can be encoded efficiently in the reservoir dynamics. We also show that this information can be used by readout structures originated by an evolutionary process, in a novel extension of the reservoir-computing paradigm that has been absent so far in its application to computational neuroscience and machine learning, where the evolutionary approach was not directly applicable or evident.

Finally, in Chapter 4 we characterised the mechanism underlying the growth oscillations in biofilms of *B. subtilis*, which we discovered in collaboration with the laboratory of Gürol Süel in the University of California San Diego. Our study revealed that these oscillations arise from a metabolic co-dependence between different subpopulations of the biofilm, depending on their distance to the colony edge. Specifically, cells in the interior of the biofilm depend on the glutamate that diffuses inwards without being consumed by cells on the periphery, and cells on the periphery depend on the ammonium that cells in the interior synthesise from that glutamate. Thus, when interior cells starve, they stop producing ammonium, which forces peripheral cells to stop growing and lets glutamate to flow inwards. From a conceptual point of view, this system is an instance of a spatially-extended negative-feedback loop, and thereby originates collective metabolic oscillations.

Using a mathematical model we predicted that in the event of an external attack the cells sheltered in the interior of the colony are able to regenerate it, but that if the co-dependence is interrupted the interior starves

and the biofilm becomes vulnerable. Both predictions were experimentally validated, which confirms that the oscillations maintain a dynamic balance between the two conflicting interests of the biofilm: growth and protection. Finally, we propose that the size-dependent onset of the oscillations and period variation is a consequence of the relative size of the interior and peripheral populations, which changes as the biofilm grows. Interestingly, all these results open the possibility to new strategies to cope with the resistance of biofilms to external aggressions such as antibiotics and antimicrobial agents.

Future outlook

As almost always happens, informative answers open perspectives that point at new unknowns and provoke new questions. The same occurs with the findings and conclusions from this thesis. After characterising the dynamics of the GAL regulatory network, we are now in position to explore how mutations in different parts of the system combine, and whether they produce epistatic effects.

Similarly, once we have proved that gene regulatory networks could in theory work as computational reservoirs, we would like to try to pursue an experimental validation. A complementary approach would be to explore if the rate of fixations of mutations is significantly different in the distinct parts of the network. Specifically, we expect mutations in the genes of the reservoir to be much rarer, as they would affect the dynamics of the whole netowrk. On the other hand, mutations in the genes that form the readout should have a much more local effect, and could be interpreted as part of the training process. We are also interested in exploring computationally the determinants that could allow the evolution of not only of the readout, but of a reservoir itself.

Finally, by simplifying the biofilm as two discrete subpopulations, we discovered that the growth oscillations are caused by a spatially-extended negative-feedback loop. Now, a spatially-continuous model of the biofilm would allow the characterisation of this spatially-extended feedback loop.

A Bow-tie structure of biological networks

In this appendix, the regulatory networks analysed in Chapter 3 are characterized in terms of the bow-tie network structure. This classification of the nodes of the network was initially proposed to describe the world wide web (Broder et al., 2000). Whether the edges of the network refer to hyperlinks connecting web pages or to regulatory interactions in a gene network, the bow-tie classification helps to characterise the structure of the information flow in the network.

Figure A.1 schematises the different groups of nodes defined by the modified version of the bow-tie scheme used here. Instead of considering a central strongly connected component, we defined our bow-tie groups around the recurrent group of nodes found after pruning the network (Section 3.1). Table A.1 shows the number of nodes counted in each of the defined categories.

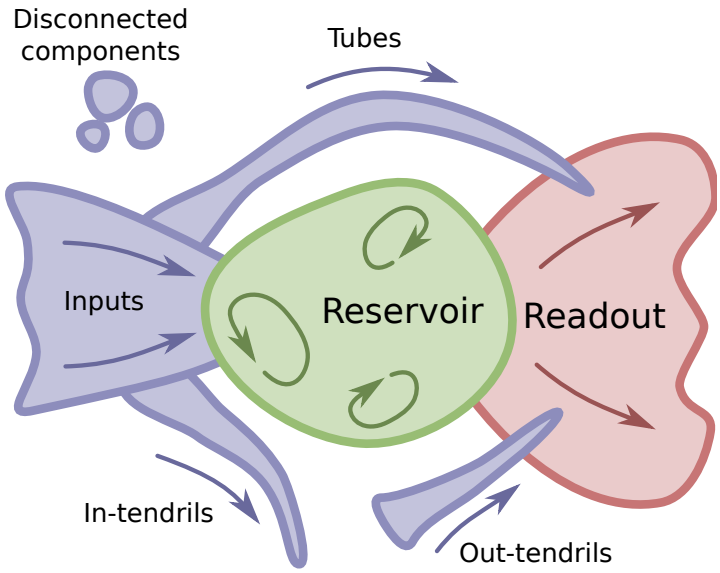

Figure A.1: Scheme of the bow-tie structure classification of nodes. First the recurrent core, or reservoir, of the network is found as described in Section 3.1. Then, all those nodes that can be reached from at least one node in the reservoir following the directed edges are considered part of the readout. All nodes that can reach at least one node in the reservoir are counted as input nodes. Those nodes that can be reached from the input group and can reach the readout group are defined as tubes. All nodes that can either be reached from the input or reach the readout but not both are considered in-tendrils and out-tendrils, respectively. Finally, the rest of the nodes are grouped as others.

	B. subtilis	*E. coli*	*S. cerevisiae*	*D. melanogaster*	*H. sapiens*
Total nodes	921	3243	6725	9440	16356
Input	6	2	0	0	400
Reservoir	13	70	289	486	207
Readout	1037	6158	12855	17516	26946
Tubes	1	1	0	0	1
in-Tendrils	161	3	0	0	2069
out-Tendrils	61	20	0	13	128
Other	107	7	0	138	52

Table A.1: Sizes of the different groups of nodes defined by bow-tie classification.

Contributions

Scientific articles

- J. Liu, **M. Gabalda-Sagarra**[*], A. Prindle[*], J. Humphries[*], M. Asally[*], D. D. Lee, S. Ly, J. Garcia-Ojalvo, G. M. Suel (2015). "Metabolic co-dependence gives rise to collective oscillations within biofilms". *Nature*, 523(7562), 550-4. http://doi.org/10.1038/nature14660
  [*] These authors contributed equally to this work, order changed for emphasis.
- **M. Gabalda-Sagarra**, L. Carey, J. Garcia-Ojalvo (2014). "State-dependent information processing in gene regulatory networks". *bioRxiv* preprint 10.1101/010124. http://dx.doi.org/10.1101/010124
  to be submitted to *PLOS comp bio*.
- **M. Gabalda-Sagarra**, A. New, B. Lehner, J. Garcia-Ojalvo. (In preparation) "Dissection of a Stochastic Memory Circuit".

Other articles not related with this thesis

- J. Liu, R. Martinez-Corral, A. Prindle, **M. Gabalda-Sagarra**, J. Garcia-Ojalvo, G. Suel. (Under revision) "Coupling between distant biofilms and emergence of nutrient-sharing". *Science*.
- **M. Gabalda-Sagarra**, A. MacLean, J. Garcia-Ojalvo, M. Stumpf. (In preparation) "Mind the Z: approximate Bayesian computation with cells in the space".

Oral communications

- **M. Gabalda-Sagarra**, L. Carey, J. Garcia-Ojalvo. "State-dependent information processing in gene regulatory networks". Advanced Lecture Course on Computational Systems Biology (CompSysBio). Aussois, France 2015.

- J. Liu, **M. Gabalda-Sagarra**[*], A. Prindle[*], J. Humphries[*], M. Asally[*], D. D. Lee, S. Ly, J. Garcia-Ojalvo, G. M. Suel. "Long-range synchronized oscillations compensate biofilm growth and resilience". 9th q-bio summer school. Fort Colins, Colorado, United States, 2015.
  [*] These authors contributed equally to this work.
- **M. Gabalda-Sagarra**, L. Carey, J. Garcia-Ojalvo. "State-dependent information processing in gene regulatory networks". 25th International Conference on Artificial Neural Networks (ICANN). Barcelona, Spain 2016.
- **M. Gabalda-Sagarra**, L. Carey, J. Garcia-Ojalvo. "State-dependent information processing in gene regulatory networks". International Conference on Systems Biology (ICSB). Barcelona, Spain 2016.

Posters

- **M. Gabalda-Sagarra**, J. Garcia-Ojalvo. "Reservoir Computing Functionality of Cellular Regulatory Networks". European Conference on Complex Systems. Barcelona, Spain 2013.
- **M. Gabalda-Sagarra**, J. Garcia-Ojalvo. "Reservoir Computing Functionality of Cellular Regulatory Networks". XIX Congreso de Física Estadística. Ourense, Spain 2014.
- **M. Gabalda-Sagarra**, L. Carey, J. Garcia-Ojalvo. "State-dependent information processing in gene regulatory networks". Advanced Lecture Course on Computational Systems Biology (CompSysBio). Aussois, France 2015.
- **M. Gabalda-Sagarra**, L. Carey, J. Garcia-Ojalvo. "State-dependent information processing in gene regulatory networks". 9th q-bio summer school. Fort Colins, Colorado, United States, 2015.
- J. Liu, **M. Gabalda-Sagarra**[*], A. Prindle[*], J. Humphries[*], M. Asally[*], D. D. Lee, S. Ly, J. Garcia-Ojalvo, G. M. Suel. "Long-range synchronized oscillations compensate biofilm growth and resilience". Biological Oscillators: Design, Mechanism, Function. EMBL Heidelberg, Germany 2015.
  [*] These authors contributed equally to this work.
- **M. Gabalda-Sagarra**, L. Carey, J. Garcia-Ojalvo. "State-dependent information processing in gene regulatory networks" Biological Oscillators: Design, Mechanism, Function. EMBL Heidelberg, Germany 2015.

Bibliography

Abramczyk, D., Holden, S., Page, C. J., and Reece, R. J. (2012). Interplay of a ligand sensor and an enzyme in controlling expression of the Saccharomyces cerevisiae GAL genes. *Eukaryotic Cell*, 11(3):334–342. (cit. on pp. 13, 52, and 53)

Acar, M., Becskei, A., and van Oudenaarden, A. (2005). Enhancement of cellular memory by reducing stochastic transitions. *Nature*, 435(7039):228–32. (cit. on pp. 15, 16, 34, 35, 43, 46, 50, 51, 52, 54, and 70)

Al-Shyoukh, I., Yu, F., Feng, J., Yan, K., Dubinett, S., Ho, C.-M., Shamma, J. S., and Sun, R. (2011). Systematic quantitative characterization of cellular responses induced by multiple signals. *BMC systems biology*, 5:88. (cit. on pp. 3)

Alberts, B. (1998). The Cell as a Collection of Protein Machines. *Cell*, 92:291–294. (cit. on pp. 2)

Almaas, E., Kovács, B., Vicsek, T., Oltvai, Z. N., and Barabási, A.-L. (2004). Global organization of metabolic fluxes in the bacterium Escherichia coli. *Nature*, 427(6977):839–43. (cit. on pp. 3)

Alon, U. (2003). Biological networks: the tinkerer as an engineer. *Science (New York, N.Y.)*, 301(5641):1866–7. (cit. on pp. 2)

Alon, U. (2007a). *An Introduction to Systems Biology: Design Principles of Biological Circuits*, volume 10. Chapman and Hall/CRC, Boca Raton, 1 edition. (cit. on pp. 26 and 98)

Alon, U. (2007b). Network motifs: theory and experimental approaches. *Nat Rev Genet*, 8(6):450–61. (cit. on pp. 2)

Anyan, M. E., Amiri, A., Harvey, C. W., Tierra, G., Morales-Soto, N., Driscoll, C. M., Alber, M. S., and Shrout, J. D. (2014). Type IV pili interactions promote intercellular association and moderate swarming of Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences of the United States of America*, 111(50):201414661. (cit. on pp. 138)

Armus, H. L., Montgomery, A. R., and Gurney, R. L. (2006). Discrimination Learning and Extinction in Paramecia (P. caudatum). *Psychological Reports*, 98(3):705–711. (cit. on pp. 99)

Asally, M., Kittisopikul, M., Rué, P., Du, Y., Hu, Z., Çağatay, T., Robinson, A. B., Lu, H., Garcia-Ojalvo, J., and Süel, G. M. (2012). Localized cell death focuses mechanical forces during 3D patterning in a biofilm. *Proceedings of the National Academy of Sciences of the United States of America*, 109(46):18891–6. (cit. on pp. 6 and 102)

Astrom, K. J. and Murray, R. M. (2012). *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton Univ Press, Princeton. (cit. on pp. 7)

Atiya, A. F. and Parlos, A. G. (2000). New results on recurrent network training: unifying the algorithms and accelerating convergence. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 11(3):697–709. (cit. on pp. 85)

Barabási, A.-L. and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature reviews. Genetics*, 5(2):101–13. (cit. on pp. 28)

Bardwell, L., Zou, X., Nie, Q., and Komarova, N. L. (2007). Mathematical models of specificity in cell signaling. *Biophysical journal*, 92(10):3425–41. (cit. on pp. 3)

Bauer, A. L., Jackson, T. L., Jiang, Y., and Rohlf, T. (2010). Receptor cross-talk in angiogenesis: mapping environmental cues to cell phenotype using a stochastic, Boolean signaling network model. *Journal of theoretical biology*, 264(3):838–46. (cit. on pp. 24)

Becskei, A., Séraphin, B., Serrano, L., A-Mohammadi, S., Hawkins, R., Bateman, E., Becskei, A., Serrano, L., Belli, G., Gari, E., Piedrafita, L., Aldea, M., Herrero, E., Bouhassira, E., Westerman, K., Leboulch, P., Cormack, B., Bertram, G., Egerton, M., Gow, N., Falkow, S., Brown, A., Elowitz, M., Leibler, S., Ferrell, J., Machleder, E., Fiering, S., Whitelaw, E., Martin, D., Forsberg, E., Zaboikina, T., Versaw, W., Ahn, N., Bresnick, E., Freeman, M., Gardner, T., Cantor, C., Collins, J., Gari, E., Piedrafita, L., Aldea, M., Herrero, E., Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., Bujard, H., Gunge, N., Hasty, J., Pradines, J., Dolnik, M., Collins, J., Hume, D., Keller, A., Ko, M., Nakauchi, H., Takahashi, N., Kringstein, A., Rossi, F., Hofmann, A., Blau, H., Lohr, D., Venkov, P., Zlatanova, J., Novick, A., Weiner, M., Osheim, Y., Mougey, E., Windle, J., Anderson, M., O'Reilly, M., Miller, O., Beyer, A., Sollner-Webb, B., Polach, K., Widom, J., Rossi, F., Kringstein, A., Spicher, A., Guicherit, O., Blau, H., Smolen, P., Baxter, D., Byrne, J., Struhl, K., Thieffry, D., Huerta, A., Perez-Rueda, E., Collado-Vides, J., Vashee, S., Melcher, K., Ding, W., Johnston, S., Kodadek, T., Walters, M., Fiering, S., Eidemiller, J., Magis, W., Groudine, M., Martin, D., Walz, D., Caplan, S., Wang, J., Ellwood, K., Lehman, A., Carey, M., and She, Z. (2001). Positive feedback in eukaryotic gene networks: cell differentiation by

graded to binary response conversion. *The EMBO Journal*, 20(10):2528–2535. (cit. on pp. 10)

Becskei, a. and Serrano, L. (2000). Engineering stability in gene networks by autoregulation. *Nature*, 405(June):590–593. (cit. on pp. 7)

Belitsky, B. R. and Sonenshein, A. L. (1998). Role and regulation of Bacillus subtilis glutamate dehydrogenase genes. *Journal of bacteriology*, 180(23):6298–305. (cit. on pp. 105)

Bennett, M. R., Pang, W. L., Ostroff, N. a., Baumgartner, B. L., Nayak, S., Tsimring, L. S., and Hasty, J. (2008). Metabolic gene regulation in a dynamically changing environment. *Nature*, 454(7208):1119–22. (cit. on pp. 15)

Berg, H. and Purcell, E. (1977). Physics of chemoreception. *Biophysical Journal*, 20(2):193–219. (cit. on pp. 25)

Berk, V., Fong, J. C. N., Dempsey, G. T., Develioglu, O. N., Zhuang, X., Liphardt, J., Yildiz, F. H., and Chu, S. (2012). Molecular Architecture and Assembly Principles of Vibrio cholerae Biofilms. *Science*, 337(6091):236–239. (cit. on pp. 101 and 103)

Betz, A. and Chance, B. (1965). Phase Relationship of Glycolytic Intermediates in Yeast Cells with Oscillatory Metabolic Control. *Archives of biochemistry and biophysics*, 109:585–94. (cit. on pp. 21)

Bhalla, U. S. and Iyengar, R. (1999). Emergent Properties of Networks of Biological Signaling Pathways. *Science Bioinorg. Chem. Chem. Biol. Nature Beechem Biophys. J. Proc. Natl. Acad. Sci. U.S.A. J. Phys. Chem*, 381(102):1247–12759. (cit. on pp. 7, 23, and 28)

Bhat, P. J. and Hopper, J. E. (1992). Overproduction of the GAL1 or GAL3 protein causes galactose-independent activation of the GAL4 protein: evidence for a new model of induction for the yeast GAL/MEL regulon. *Molecular and cellular biology*, 12(6):2701–2707. (cit. on pp. 13, 52, and 54)

Boedecker, J., Obst, O., Lizier, J. T., Mayer, N. M., and Asada, M. (2012). Information processing in echo state networks at the edge of chaos. *Theory in Biosciences*, 131(3):205–213. (cit. on pp. 32 and 89)

Boogerd, F. C., Ma, H., Bruggeman, F. J., van Heeswijk, W. C., García-Contreras, R., Molenaar, D., Krab, K., and Westerhoff, H. V. (2011). AmtB-mediated NH 3 transport in prokaryotes must be active and as a consequence regulation of transport by GlnK is mandatory to limit futile cycling of NH4+/NH3. *FEBS Letters*, 585(1):23–28. (cit. on pp. 108)

Bornholdt, S. and Sneppen, K. (2000). Robustness as an evolutionary principle. *Proceedings. Biological sciences / The Royal Society*, 267(1459):2281–2286. (cit. on pp. 23)

Branda, S. S., Gonzalez-Pastor, J. E., Ben-Yehuda, S., Losick, R., and Kolter, R. (2001). Fruiting body formation by Bacillus subtilis. *Proceedings of the National Academy of Sciences*, 98(20):11621–11626. (cit. on pp. 3, 101, and 105)

Brandman, O., Ferrell, J. E., Li, R., and Meyer, T. (2005). Interlinked Fast and Slow Positive Feedback Loops Drive Reliable Cell Decisions. *Science*, 310(5747):496–498. (cit. on pp. 10)

Brandman, O. and Meyer, T. (2008). Feedback loops shape cellular signals in space and time. *Science (New York, N.Y.)*, 322(5900):390–5. (cit. on pp. 7)

Broach, J. R. (2012). Nutritional control of growth and development in yeast. *Genetics*, 192(1):73–105. (cit. on pp. 33)

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., and Wiener, J. (2000). Graph structure in the Web. *Computer Networks*, 33(1):309–320. (cit. on pp. 81 and 147)

Broome, B. M., Jayaraman, V., and Laurent, G. (2006). Encoding and decoding of overlapping odor sequences. *Neuron*, 51(4):467–82. (cit. on pp. 31)

Bryan, A. K., Goranov, A., Amon, A., and Manalis, S. R. (2010). Measurement of mass, density, and volume during the cell cycle of yeast. *Proc Natl Acad Sci U S A*, 107(3):999–1004. (cit. on pp. 60)

Bryant, G. O., Prabhu, V., Floer, M., Wang, X., Spagna, D., Schreiber, D., and Ptashne, M. (2008). Activator Control of Nucleosome Occupancy in Activation and Repression of Transcription. *PLoS Biology*, 6(12):e317. (cit. on pp. 14)

Bundschuh, R., Hayot, F., and Jayaprakash, C. (2003). Fluctuations and Slow Variables in Genetic Networks. *Biophysical Journal*, 84(3):1606–1615. (cit. on pp. 58)

Buonomano, D. V. and Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. *Nature reviews. Neuroscience*, 10(2):113–25. (cit. on pp. 30, 31, 32, 75, 76, and 97)

Büsing, L., Schrauwen, B., and Legenstein, R. (2010). Connectivity, dynamics, and memory in reservoir computing with binary and analog neurons. *Neural computation*, 22(5):1272–311. (cit. on pp. 99)

Cai, L. and Tu, B. P. (2012). Driving the cell cycle through metabolism. *Annual review of cell and developmental biology*, 28:59–87. (cit. on pp. 33)

Cannon, W. B. (1932). *The wisdom of the body*. New York, NY, US: W W Norton & Co. (cit. on pp. 25)

Castorph, H. and Kleiner, D. (1984). Some properties of a Klebsiella pneumoniae ammonium transport negative mutant (Amt-). *Archives of microbiology*, 139(2-3):245–7. (cit. on pp. 108)

Chambers, L. D., Stokes, K. R., Walsh, F. C., and Wood, R. J. K. (2006). Modern approaches to marine antifouling coatings. *Surface and Coatings Technology*, 201(6):3642–3652. (cit. on pp. 5)

Chang, L. K., Putcha, G. V., Deshmukh, M., and Johnson, E. M. (2002). Mitochondrial involvement in the point of no return in neuronal apoptosis. *Biochimie*, 84(2-3):223–231. (cit. on pp. 10)

Chen, Z., Odstrcil, E. a., Tu, B. P., and McKnight, S. L. (2007). Restriction of DNA replication to the reductive phase of the metabolic cycle protects genome integrity. *Science (New York, N.Y.)*, 316(5833):1916–1919. (cit. on pp. 17)

Cheong, R., Hoffmann, A., and Levchenko, A. (2008). Understanding NF-kappaB signaling via mathematical modeling. *Molecular systems biology*, 4(1):192. (cit. on pp. 19)

Chialvo, D. R. (2010). Emergent complex neural dynamics. *Nature Physics*, 6(10):744–750. (cit. on pp. 23 and 28)

Chubukov, V., Gerosa, L., Kochanowski, K., and Sauer, U. (2014). Coordination of microbial metabolism. *Nature Reviews Microbiology*, 12(5):327–340. (cit. on pp. 33)

Collins, J. J., Gardner, T. S., and Cantor, C. R. (2000). Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339–342. (cit. on pp. 10)

Commichau, F. M., Gunka, K., Landmann, J. J., and Stülke, J. (2008). Glutamate metabolism in bacillus subtilis: Gene expression and enzyme activities evolved to avoid futile cycles and to allow rapid responses to perturbations of the system. *Journal of Bacteriology*, 190(10):3557–3564. (cit. on pp. 124)

Costerton, J. W., Stewart, P. S., and Greenberg, E. P. (1999). Bacterial biofilms: a common cause of persistent infections. *Science (New York, N.Y.)*, 284(5418):1318–22. (cit. on pp. 101 and 103)

Cotterell, J. and Sharpe, J. (2010). An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients. *Molecular systems biology*, 6(425):425. (cit. on pp. 98)

Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J., and Palsson, B. O. (2004). Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, 429(6987):92–6. (cit. on pp. 23)

Dambre, J., Verstraeten, D., Schrauwen, B., and Massar, S. (2012). Information processing capacity of dynamical systems. *Scientific reports*, 2:514. (cit. on pp. 99)

Danial, N. N., Gramm, C. F., Scorrano, L., Zhang, C.-Y., Krauss, S., Ranger, A. M., Datta, S. R., Greenberg, M. E., Licklider, L. J., Lowell, B. B., Gygi, S. P., and Korsmeyer, S. J. (2003). BAD and glucokinase reside in a mitochondrial complex that integrates glycolysis and apoptosis. *Nature*, 424(6951):952–6. (cit. on pp. 2 and 23)

Davidich, M. I. and Bornholdt, S. (2008). Boolean network model predicts cell cycle sequence of fission yeast. *PloS one*, 3(2):e1672. (cit. on pp. 23)

Davies, D. (2003). Understanding biofilm resistance to antibacterial agents. *Nature reviews. Drug discovery*, 2(2):114–22. (cit. on pp. 101)

De Monte, S., D'Ovidio, F., Danø, S., and Sørensen, P. G. (2007). Dynamical quorum sensing: Population density encoded in cellular dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 104(47):18377–81. (cit. on pp. 21 and 22)

Detsch, C. and Stülke, J. (2003). Ammonium utilization in Bacillus subtilis: Transport and regulatory functions of NrgA and NrgB. *Microbiology*, 149(11):3289–3297. (cit. on pp. 124)

Dhar, R., Sägesser, R., Weikert, C., and Wagner, A. (2013). Yeast adapts to a changing stressful environment by evolving cross-protection and anticipatory gene regulation. *Molecular biology and evolution*, 30(3):573–88. (cit. on pp. 27 and 74)

Dobrin, R., Beg, Q. K., Barabási, A.-L., and Oltvai, Z. N. (2004). Aggregation of topological motifs in the Escherichia coli transcriptional regulatory network. *BMC bioinformatics*, 5:10. (cit. on pp. 3)

Donlan, R. M. and Costerton, J. W. (2002). Biofilms: Survival mechanisms of clinically relevant microorganisms. (cit. on pp. 101)

Dublanche, Y., Michalodimitrakis, K., Kummerer, N., Foglierini, M., and Serrano, L. (2006). Noise in transcription negative feedback loops: simulation and experimental analysis. *Mol Syst Biol*, 2:41. (cit. on pp. 52)

Duysens, L. N. M. and Amesz, J. (1957). Fluorescence Spectrophotometry of Reduced Phosphopyridine Nucleotide in Intact Cells in the Near-Ultraviolet and Visible Region. *Biophys. Acta J. Biol. Chem. J. Biol. Chem. Biochem. J. 9 j. B. S. HALDANE, Enzymes YON EULER AND K. JOSEPHSON, Z. physiol. Chem. J. Am. Chem. Soc. Biochem. Z*, 24(62):27–273. (cit. on pp. 21)

Egriboz, O., Jiang, F., and Hopper, J. E. (2011). Rapid GAL gene switch of Saccharomyces cerevisiae depends on nuclear Gal3, not nucleocytoplasmic trafficking of Gal3 and Gal80. *Genetics*, 189(3):825–836. (cit. on pp. 13 and 54)

Eldar, A. (2011). Social conflict drives the evolutionary divergence of quorum sensing. *Proceedings of the National Academy of Sciences*, 108(33):13635–13640. (cit. on pp. 101)

Eldar, A. and Elowitz, M. B. (2010). Functional roles for noise in genetic circuits. *Nature*, 467(7312):167–173. (cit. on pp. 25)

Epstein, A. K., Pokroy, B., Seminara, A., and Aizenberg, J. (2011). Bacterial biofilm shows persistent resistance to liquid wetting and gas penetration. *Proc Natl Acad Sci U S A*, 108(3):995–1000. (cit. on pp. 6)

Escalante-chong, R., Savir, Y., Carroll, S. M., Ingraham, J. B., Wang, J., and Marx, C. J. (2014). Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. 112(5). (cit. on pp. 15)

Espeland, E. M. and Wetzel, R. G. (2001). Complexation, stabilization, and UV photolysis of extracellular and surface-bound glucosidase and alkaline phosphatase: Implications for biofilm microbiota. *Microbial Ecology*, 42(4):572–585. (cit. on pp. 4)

Espinar, L., Dies, M., Cagatay, T., Süel, G. M., and Garcia-Ojalvo, J. (2013). Circuit-level input integration in bacterial gene regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 110(17):7091–6. (cit. on pp. 2)

Fauré, A., Naldi, A., Chaouiya, C., and Thieffry, D. (2006). Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics (Oxford, England)*, 22(14):e124–31. (cit. on pp. 23)

Ferrell, J. E. (2002). Self-perpetuating states in signal transduction: Positive feedback, double-negative feedback and bistability. *Current Opinion in Cell Biology*, 14(2):140–148. (cit. on pp. 11)

Ferrell, J. E. and Machleder, E. M. (1998). The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes. *Science (New York, N.Y.)*, 280(5365):895–898. (cit. on pp. 8, 9, and 10)

Ferrell, J. E. and Xiong, W. (2001). Bistability in cell signaling: How to make continuous processes discontinuous, and reversible processes irreversible. *Chaos*, 11(1):227–236. (cit. on pp. 11)

Flemming, H. C. (2002). Biofouling in water systems - Cases, causes and countermeasures. *Applied Microbiology and Biotechnology*, 59(6):629–640. (cit. on pp. 5)

Freddolino, P. L. and Tavazoie, S. (2012). Beyond homeostasis: a predictive-dynamic framework for understanding cellular behavior. *Annual review of cell and developmental biology*, 28:363–84. (cit. on pp. 26)

Gagn, C. (2012). DEAP : Evolutionary Algorithms Made Easy. *Journal of Machine Learning Research*, 13:2171–2175. (cit. on pp. 95)

Gallicchio, C. and Micheli, A. (2016). Deep Reservoir Computing: A Critical Analysis. (April):27–29. (cit. on pp. 99)

Gend, C. V., Heidelberg, D., and Kummer, U. (2001). STODE - automatic stochastic simulation of systems described by differential equations. In *Proceedings of the 2nd International Conference on Systems Biology*, pages 326–332. (cit. on pp. 58)

Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K.-K., Cheng, C., Mu, X. J., Khurana, E., Rozowsky, J., Alexander, R., Min, R., Alves, P., Abyzov, A., Addleman, N., Bhardwaj, N., Boyle, A. P., Cayting, P., Charos, A., Chen, D. Z.,

Cheng, Y., Clarke, D., Eastman, C., Euskirchen, G., Frietze, S., Fu, Y., Gertz, J., Grubert, F., Harmanci, A., Jain, P., Kasowski, M., Lacroute, P., Leng, J., Lian, J., Monahan, H., O'Geen, H., Ouyang, Z., Partridge, E. C., Patacsil, D., Pauli, F., Raha, D., Ramirez, L., Reddy, T. E., Reed, B., Shi, M., Slifer, T., Wang, J., Wu, L., Yang, X., Yip, K. Y., Zilberman-Schapira, G., Batzoglou, S., Sidow, A., Farnham, P. J., Myers, R. M., Weissman, S. M., and Snyder, M. (2012). Architecture of the human regulatory network derived from ENCODE data. *Nature*, 489(7414):91–100. (cit. on pp. 78)

Ghaemmaghami, S., Huh, W.-k., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. (2003). Global analysis of protein expression in yeast. *Nature*, 425(6959):737–41. (cit. on pp. 64)

Gilbert, P., Allison, D., and McBain, A. (2002). Biofilms in vitro and in vivo: do singular mechanisms imply cross-resistance? *Journal of Applied Microbiology*, 92(s1):98S–110S. (cit. on pp. 5)

Gillespie, D. T. (2007). Stochastic simulation of chemical kinetics. *Annual review of physical chemistry*, 58:35–55. (cit. on pp. 57)

Giniger, E., Varnum, S. M., and Ptashne, M. (1985). Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell*, 40(4):767–774. (cit. on pp. 13)

Golden, S. S., Ishiura, M., Johnson, C. H., and Kondo, T. (1997). Cyanobacterial Circadian Rhythms. *Annual review of plant physiology and plant molecular biology*, 48:327–354. (cit. on pp. 17, 73, and 98)

Grasman, J. (1984). The mathematical modeling of entrained biological oscillators. *Bulletin of Mathematical Biology*, 46(3):407–422. (cit. on pp. 21)

Gregor, T., Fujimoto, K., Masaki, N., and Sawai, S. (2010). The onset of collective behavior in social amoebae. *Science (New York, NY)*, 328(5981):1021–1025. (cit. on pp. 101)

Guidi, G. M. and Goldbeter, A. (1997). Bistability without Hysteresis in Chemical Reaction Systems: A Theoretical Analysis of Irreversible Transitions between Multiple Steady States. *The Journal of Physical Chemistry A*, 101(49):9367–9376. (cit. on pp. 10)

Gunka, K. and Commichau, F. M. (2012). Control of glutamate homeostasis in Bacillus subtilis: a complex interplay between ammonium assimilation, glutamate biosynthesis and degradation. *Molecular Microbiology*, 85(2):213–224. (cit. on pp. 105)

Hall-Stoodley, L., Costerton, J. W., and Stoodley, P. (2004). Bacterial biofilms: from the natural environment to infectious diseases. *Nature reviews. Microbiology*, 2(2):95–108. (cit. on pp. 101 and 103)

Hänggi, P., Talkner, P., and Borkovec, M. (1990). Reaction-rate theory: fifty years after Kramers. *Reviews of Modern Physics*, 62(2):251–341. (cit. on pp. 12, 47, and 50)

Hansen, N. and Ostermeier, A. (2001). Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation*, 9(2):159–195. (cit. on pp. 95 and 96)

Hartwell, L. H., Hopfield, J. J., Leibler, S., and Murray, A. W. (1999). From molecular to modular cell biology. *Nature*, 402(6761 Suppl):C47–C52. (cit. on pp. 2)

Hawkins, K. M. and Smolke, C. D. (2006). The regulatory roles of the galactose permease and kinase in the induction response of the GAL network in Saccharomyces cerevisiae. *Journal of Biological Chemistry*, 281(19):13485–13492. (cit. on pp. 52)

Helikar, T., Konvalina, J., Heidel, J., and Rogers, J. a. (2008). Emergent decision-making in biological signal transduction networks. *Proceedings of the National Academy of Sciences of the United States of America*, 105(6):1913–8. (cit. on pp. 23 and 24)

Hennessey, T. M., Rucker, W. B., and McDiarmid, C. G. (1979). Classical conditioning in paramecia. *Animal Learning & Behavior*, 7(4):417–423. (cit. on pp. 99)

Hibbing, M. E., Fuqua, C., Parsek, M. R., and Peterson, S. B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nature reviews. Microbiology*, 8(1):15–25. (cit. on pp. 3 and 101)

Høiby, N., Bjarnsholt, T., Givskov, M., Molin, S., and Ciofu, O. (2010). Antibiotic resistance of bacterial biofilms. *International journal of antimicrobial agents*, 35(4):322–32. (cit. on pp. 5 and 122)

Horak, J. and Wolf, D. H. (1997). Catabolite inactivation of the galactose transporter in the yeast Saccharomyces cerevisiae: ubiquitination, endocytosis, and degradation in the vacuole. *Journal of bacteriology*, 179(5):1541–9. (cit. on pp. 14)

Isaacs, F. J., Hasty, J., Cantor, C. R., and Collins, J. J. (2003). Prediction and measurement of an autoregulatory genetic module. *Proceedings of the National Academy of Sciences of the United States of America*, 100(13):7714–7719. (cit. on pp. 7)

Jaeger, H. (2001). The "echo state" approach to analysing and training recurrent neural networks-with an erratum note'. Technical report, Fraunhofer Institute for Autonomous Intelligent Systems. (cit. on pp. 29, 75, and 83)

Jaeger, H. (2002a). Adaptive Nonlinear System Identification with Echo State Networks. *Advances in neural information processing systems*, 4:593–600. (cit. on pp. 85)

Jaeger, H. (2002b). *Short term memory in echo state networks*. (cit. on pp. 89)

Jayakumar, A., Schulman, I., MacNeil, D., and Barnes, E. M. (1986). Role of the Escherichia coli glnALG operon in regulation of ammonium transport. *Journal of bacteriology*, 166(1):281–4. (cit. on pp. 108)

Jensen, M. H. and Krishna, S. (2012). Inducing phase-locking and chaos in cellular oscillators by modulating the driving stimuli. *FEBS Letters*, 586(11):1664–1668. (cit. on pp. 21)

Johnston, M. (1987). A model fungal gene regulatory mechanism: the GAL genes of Saccharomyces cerevisiae. *Microbiological reviews*, 51(4):458–76. (cit. on pp. 52)

Johnston, M., Flick, J. S., and Pextont, T. (1994). Multiple Mechanisms Provide Rapid and Stringent Glucose Repression of GAL Gene Expression in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, 14(6):3834–3841. (cit. on pp. 61)

Jones, B., Stekel, D., Rowe, J., and Fernando, C. (2007). Is there a Liquid State Machine in the Bacterium Escherichia Coli? In *2007 IEEE Symposium on Artificial Life*, pages 187–191. IEEE. (cit. on pp. 98)

Kalisky, T. and Quake, S. R. (2011). Single-cell genomics. *Nature methods*, 8(4):311–4. (cit. on pp. 25)

Kellershohn, N. and Laurent, M. (2001). Prion diseases: dynamics of the infection and properties of the bistable transition. *Biophysical journal*, 81(November):2517–2529. (cit. on pp. 10)

Keseler, I. M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Muñiz-Rascado, L., Bonavides-Martinez, C., Paley, S., Krummenacker, M., Altman, T., Kaipa, P., Spaulding, A., Pacheco, J., Latendresse, M., Fulcher, C., Sarker, M., Shearer, A. G., Mackie, A., Paulsen, I., Gunsalus, R. P., and Karp, P. D. (2011). EcoCyc: a comprehensive database of Escherichia coli biology. *Nucleic acids research*, 39(Database issue):D583–90. (cit. on pp. 76 and 90)

Kim, M., Zhang, Z., Okano, H., Yan, D., Groisman, A., and Hwa, T. (2012). Need-based activation of ammonium uptake in Escherichia coli. *Molecular systems biology*, 8(616):616. (cit. on pp. 108 and 116)

Kleiner, D. (1985). Bacterial ammonium transport. *FEMS Microbiology Letters*, 32(2):87–100. (cit. on pp. 108)

Klemm, K. and Bornholdt, S. (2005). Topology of biological networks and reliability of information processing. *Proceedings of the National Academy of Sciences of the United States of America*, 102(51):18414–9. (cit. on pp. 23)

Kramers, H. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304. (cit. on pp. 47)

Kubo, R. (1966). The fluctuation-dissipation theorem. *Reports on Progress in Physics*, 29(1):306. (cit. on pp. 48)

Laurent, M. and Kellershohn, N. (1999). Multistability: A major means of differentiation and evolution in biological systems. *Trends in Biochemical Sciences*, 24(11):418–422. (cit. on pp. 10)

Le Magrex-Debar, E., Lemoine, J., Gellé, M. P., Jacquelin, L. F., and Choisy, C. (2000). Evaluation of biohazards in dehydrated biofilms on foodstuff packaging. *International Journal of Food Microbiology*, 55(1-3):239–243. (cit. on pp. 4)

Legewie, S., Blüthgen, N., and Herzel, H. (2006). Mathematical modeling identifies inhibitors of apoptosis as mediators of positive feedback and bistability. *PLoS Computational Biology*, 2(9):1061–1073. (cit. on pp. 10)

Leid, J. G., Shirtliff, M. E., Costerton, J. W., and Stoodley, P. (2002). Human leukocytes adhere to, penetrate, and respond to *Staphylococcus aureus* biofilms. *Infection and immunity*, 70(11):6339–6345. (cit. on pp. 4)

Lenski, R. E. and Travisano, M. (1994). Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences of the United States of America*, 91(15):6808–14. (cit. on pp. 25)

Lev Bar-Or, R., Maya, R., Segel, L. A., Alon, U., Levine, A. J., and Oren, M. (2000). Generation of oscillations by the p53-Mdm2 feedback loop: a theoretical and experimental study. *Proceedings of the National Academy of Sciences of the United States of America*, 97(21):11250–11255. (cit. on pp. 18)

Lewis, J. (2003). Autoinhibition with transcriptional delay: A simple mechanism for the zebrafish somitogenesis oscillator. *Current Biology*, 13(16):1398–1408. (cit. on pp. 17)

Lidstrom, M. E. and Konopka, M. C. (2010). The role of physiological heterogeneity in microbial population behavior. *Nature chemical biology*, 6(10):705–712. (cit. on pp. 25)

Liu, J., Prindle, A., Humphries, J., Gabalda-Sagarra, M., Asally, M., Lee, D.-y. D., Ly, S., Garcia-Ojalvo, J., and Süel, G. M. (2015). Metabolic co-dependence gives rise to collective oscillations within biofilms. *Nature*, 523(7562):550–4. (cit. on pp. 104, 107, 123, 125, 126, 129, and 138)

Locke, J. C. W. and Elowitz, M. B. (2009). Using movies to analyse gene circuit dynamics in single cells. *Nature reviews. Microbiology*, 7(5):383–92. (cit. on pp. 25)

Lohr, D., Venkov, P., and Zlatanova, J. (1995). Transcriptional regulation in the yeast GAL gene family: a complex genetic network. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 9(9):777–87. (cit. on pp. 13, 34, 40, and 64)

Lukoševičius, M. and Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149. (cit. on pp. 30, 76, and 83)

Lynch, A. S. and Robertson, G. T. (2008). Bacterial and Fungal Biofilm Infections. *Annual Review of Medicine*, 59(1):415–428. (cit. on pp. 5)

Ma, L., Wagner, J., Rice, J. J., Hu, W., Levine, A. J., and Stolovitzky, G. A. (2005). A plausible model for the digital response of p53 to DNA damage. *Proceedings of the*

National Academy of Sciences of the United States of America, 102(40):14266–71. (cit. on pp. 18)

Maass, W., Natschlager, T., and Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.*, 14(11):2531–2560. (cit. on pp. 29 and 75)

Mah, T. F. and O'Toole, G. A. (2001). Mechanisms of biofilm resistance to antimicrobial agents. *Trends in microbiology*, 9(1):34–9. (cit. on pp. 5 and 122)

Mattick, J. S. (2001). Non-coding RNAs: the architects of eukaryotic complexity. *EMBO reports*, 2(11):986–91. (cit. on pp. 28 and 74)

Mattila-Sandholm, T. and Wirtanen, G. (1992). Biofilm formation in the industry: A review. *Food Reviews International*, 8(4):573–603. (cit. on pp. 5)

McAdams, H. H. and Shapiro, L. (2003). A bacterial cell-cycle regulatory network operating in time and space. *Science (New York, N.Y.)*, 301(5641):1874–7. (cit. on pp. 2)

McGregor, S., Vasas, V., Husbands, P., and Fernando, C. (2012). Evolution of associative learning in chemical networks. *PLoS computational biology*, 8(11):e1002739. (cit. on pp. 98)

McNeill, K. and Hamilton, I. R. (2003). Acid tolerance response of biofilm cells of Streptococcus mutans. *FEMS Microbiology Letters*, 221(1):25–30. (cit. on pp. 4)

Mendoza, L., Thieffry, D., and Alvarez-Buylla, E. R. (1998). Genetic control of flower morphogenesis in Arabidopsis thaliana: a logical analysis. *Bioinformatics (Oxford, England)*, 15(7-8):593–606. (cit. on pp. 23)

Mitchell, A., Romano, G. H., Groisman, B., Yona, A., Dekel, E., Kupiec, M., Dahan, O., and Pilpel, Y. (2009). Adaptive prediction of environmental changes by microorganisms. *Nature*, 460(7252):220–4. (cit. on pp. 74 and 98)

Mitrophanov, A. Y. and Groisman, E. A. (2008). Positive feedback in cellular control systems. *BioEssays*, 30(6):542–555. (cit. on pp. 7)

Monk, N. A. (2003). Oscillatory Expression of Hes1, p53, and NF-κB Driven by Transcriptional Time Delays. *Current Biology*, 13(16):1409–1413. (cit. on pp. 17, 18, and 19)

Mori, T. and Johnson, C. H. (2001). Circadian programming in cyanobacteria. *Seminars in cell & developmental biology*, 12(4):271–8. (cit. on pp. 73 and 98)

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 5(7):621–628. (cit. on pp. 63)

Nakano, M. M., Yang, F., Hardin, P., and Zuber, P. (1995). Nitrogen regulation of nasA and the nasB operon, which encode genes required for nitrate assimilation in Bacillus subtilis. *Journal of bacteriology*, 177(3):573–9. (cit. on pp. 105 and 107)

Nelson, D. E., Ihekwaba, A. E. C., Elliott, M., Johnson, J. R., Kell, D. B., and White, M. R. H. (2004). Oscillations in NF- . B Signaling Control the Dynamics of Gene Expression. 306(October):704–709. (cit. on pp. 19)

Novák, B. and Tyson, J. J. (1993). Numerical analysis of a comprehensive model of M-phase control in Xenopus oocyte extracts and intact embryos. *Journal of Cell Science*, 106:1153–1168. (cit. on pp. 10)

Novák, B. and Tyson, J. J. (2008). Design principles of biochemical oscillators. *Nature reviews. Molecular cell biology*, 9(12):981–991. (cit. on pp. 7, 17, 19, and 20)

Oliveira, N. M., Niehus, R., and Foster, K. R. (2014). Evolutionary limits to cooperation in microbial communities. *Proceedings of the National Academy of Sciences of the United States of America*, 111(50):17941–6. (cit. on pp. 3 and 101)

Olson, B., Hashmi, I., Molloy, K., and Shehu, A. (2012). Basin Hopping as a General and Versatile Optimization Framework for the Characterization of Biological Macromolecules. *Advances in Artificial Intelligence*, 2012:1–19. (cit. on pp. 39)

Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L., and Adhya, S. (2005). Switches in bacteriophage lambda development. *Annual review of genetics*, 39:409–429. (cit. on pp. 10)

Peng, G. and Hopper, J. E. (2000). Evidence for Gal3p ' s Cytoplasmic Location and Gal80p ' s Dual Cytoplasmic-Nuclear Location Implicates New Mechanisms for Controlling Gal4p Activity in Saccharomyces cerevisiae Evidence for Gal3p ' s Cytoplasmic Location and Gal80p ' s Dual Cytoplasmic-N. 20(14):5140–5148. (cit. on pp. 54)

Peng, W., Liu, P., Xue, Y., and Acar, M. (2015). Evolution of gene network activity by tuning the strength of negative-feedback regulation. *Nature communications*, 6:6226. (cit. on pp. 15 and 70)

Phillips, P. C. (2008). Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, 9(11):855–867. (cit. on pp. 35)

Platt, A. and Reece, R. J. (1998). The yeast galactose genetic switch is mediated by the formation of a Gal4p-Gal80p-Gal3p complex. *EMBO Journal*, 17(14):4086–4091. (cit. on pp. 13)

Pomerening, J. R., Sun, Y. K., and Ferrell, J. E. (2005). Systems-level dissection of the cell-cycle oscillator: Bypassing positive feedback produces damped oscillations. *Cell*, 122(4):565–578. (cit. on pp. 20)

Raj, A. and van Oudenaarden, A. (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*, 135(2):216–226. (cit. on pp. 25)

Ramsey, S. a., Smith, J. J., Orrell, D., Marelli, M., Petersen, T. W., de Atauri, P., Bolouri, H., and Aitchison, J. D. (2006). Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast. *Nature genetics*, 38(9):1082–7. (cit. on pp. 34 and 52)

Rao, C. V. and Arkin, A. P. (2003). Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *Journal of Chemical Physics*, 118(11):4999–5010. (cit. on pp. 58)

Raser, J. M. (2005). Noise in Gene Expression: Origins, Consequences, and Control. *Science*, 309(5743):2010–2013. (cit. on pp. 25)

Rasmussen, B. (2000). Filamentous microfossils in a 3,235-million-year-old volcanogenic massive sulphide deposit. *Nature*, 405(6787):676–679. (cit. on pp. 4)

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science (New York, N.Y.)*, 297(5586):1551–5. (cit. on pp. 3)

Richard, P. (1996). Sustained Oscillations in Free-energy State and Hexose Phosphates in Yeast. *Yeast*, 0061(June):731–740. (cit. on pp. 21)

Richard, P. (2003). The rhythm of yeast. *FEMS Microbiology Reviews*, 27(4):547–557. (cit. on pp. 17)

Rodan, A. and Tino, P. (2011). Minimum complexity echo state network. *IEEE transactions on neural networks*, 22(1):131–44. (cit. on pp. 78 and 87)

Roenneberg, T., Dragovic, Z., and Merrow, M. (2005). Demasking biological oscillators: properties and principles of entrainment exemplified by the Neurospora circadian clock. *Proceedings of the National Academy of Sciences of the United States of America*, 102(21):7742–7. (cit. on pp. 21)

Rondina, G. G. and Da Silva, J. L. F. (2013). Revised basin-hopping monte carlo algorithm for structure optimization of clusters and nanoparticles. *Journal of Chemical Information and Modeling*, 53(9):2282–2298. (cit. on pp. 39)

Rosenfeld, N., Elowitz, M. B., and Alon, U. (2002). Negative autoregulation speeds the response times of transcription networks. *Journal of Molecular Biology*, 323(5):785–793. (cit. on pp. 7)

Rossi, G. and Ferrando, R. (2009). Searching for low-energy structures of nanoparticles: a comparison of different methods and algorithms. *Journal of physics. Condensed matter : an Institute of Physics journal*, 21(8):084208. (cit. on pp. 39)

Roy, S., Ernst, J., Kharchenko, P. V., Kheradpour, P., Negre, N., Eaton, M. L., Landolin, J. M., Bristow, C. a., Ma, L., Lin, M. F., Washietl, S., Arshinoff, B. I., Ay, F., Meyer, P. E., Robine, N., Washington, N. L., Di Stefano, L., Berezikov, E., Brown, C. D., Candeias, R., Carlson, J. W., Carr, A., Jungreis, I., Marbach, D., Sealfon, R.,

Tolstorukov, M. Y., Will, S., Alekseyenko, A. a., Artieri, C., Booth, B. W., Brooks, A. N., Dai, Q., Davis, C. a., Duff, M. O., Feng, X., Gorchakov, A. a., Gu, T., Henikoff, J. G., Kapranov, P., Li, R., MacAlpine, H. K., Malone, J., Minoda, A., Nordman, J., Okamura, K., Perry, M., Powell, S. K., Riddle, N. C., Sakai, A., Samsonova, A., Sandler, J. E., Schwartz, Y. B., Sher, N., Spokony, R., Sturgill, D., van Baren, M., Wan, K. H., Yang, L., Yu, C., Feingold, E., Good, P., Guyer, M., Lowdon, R., Ahmad, K., Andrews, J., Berger, B., Brenner, S. E., Brent, M. R., Cherbas, L., Elgin, S. C. R., Gingeras, T. R., Grossman, R., Hoskins, R. a., Kaufman, T. C., Kent, W., Kuroda, M. I., Orr-Weaver, T., Perrimon, N., Pirrotta, V., Posakony, J. W., Ren, B., Russell, S., Cherbas, P., Graveley, B. R., Lewis, S., Micklem, G., Oliver, B., Park, P. J., Celniker, S. E., Henikoff, S., Karpen, G. H., Lai, E. C., MacAlpine, D. M., Stein, L. D., White, K. P., and Kellis, M. (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science (New York, N.Y.)*, 330(6012):1787–97. (cit. on pp. 76)

Rué, P., Pons, a. J., Domedel-Puig, N., and García-Ojalvo, J. (2010). Relaxation dynamics and frequency response of a noisy cell signaling network. *Chaos (Woodbury, N.Y.)*, 20(4):045110. (cit. on pp. 23 and 24)

Saez-Rodriguez, J., Simeoni, L., Lindquist, J. A., Hemenway, R., Bommhardt, U., Arndt, B., Haus, U.-U., Weismantel, R., Gilles, E. D., Klamt, S., and Schraven, B. (2007). A logical model provides insights into T cell receptor signaling. *PLoS computational biology*, 3(8):e163. (cit. on pp. 24)

Saigusa, T., Tero, A., Nakagaki, T., and Kuramoto, Y. (2008). Amoebae anticipate periodic events. *Physical Review Letters*, 100(1):1–4. (cit. on pp. 99)

Samaga, R., Saez-Rodriguez, J., Alexopoulos, L. G., Sorger, P. K., and Klamt, S. (2009). The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. *PLoS computational biology*, 5(8):e1000438. (cit. on pp. 24)

Sanft, K. R., Gillespie, D. T., and Petzold, L. R. (2011). Legitimacy of the stochastic Michaelis-Menten approximation. *IET systems biology*, 5(1):58. (cit. on pp. 58)

Santillán, M., Mackey, M., and Zeron, E. (2007). Origin of Bistability in the lac Operon. *Biophysical Journal*, 92(11):3830–3842. (cit. on pp. 10)

Schild, S., Tamayo, R., Nelson, E. J., Qadri, F., Calderwood, S. B., and Camilli, A. (2007). Genes Induced Late in Infection Increase Fitness of Vibrio cholerae after Release into the Environment. *Cell Host and Microbe*, 2(4):264–277. (cit. on pp. 74)

Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. *Nature genetics*, 31(1):64–8. (cit. on pp. 98)

Sierro, N., Makita, Y., de Hoon, M., and Nakai, K. (2008). DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information. *Nucleic acids research*, 36(Database issue):D93–6. (cit. on pp. 76)

Simon, I., Barnett, J., Hannett, N., Harbison, C. T., Rinaldi, N. J., Volkert, T. L., Wyrick, J. J., Zeitlinger, J., Gifford, D. K., Jaakkola, T. S., and Young, R. A. (2001). Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell*, 106(6):697–708. (cit. on pp. 2)

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., and Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nature reviews. Genetics*, 15(2):121–32. (cit. on pp. 63)

Sorek, M., Balaban, N. Q., and Loewenstein, Y. (2013). Stochasticity, bistability and the wisdom of crowds: a model for associative learning in genetic regulatory networks. *PLoS computational biology*, 9(8):e1003179. (cit. on pp. 27 and 98)

Stannek, L., Thiele, M. J., Ischebeck, T., Gunka, K., Hammer, E., Völker, U., and Commichau, F. M. (2015). Evidence for synergistic control of glutamate biosynthesis by glutamate dehydrogenases and glutamate in Bacillus subtilis. *Environmental microbiology*, 17(9):3379–90. (cit. on pp. 105)

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., and Gilles, E. D. (2002). Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912):190–3. (cit. on pp. 28 and 74)

Stewart, P. S. (2002). Mechanisms of antibiotic resistance in bacterial biofilms. *International journal of medical microbiology : IJMM*, 292(2):107–13. (cit. on pp. 5 and 122)

Stewart, P. S. and Costerton, J. W. (2001). Antibiotic resistance of bacteria in biofilms. *Lancet*, 358(9276):135–138. (cit. on pp. 4)

Stockwell, S. R., Landry, C. R., and Rifkin, S. a. (2015). The yeast galactose network as a quantitative model for cellular memory. *Mol. BioSyst.*, 11(1):28–37. (cit. on pp. 14 and 34)

Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S., and Hasty, J. (2008). A fast, robust and tunable synthetic gene oscillator. *Nature*, 456(7221):516–519. (cit. on pp. 20)

Swain, P. S., Elowitz, M. B., and Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12795–800. (cit. on pp. 25)

Tagkopoulos, I., Liu, Y.-C., and Tavazoie, S. (2008). Predictive behavior within microbial genetic networks. *Science*, 320(5881):1313–7. (cit. on pp. 27, 74, and 98)

Tanaka, G., Nakane, R., Yamane, T., Nakano, D., Takeda, S., Nakagawa, S., and Hirose, A. (2016). Exploiting Heterogeneous Units for Reservoir Computing with Simple Architecture. volume 1, pages 187–194. (cit. on pp. 99)

Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X. S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule

sensitivity in single cells. *Science (New York, N.Y.)*, 329(5991):533–8. (cit. on pp. 25)

Teitzel, G. and Parsek, M. (2003). Heavy metal resistance of biofilm and planktonic *Pseudomonas aeruginosa*. *Appl. Environ. Microbiol.*, 69(4):2313–2320. (cit. on pp. 4)

Teixeira, M. C., Monteiro, P. T., Guerreiro, J. F., Gonçalves, J. P., Mira, N. P., Dos Santos, S. C., Cabrito, T. R., Palma, M., Costa, C., Francisco, A. P., Madeira, S. C., Oliveira, A. L., Freitas, A. T., and Sá-Correia, I. (2014). The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in Saccharomyces cerevisiae. *Nucleic acids research*, 42(1):D161–6. (cit. on pp. 76)

Tian, T. and Burrage, K. (2004). Bistability and switching in the lysis/lysogeny genetic regulatory network of bacteriophage λ. *Journal of Theoretical Biology*, 227(2):229–237. (cit. on pp. 10)

Tyson, J. J., Hong, C. I., Thron, C. D., and Novak, B. (1999). A simple model of circadian rhythms based on dimerization and proteolysis of PER and TIM. *Biophysical journal*, 77(5):2411–7. (cit. on pp. 17 and 18)

Venturelli, O. S., El-Samad, H., and Murray, R. M. (2012). Synergistic dual positive feedback loops established by molecular sequestration generate robust bimodal response. *Proceedings of the National Academy of Sciences of the United States of America*, 109(48):E3324–33. (cit. on pp. 15, 34, 52, 53, 60, and 61)

Venturelli, O. S., Zuleta, I., Murray, R. M., and El-Samad, H. (2015). Population diversification in a yeast metabolic program promotes anticipation of environmental shifts. *PLoS biology*, 13(1):e1002042. (cit. on pp. 15)

Verstraeten, D., Schrauwen, B., D'Haene, M., and Stroobandt, D. (2007). An experimental unification of reservoir computing methods. *Neural networks*, 20(3):391–403. (cit. on pp. 30 and 31)

Vlamakis, H., Aguilar, C., Losick, R., and Kolter, R. (2008). Control of cell fate by the formation of an architecturally complex bacterial community. *Genes and Development*, 22(7):945–953. (cit. on pp. 101)

Wales, D. and Doye, J. P. K. (1997). Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *Journal of Physical Chemistry A*, 101(97):5111–5116. (cit. on pp. 39)

Wales, D. J. (2010). Energy Landscapes and Structure Prediction Using Basin-Hopping. In *Modern Methods of Crystal Structure Prediction*, pages 29–54. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. (cit. on pp. 39)

Wall, M. E., Hlavacek, W. S., and Savageau, M. a. (2004). Design of gene circuits: lessons from bacteria. *Nature reviews. Genetics*, 5(1):34–42. (cit. on pp. 2)

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10(1):57–63. (cit. on pp. 63)

Watnick, P. and Kolter, R. (2000). Biofilm, city of microbes. *Journal of Bacteriology*, 182(10):2675–2679. (cit. on pp. 5)

Webb, J. S., Givskov, M., and Kjelleberg, S. (2003). Bacterial biofilms: Prokaryotic adventures in multicellularity. *Current Opinion in Microbiology*, 6(6):578–585. (cit. on pp. 3 and 4)

Westall, F., De Wit, M. J., Dann, J., Van der Gaast, S., De Ronde, C. E. J., and Gerneke, D. (2001). Early archean fossil bacteria and biofilms in hydrothermally-influenced sediments from the Barberton greenstone belt, South Africa. *Precambrian Research*, 106(1-2):93–116. (cit. on pp. 4)

Wightman, R., Bell, R., and Reece, R. J. (2008). Localization and interaction of the proteins constituting the GAL genetic switch in Saccharomyces cerevisiae. *Eukaryotic Cell*, 7(12):2061–2068. (cit. on pp. 54)

Wilking, J. N., Zaburdaev, V., De Volder, M., Losick, R., Brenner, M. P., and Weitz, D. A. (2013). Liquid transport facilitated by channels in Bacillus subtilis biofilms. *Proceedings of the National Academy of Sciences*, 110(3):848–852. (cit. on pp. 5 and 102)

Winfree, A. T. (2002). On emerging coherence. *Science (New York, N.Y.)*, 298(5602):2336–7. (cit. on pp. 21)

Wingreen, N. S. and Levin, S. A. (2006). Cooperation among microorganisms. (cit. on pp. 101)

Wolf, D. M., Fontaine-Bodin, L., Bischofs, I., Price, G., Keasling, J., and Arkin, A. P. (2008). Memory in microbes: quantifying history-dependent behavior in a bacterium. *PloS one*, 3(2):e1700. (cit. on pp. 28 and 74)

Wyffels, F., Schrauwen, B., and Stroobandt, D. (2008). Stable Output Feedback in Reservoir Computing Using Ridge Regression. In Kůrková, V., Neruda, R., and Koutník, J., editors, *Artificial Neural Networks - ICANN 2008*, volume 5163 of *Lecture Notes in Computer Science*, pages 808–817. Springer Berlin Heidelberg, Berlin, Heidelberg. (cit. on pp. 84)

Xiong, W. and Ferrell, J. E. J. (2003). A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature*, 426(November):460–465. (cit. on pp. 8 and 10)

Yildiz, F. H. and Visick, K. L. (2009). Vibrio biofilms: so much the same yet so different. (cit. on pp. 101)

Zacharioudakis, I., Gligoris, T., and Tzamarias, D. (2007). A Yeast Catabolic Enzyme Controls Transcriptional Memory. *Current Biology*, 17(23):2041–2046. (cit. on pp. 13, 52, and 53)

Zeigler, D. R., Prágai, Z., Rodriguez, S., Chevreux, B., Muffler, A., Albert, T., Bai, R., Wyss, M., and Perkins, J. B. (2008). The origins of 168, W23, and other Bacillus subtilis legacy strains. *Journal of bacteriology*, 190(21):6983–95. (cit. on pp. 105)

Zenke, F. T., Engles, R., Vollenbroich, V., Meyer, J., Hollenberg, C. P., and Breunig, K. D. (1996). Activation of Gal4p by galactose-dependent interaction of galactokinase and Gal80p. *Science (New York, N.Y.)*, 272(5268):1662–1665. (cit. on pp. 13)