Contribution of gene duplications to the evolution of genetic networks

by

Bernardo Fabián Pando

Lic. in Physics
Universidad de Buenos Aires, 2003

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics
at the
Massachusetts Institute of Technology

June 2010

Signature of author: _____

Department of Physics
April 22, 2010

Certified by: _____

Alexander van Oudenaarden
Professor of Physics and Biology
Thesis supervisor

Accepted by: _____

Krishna Rajagopal
Professor of Physics
Associate Department Head for Education

Contribution of gene duplications to the evolution of genetic networks

by

Bernardo Fabián Pando

Submitted to the Department of Physics
on April 22, 2010 in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy in Physics

Abstract

Exploring the forces that drive evolution at the gene network level and investigating underlying principles behind this process are fundamental questions in the context of understanding how evolution shapes transcriptional circuits. In this thesis I present two different explorations along these lines with special emphasis on the contribution of gene dosage variations to the alteration of phenotypes.

On one hand I describe the design of an experimental system for observing evolution *in vivo* in the yeast *Saccharomyces cerevisiae*, the construction of a simple two-component genetic system and how I used the setup to explore its adaptative capabilities. An external inducer allowed me to tune the basal state of the system and by doing this I was able to tune the relative contribution of gene duplications and point mutations to the evolution of the system against an imposed fitness defect. This illustrates how the number of evolutionary solutions available against an imposed fitness constraint depends on the operating point of the underlying circuit.

Increasing in complexity I then describe an analysis of the effect of gene dosage variations in the context of the galactose uptake network in the same organism. This network is composed of four regulatory elements and it contains several feedback loops built into it, which makes its analysis nontrivial. The effect of dosage variations was explored experimentally by systematically deleting one of two copies of each regulatory gene in a diploid background. Surprisingly the system turned out to be invariant to proportional changes in all its regulatory elements, a property that we call network-dosage invariance. I developed a modeling framework for rationalizing these observations and found that the presence of both an activator and inhibitor interacting with a 1-to-1 stochiometry as well as certain topological constraints are requirements for such a behavior. This provides insight into what kind of regulatory circuits are robust to global effects like genomic duplications events, ploidy changes or global variations in the concentration of transcription factors.

This work could be extended to the study of more complicated circuits, allowing the systematic exploration of evolutionary properties of small scale genetic systems.

Thesis Supervisor: Alexander van Oudenaarden
Title: Professor of Physics and Biology

Para Ani

Contents

Acknowledgments

This thesis would not have been possible without the help of many people to whom I would like to extend my thanks.

First, Alexander van Oudenaarden, my thesis advisor, who welcomed me into his lab and gave me great freedom and support to pursue my research interests. He personally taught me many of the experimental techniques that made this thesis possible and he's been always available for deep scientific discussions. He created a great group at MIT that makes significant contributions in the area of quantitative biology and which I am very proud to be part of.

Many of the members of the group have contributed significantly to the work presented in this thesis. I would like to specially thank Murat Acar, a former graduate student in the lab, who led the experimental effort of the work I present in Chapter 3. During the time the project developed he was working in the lab of Michael Elowitz, to whom I am also grateful for his support and for several discussions we had that allowed that project to move forward.

Over the course of the years I collaborated with many of the members of the lab and the MIT community in general. I worked specially close to Qiong Yang with whom I collaborated on the analysis of the coupling of circadian and cell cycle clocks of the cyanobacterium *Synechococcus elongatus*. I also had the pleasure to work closely with Krishanu Saha, Jacob Hanna and Rudolf Jaenisch on the analysis of the process of induced pluripotency in ProB cells. Juan Pedraza, Benjamin Kaufmann, Jerome Mettetal, Allen Lee, Carlos Gomez Uribe, Shankar Mukherji, Jeff Gore, Rui Zhen Tan, Gregor

Neuert, Mei Lyn Ong, Jeroen van Zon, Miaoqing Fang, and Christoph Engert are some of the members of the lab with whom I interacted the most and I am grateful for all the exchange of ideas that we shared.

I would like to thank many of the MIT professors for their lectures and the knowledge they were able to share with me. I specially enjoyed lectures given by Mehran Kardar, Leonid Mirny, Scott Hughes, Alan Guth, Edward Farhi, Martin Bazant, Seth Lloyd, Peter Shor, Rodolfo Rosales and George Haller.

During my years at MIT I got the pleasure to assist in teaching the Systems Biology graduate course that Alexander van Oudenaarden developed. I found this experience most gratifying, in particular due to the fact that the course promotes interdisciplinary research and is making significant contributions in introducing more quantitative tools and ideas into the field of Biology. I would like to thank Alexander again for giving me this opportunity and all the students from diverse fields of science that decided to participate in the class making it such an enjoyable experience.

Finally I would like to thank Ani for all her support, patience and love throughout these years.

Foreword

Since the last years of my undergraduate education I've been involved in research in the areas of Biophysics and Systems Biology, contributing the intuition and quantitative and analytical skills that I acquired during my education in Physics and Mathematics to the study of biological problems.

In Alexander van Oudenaarden's lab at MIT I continued to developed these interests, gaining knowledge of experimental techniques and making contributions to different projects mainly on the quantitative and modeling side.

Shortly after joining the lab I started working on applying ideas from control theory to the analysis of the behavior of the galactose uptake network in the yeast *Saccharomyces cerevisiae*. I explored formulations of stochastic chemical kinetics in spatially distributed systems. I contributed models for describing the kinetics of reprogramming of mammalian somatic cells into states of induced pluripotency [1]. I analyzed the coupling between the circadian and cell cycle oscillators in the cyanobacterium *Synechococcus elongatus* [2]. And I also contributed models to the description of the transcriptional transition of Th0 cells into other fates.

Since a few years ago, I became interested in the dynamics of the evolution of genetic networks and I developed tools for exploring this process *in vivo* in the yeast *Saccharomyces cerevisiae*. The main questions I got interested in exploring have to do with how different properties of the evolution of a given network are related to its topological properties. In this thesis I discuss some aspects of this with special emphasis on the contribution of gene duplications as potential evolutionary mechanisms.

Chapter 1

Introduction

Exploring the forces that drive evolution at the gene network level and investigating underlying principles behind this process are fundamental questions in the context of understanding how evolution shapes transcriptional circuits [3, 4]. Do larger networks evolve functionalities faster than smaller networks? Or are they more robust to evolutionary forces? Does the topology of the network make some components more sensitive to mutational changes than others? Are circuits in which the components interact in a cascade-like fashion more likely to evolve than circuits in which each component interacts with all the others? Are there circuits in which gene duplications would be favored over point mutations as feasible solutions for overcoming some fitness defect?

These and related questions have been traditionally attacked from the point of view of comparative proteomics [5] and genomics [6, 7] and over the last decades, with the deployment of high-throughput techniques [8], progress in this front has accelerated. However, these techniques rely on the observation of signatures that evolution, as occurred in the past, imprinted in present-day samples. Approaches in which evolution is observed in a laboratory as it happens [9–12] allow one to complement the methods mentioned above and shed more light on the dynamics of evolutionary processes and the underlying driving forces behind them.

The study of adaptation of genetic networks *in vivo* provides us with

further insights into the dynamics and driving forces behind evolutionary processes and complements more traditional studies based on comparative genomics and *in silico* evolution, allowing us to test relevant hypotheses by observing the process as it happens. Microorganisms make ideal model systems for studying this process given their fast cell division timescales and the use of tools from synthetic biology and bioengineering allow one to isolate evolutionary pressure to small-scale gene networks. The application of controlled culturing techniques makes it possible to maintain the imposed pressure selectively over long periods of time and by combining these techniques it becomes feasible to observe, study and quantify the evolution of genetic networks in a controlled manner. The lessons that we will learn by applying this approach and observing evolution taking place before our eyes will deepen our understanding not only of the evolution of genetic networks but also of behavior in analogous systems in which features of networks of interacting agents get selectively enhanced over time.

There are a multitude of ways in which a given organism or genetic circuit could evolve: point mutations [13–16], gene duplications [11, 16–20], changes in ploidy [21, 22], chromosomal crossovers [13, 23] and horizontal gene transfer [24], among others. It is of interest to understand under what situations some of these events will be more likely to produce an evolutionary advantage to some organism and in this thesis I discuss some aspects of the question of the contribution of gene duplications to the evolution of genetic networks using a combination of experimental evolution in the yeast *S. cerevisiae*, molecular techniques and mathematical modeling. In Chapter 2 I present the design of a simple synthetic genetic system and the exploration of the ways in which it adapted to an imposed fitness constraint, with special emphasis on the effect of gene duplications as viable evolutionary solutions. And, increasing in complexity, in Chapter 3 I discuss the effects of gene dosage in the context of a more complex genetic network, the galactose uptake system, paying particular attention to the question of network-dosage invariance, that is: under which situations a phenotype produced by a genetic system will be robust to proportional changes in the dosage of all the genes involved. These observations provide a starting point for future studies of the evolution of genetic circuits and the contribution of different mechanisms to the shaping of transcriptional circuits through evolutionary forces.

Chapter 2

Gene duplications in the adaptation of a small synthetic gene circuit

2.1 Introduction

Point mutations [13–16] and gene duplications [11, 16–20] are two of the major mechanisms that drive the evolution of genetic networks but how these processes determine the dynamics of adaptation is poorly understood. Here I present an exploration of the relative contributions of these two mechanisms to the adaptation dynamics of a synthetic gene circuit in the budding yeast *Saccharomyces cerevisiae*, an asexual model system, using an experimental evolution approach [9, 10, 12, 25–30].

In this circuit a synthetic transcriptional activator, regulated by an extracellular inducer, drives the expression of an essential gene involved in uracil synthesis. In the absence of inducer, cells do not produce enough uracil which results in a growth deficit with respect to wildtype strains. Nevertheless, yeast populations cultured continually in this environment irreversibly adapt, approaching wildtype growth rates after a few days.

We found that a narrow spectrum of point mutations in the transcriptional activator explains this recovery and that many mutants recover by inverting the logic of the transcriptional activator. In the presence of a low inducer concentration we observed an increase in the effective adaptation rate and in addition to a similar mutation spectrum we found gene duplication events of the transcriptional activator, providing a plausible explanation for the faster adaptation dynamics.

Our work suggests that the effective rate of generation of fitter phenotypes is determined by a combination of the point mutation and gene duplication rates and that their relative contribution is strongly dependent on the coupling between genes in the network. This provides a starting point for unraveling the relative contributions of point mutations and gene duplications during the evolution of gene networks.

2.2 Construction and characterization of the genetic system

In order to explore the effect of different evolutionary mechanisms on a genetic network we decided to use a very simple engineered genetic model so that we could have most of the system under our control. The idea was to create a system of just two genes in which one of them, constitutively expressed, would enhance production of the other. We wanted to be able to tune the coupling between the two genes and also to tightly couple the expression level of the downstream gene to growth so that it would have a significant impact on fitness, allowing us to obtain fixation timescales compatible with times accessible in laboratory setups.

A plasmid derived from pRS402 [31, 32] containing the *MYO2* promoter driving the *rtTA*(*S2*) gene followed by the *CYC1* transcriptional terminator [33] was integrated into the genomic *ade2* locus of the *Saccharomyces cerevisiae* W303 mat a strain by homologous recombination and selection in media lacking adenine. A transformant consisting of a single integration was selected by Southern Blot analysis yielding an strain that we named BP82.3. Constructs consisting of the *TET07* promoter driving the *URA3* gene from

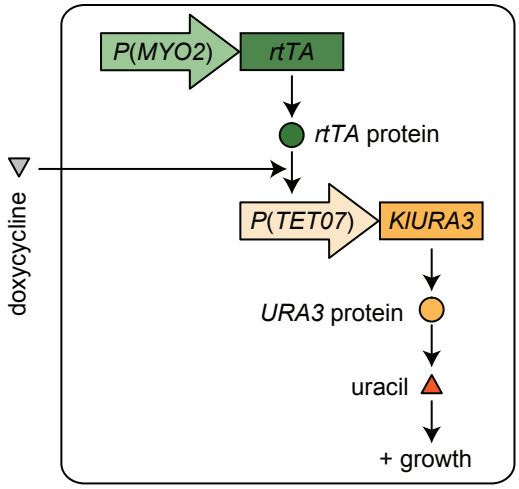

Figure 2.1: Schematic of the underlying genetic system.

the related yeast *Kluyveromyces lactis* plus the *ADH1* terminator and the *TEF* promoter driving the *KanMX* gene plus the *TEF* terminator were cloned into *Escherichia coli* vectors and then extracted and fused by PCR including, at both ends, 50 base pairs of homology to the W303 sequences harboring the *URA3* coding region. This construct was incorporated into the BP82.3 strain and positive transformants were selected under the presence of Geneticin. This effectively replaced the original *URA3* coding region by the designed sequence. All integrations were verified by PCR.

The *MYO2* promoter is constitutively active in *Saccharomyces cerevisiae* and therefore our engineered strain produces the *rtTA* transcriptional activator (a fusion of a modified version of the *TET* repressor and the *VP16* transcriptional activation moeity [34]) at a constant rate. Doxycycline, when bound to the *rtTA* protein, induces a conformational change that promotes binding to the *TET07* promoter where the *VP16* moiety enhances transcription of the downstream *KlURA3* sequence. As the endogenous *URA3* gene had been deleted the *Kluyveromyces lactis URA3* mRNA (*KlURA3*) is the sole source for Ura3 protein, orotidine 5'-phosphate decarboxylase (ODCase), an enzyme necessary for catalyzing the synthesis of uracil, a nucleic acid needed for normal growth (Figure 2.1). Incidentally this protein is one

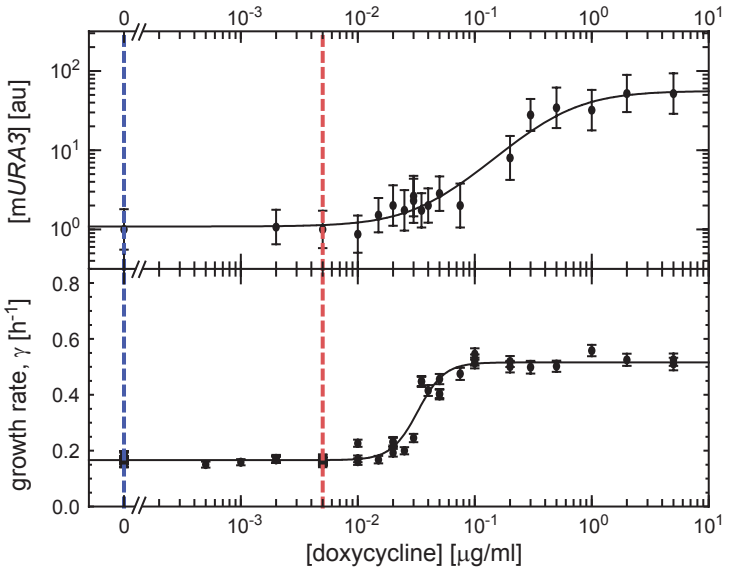

Figure 2.2: *URA3* transcript level as measured by qRT-PCR (top panel) and culture growth rate γ (bottom panel), both measured after 24 hours in media without uracil and different doxycycline concentrations. The solid lines represent fits to Hill functions as guides to the eye. The dashed vertical lines indicate the doxycycline concentrations at which the two different sets of adaptation experiments described in Section 2.3.2 were performed.

of the most proficient enzymes in nature, enhancing the reaction rate of the underlying reaction by a factor of about 10^{17} [35].

We found that the *KlURA3 mRNA* expression levels could be tuned over about a 60-fold range (Figure 2.2, top panel). In the absence of doxycycline a basal transcription is detected which provides enough Ura3 proteins for the cells to survive in media without uracil. However the growth rate is significantly lower than that observed at high concentrations of doxycycline where cultures approach growth rates typical of wildtype yeast (Figure 2.2, bottom panel).

2.3 Experimental adaptation of the genetic system to selective media

To directly observe the dynamics of adaptation we continuously monitored the growth rate γ of many independent cultures that were maintained at a constant population size of $N = (1.7 \pm 0.1)\,10^7$cells (mean $\pm$ s.e., $n = 51$) in environments with low concentrations of doxycycline where the cells experience a strong selective pressure with respect to organisms that would reproduce at wildtype growth rates.

2.3.1 Experimental setup

We performed the experiments using turbidostats [36, 37] (Figure 2.3). In these setups, cells are maintained at a constant optical density in liquid culture by regulating the dilution rate in response to an average instantaneous proxy for culture growth rate. To achieve this we first established a system for continuous measurement of the culture's relative absorption coefficient by using an ultraviolet LED - photodetector pair. In this way the electronic output of the photodetector ($v(t)$) is correlated to the optical density of the culture. When $v(t)$ exceeds a pre-set threshold a pump is activated which dilutes the culture back below the threshold. A pump constantly acting on an exhaust line set at a certain level ensures that the volume of the culture stays fixed throughout the experiment.

By recording the activity of the computer controlled pump over time, $a(t)$, we can accurately calculate the population's instantaneous growth rate, $\gamma(t)$. We did this by measuring the fraction of time the pump was actively providing fresh media during each hour-long interval and converting this into an average instantaneous pump flow rate, $\langle p(t) \rangle$, by multiplying by the pump's maximal possible flow rate. This raw pump activity is then converted into a growth rate by normalizing with the volume of the turbidostat culture using the formula, $\gamma(t) = {}^{\langle p(t) \rangle}/_{V}$ where V represents the culture volume as measured at the end of the experiment. For this work we constructed a system consisting of a computer controlling 8 chambers in parallel using simple electronics and a custom-built LabView program to control the different devices.

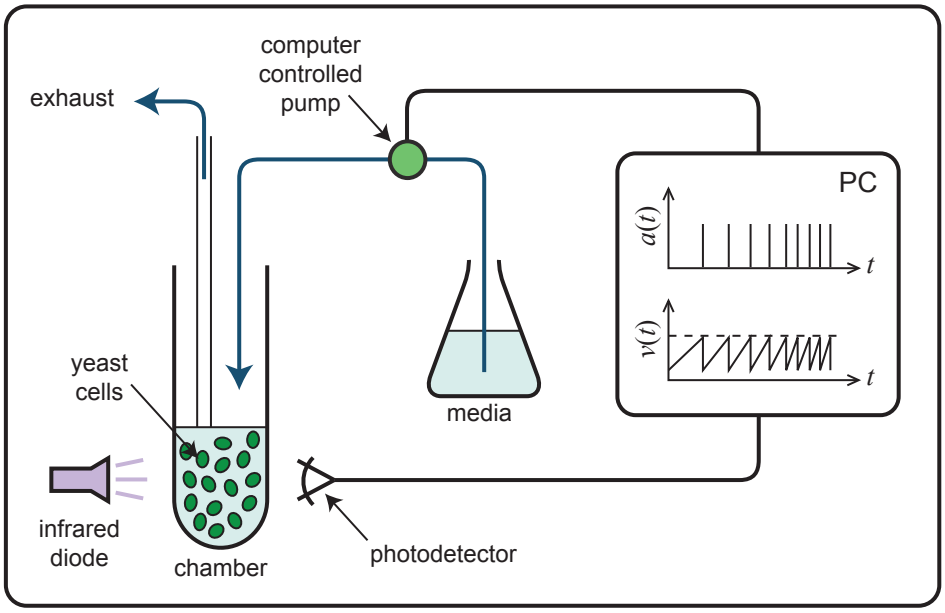

Figure 2.3: Turbidostat setup schematic. $v(t)$ represents the signal on the photodector, which is related to the optical density of the culture in the chamber. The activity of the controlling pump is denoted by $a(t)$. In the schematic we present a caricature of a situation in which the growth rate of the culture is increasing with respect to time.

We chose this experimental system so that, on one hand, we could keep the density of the culture fixed throughout the experiment, ensuring that the selective pressure was applied on the output of the genetic system of interest and that it was not drifting towards other systems like, for instance, response to culture overcrowding as might occur in chemostats. In other words, having a system in which cells can grow at a constant density irrespectively of their growth rate allows one to ignore potentially confounding density-dependent effects. On the other hand the use of turbidostats allows one to obtain measurements of growth rate dynamics with high temporal resolution, allowing for better quantification of observations.

We grew liquid cultures in synthetic dropout media with appropriate amino-acid supplements (adenine and methionine were not included in any

experiment as the constructed strains were able to synthesize these, and in some experiments, as indicated, uracil was also left out), 2 % $^{w}/_{v}$ glucose as the carbon source and different doxycycline concentrations. Before transferring them to the turbidostat chambers, each culture was grown overnight in a shaker at 30 °C in a 10 ml volume starting with a low cell density[1] so that just before the transfer into the turbidostat they had not reached stationary phase ($OD_{600} < 2$). Next, cultures were washed with their prospective in-turbidostat media and transferred to turbidostat chambers. The turbidostat maintains the culture at a constant volume and, after a transient stabilization period, constant optical density levels ($0.05 < OD_{600} < 0.4$) where cells do not experience the effect of nutrients depletion. By determining the volume of the chamber and the optical density at the end of the experiment as well as a an experimentally measured conversion factor between these two quantities[2] we estimated the size of the population N that was cultured in each chamber.

2.3.2 Cultures adapt to the selective environment

Figure 2.4a displays the growth dynamics of 26 independent populations that were cultured in the absence of uracil and doxycycline over a period of $\tilde{6}$ days. At $t = 0$ the turbidostats were seeded with an exponentially dividing population that had been grown in media with uracil ($\gamma_\infty = (0.49 \pm 0.02)\,h^{-1}$). During the first day in media without doxycycline we observed a transient decrease in the growth rate likely determined by the degradation dynamics of uracil reserves. After this transient all populations reached a low steady growth rate of about $0.2\,h^{-1}$ and between 1.5 and 3 days of continuous culturing all populations displayed a growth rate recovery to a level that often approached the rate observed in media with a high doxycycline concentration (Figure 2.4a, dashed line). At the end of each run cultures were frozen for later analysis.

[1] Culture densities were quantified by measuring optical density of samples at 600 nm against references consisting of growth media with no cells using a Hitachi U-1800 spectrophotometer. The OD_{600} figures reported in this document represent absorbance values at this wavelength.

[2] The conversion factor was measured by comparing the optical densities of samples with estimated OD_{600} values ranging from 10^{-5} to 10^{-2} with the number of colonies observed to grow in plates in which 50 μl of the corresponding samples had been inoculated. These experiments yielded a conversion factor of $(1.0 \pm 0.3)\,10^7\,\frac{\text{cells}}{\text{ml} \cdot \text{OD}_{600}}$ (best fit ± 95 % c.i.).

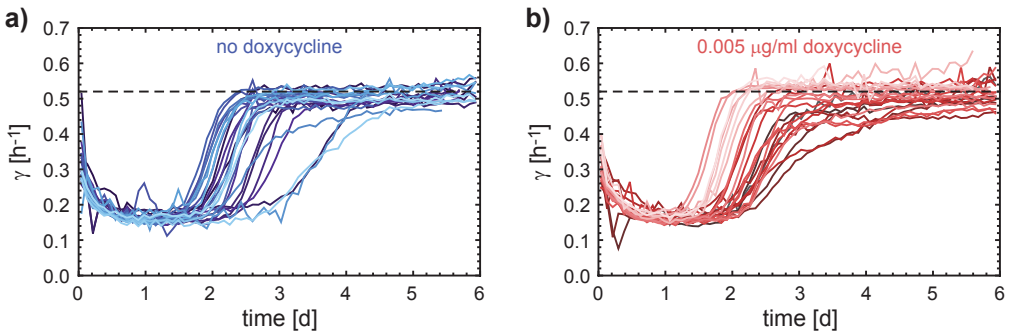

Figure 2.4: Adaptation dynamics as measured in turbidostat setups. **a**, **b**, Growth rate traces for 25 (a) and 26 (b) independent cultures grown in media with no uracil and either 0 or 0.005 μg/ml doxycycline. The dashed line indicates the growth rate of a population cultured for 24 hours in the presence of 5 μg/ml doxycycline.

In a second set of experiments we applied the same strategy to 25 populations that were grown in media without uracil but in the presence of a low concentration of 0.005 μg/ml doxycycline (Figure 2.4b). This concentration is too low to induce the TET07 promoter: the mRNA concentrations of *KlURA3* and the population growth rates after 24 hours are indistinguishable in the absence and presence of 0.005 μg/ml doxycycline (Figure 2.2). However, the system is poised closer to its induction threshold where it might be more sensitive to perturbations of the underlying regulatory circuit.

In order to test whether changes that occurred during the adaptation phase were stable, adapted cultures were transferred from the frozen stock into non-selective plates were they were allowed to grow for 2 days. They were later incubated in liquid non-selective media overnight and then transferred to media lacking uracil. After letting them grow for about 16 hours, growth rates in non-selective media were measured by keeping track of the optical density of each culture over the course of roughly 8 hours. In parallel, the same experiment was performed on the ancestor strain using media both lacking and containing uracil. This experiment mimics the behavior at the start of the adaptation experiment and in all cases the measured growth rate of the adapted cultures was significantly different than that of the ancestor and similar to wildtype values (Figure 2.5), indicating that some stable

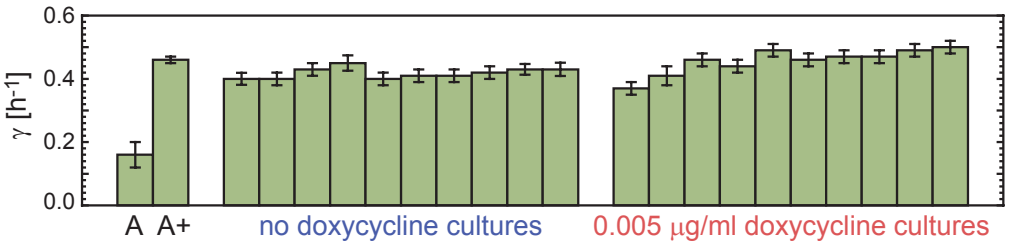

Figure 2.5: Growth rate measurements roughly 16 hours into growth in selective media (no uracil, no doxycycline) for each of ten adapted cultures characterized on each experiment as well for the ancestor strain (A) and the ancestor strain in non-selective media (A+, 5 μg/ml doxycycline). Error bars indicate the 95 % c.i. obtained by fitting an exponential growth model to measurements of optical density of the culture over a period of about 8 h.

changes had taken place in all adapted cultures.

2.4 Model for quantification of growth traces

During the adaptation experiment in turbidostat setups we have a fixed-size population of N asexual cells that, to first approximation, consists of two subgroups: the wildtype, which grows at rate γ_0, and mutant invaders that grow at rate γ_∞.

If we denote by n the number of mutants in the population, the state of the system is fully specified by this number and we can think of describing the dynamics of the system as transitions between different states that occur in an stochastic fashion. There are three possible types of events that could occur in a population with n mutants that will lead to a change of state:

1. a cell corresponding to the ancestor population divides and a mutant cell leaves the culture, leading to an unit decrease in n;

2. a mutant cell divides and one of the ancestors leaves the population, leading to an unit increase in n;

3. one of the ancestral cells mutates into the fitter phenotype, leading to an unit increase in n.

Note that in order to keep a fixed population we have to require one cell to leave the population if a cell has divided. In the turbidostat setup this is achieved by the exhaust line and to model this effect we approximate the process by thinking that as soon as one cell divides one cell is chosen at random from the overall population and leaves the culture.

To simplify the description we will consider that all these processes are independent of each other and independent of the timing of any previous events, allowing a description in terms of a Markov process [38, 39]. Each of the processes described above occurs with the following state-dependent rates ρ_i:

$$\rho_1 = [\gamma_0(N-n)] \times \left[\frac{n}{N+1}\right] \simeq \gamma_0 \frac{n(N-n)}{N}, \tag{2.1}$$

$$\rho_2 = [\gamma_\infty n] \times \left[\frac{N-n}{N+1}\right] \simeq \gamma_\infty \frac{n(N-n)}{N}, \tag{2.2}$$

$$\rho_3 = \mu(N-n). \tag{2.3}$$

In (2.1, 2.2) the terms in square brackets represent the rate at which division events occur (first term) and the probability that the corresponding type of cell will be chosen to leave the population (second term). Given that $N \gg 1$, we considered the approximation $N+1 \simeq N$ in writing the final expressions in (2.1) and (2.2). In (2.3) μ represents the rate at which ancestor cells mutate into the fitter phenotype.

So, we can summarize the model using the following schematic description of the underlying Markov Chain:

$$\boxed{n-1} \xleftarrow[\gamma_0 \frac{(N-n)n}{N}]{} \boxed{n} \xrightarrow{\mu(N-n)+\gamma_\infty \frac{n(N-n)}{N}} \boxed{n+1} \tag{2.4}$$

In the deterministic limit, *i.e.* once the number of individuals in each population is large enough so that fluctuations become negligible and we can

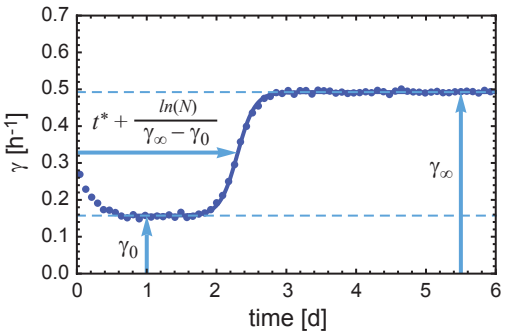

Figure 2.6: Fit of the one the traces in Figure 2.4a to the "one mutant take-over" model described by equation (2.7).

think of n as a continuous variable the evolution of the number of mutants in the population satisfies the equation

$$\frac{\mathrm{d}n}{\mathrm{d}t} = (\gamma_\infty - \gamma_0)\, n(N-n) \tag{2.5}$$

from which we can derive an explicit expression for the dynamics of the fraction $x = {}^{n}/_{N}$ of mutant cells in the population

$$x(t) = \frac{1}{1 + e^{-(\gamma_\infty - \gamma_0)(t - t_c)}} \tag{2.6}$$

where t_c is an integration constant.

Expression (2.6) permits one to calculate the instantaneous population growth rate $\gamma(t)$ as

$$\gamma(t) = \gamma_\infty x + \gamma_0(1-x) = \gamma_0 + \frac{\gamma_\infty - \gamma_0}{1 + e^{-(\gamma_\infty - \gamma_0)(t - t_c)}} \tag{2.7}$$

and this expression can be fitted to the growth rate traces obtained from the turbidostat setups. This allows one to extract the parameters γ_0, γ_∞ and t_c from each trace. In Figure 2.6 we present a typical fit of this expression to the growth rate estimates between days 1 and 5 of a sample trace, indicating how each parameter describes a different feature of the experimental curve.

In Figures 2.7, 2.8 we present the fits to all individual traces to illustrate the level of agreement between this model and the data.

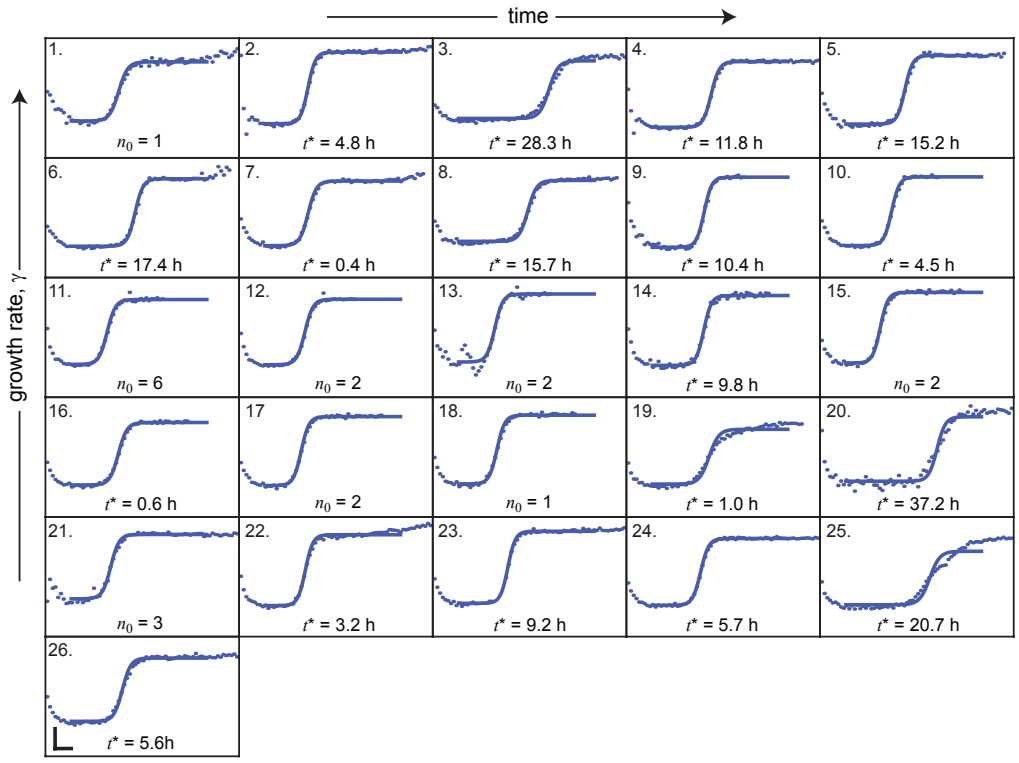

Figure 2.7: Fits of each individual growth trace obtained in the experiments performed in the absence of doxycycline to the model described by equation (2.7). For each trace, we include the inferred value of n_0 or $t^\star$ depending on whether we estimate that mutants were generated before the start of the turbidostat experiment or after. The bars in panel 26 indicate scale and represent $0.1\,\mathrm{h}^{-1}$ in the growth rate axis and 12 h in the time direction.

Knowing these parameters one can estimate the number of mutant cells at the beginning of the turbidostat experiments using the expression

$$n_0 \simeq Nx(t=0) = \frac{N}{1+e^{(\gamma_\infty-\gamma_0)t_c}}. \tag{2.8}$$

The condition $n_0 \geq 1$ implies that some mutants were present at the start of the run and therefore had been generated in the conditions of no selection (media with uracil) at which the cells had been grown overnight,

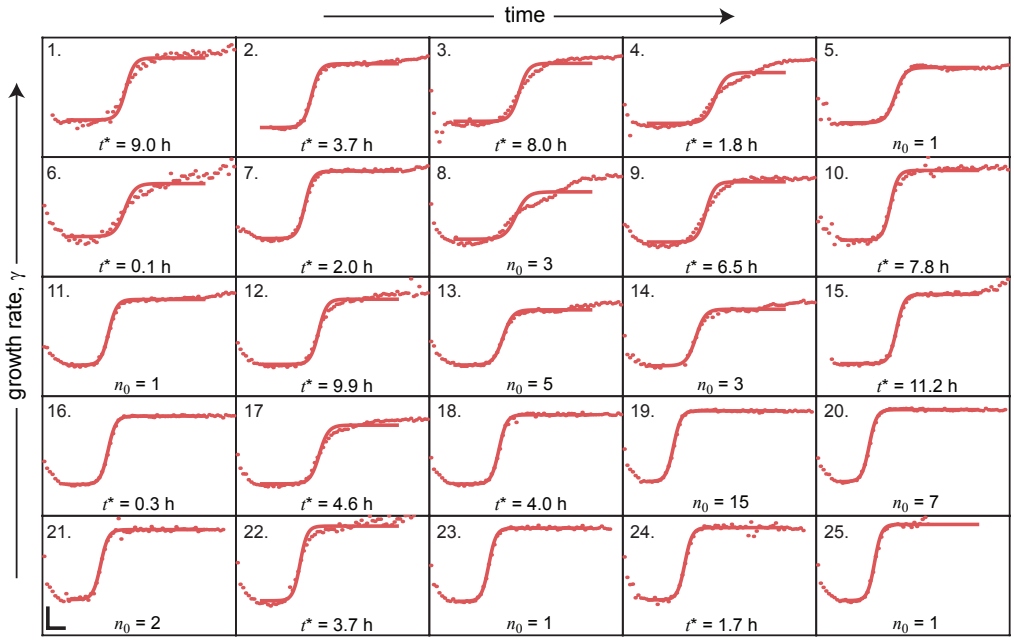

Figure 2.8: As in Figure 2.7 but for the experiments performed at a concentration of 0.005 μg/m doxycycline. The bars in panel 21 indicate scale and represent 0.1 h^{-1} in the growth rate axis and 12 h in the time direction.

before cultures had been washed and transferred to media without uracil. In those growth conditions both the wildtype and the mutant strains grow at the same rate γ_0 and the size of the population increases exponentially with time at this rate. This is the same process described by Luria and Delbrück in their seminal work [40].

In this case, one expects that the distribution of mutant cells across the cultures that were used to seed the turbidostats will follow a Luria-Delbrück distribution. Following Lea and Coulson [41–43], one can express the probability D_r of having r mutants in a population that was exponentially propagated to a size $N_0 \gg 1$ by the equations

$$\begin{cases} D_0 = e^{-m}, \\ D_r = e^{-m} \sum_{j=1}^{r} C_{j,r} \dfrac{m^j}{j!} \end{cases} \tag{2.9}$$

where $m = \frac{\mu N_0}{\gamma_0}$ and the coefficients $C_{j,r}$ are defined by the relations

$$\begin{cases} C_{0,r} = C_{r,0} = 0, \\ C_{1,r} = \dfrac{1}{r(r+1)}, \\ C_{j,r} = \dfrac{1}{j+r}\left[jC_{j-1,r-1} + (r-1)C_{j,r-1}\right]. \end{cases} \tag{2.10}$$

In our situation, the populations used for seeding the turbidostat experiments were only a fraction $\alpha = {}^{N}/_{N_0}$ of the cells that were propagated overnight, with $\alpha \simeq 0.1$. In this situation the probability P_r of finding r mutant cells at $t = 0$ in the turbidostat is

$$P_r = \sum_{k_r}^{\infty} H_{k,r}^{N_0,N} D_k \tag{2.11}$$

where $H_{k,r}^{N_0,N}$ (the hypergeometric distribution) represents the probability of getting r mutants when sampling N cells from a population of N_0 cells that contains k mutants:

$$H_{k,r}^{N_0,N} = \frac{\binom{k}{r}\binom{N_0-k}{N-r}}{\binom{N_0}{N}}. \tag{2.12}$$

In the limit in which the number of mutants is small we can approximate (2.11) by

$$P_r \simeq \sum_{k_r}^{\infty} \binom{r}{k} (1-\alpha)^{k-r} \alpha^r D_k. \tag{2.13}$$

In Figure 2.9 we present the measured distributions of n_0 as well as fits to the model described in these paragraphs, from where we can get a first estimate of mutation rates, albeit with a large uncertainty. Taking $N_0 \simeq 2\,10^8$ cells the estimates we obtained from this analysis were $\mu = (0.41 \pm 0.15)\,10^{-8}\,\mathrm{h}^{-1}$ (best estimate $\pm$ 95 % c.i.) for the experiments performed without doxycycline and $\mu = (0.60 \pm 0.15)\,10^{-8}\,\mathrm{h}^{-1}$ for the cultures adapted in $0.005\,\mu\mathrm{g/ml}$ doxycycline.

In the case in which $n_0 < 1$ we can estimate that there were no mutants at the beginning of the experiment and therefore that they had to appear

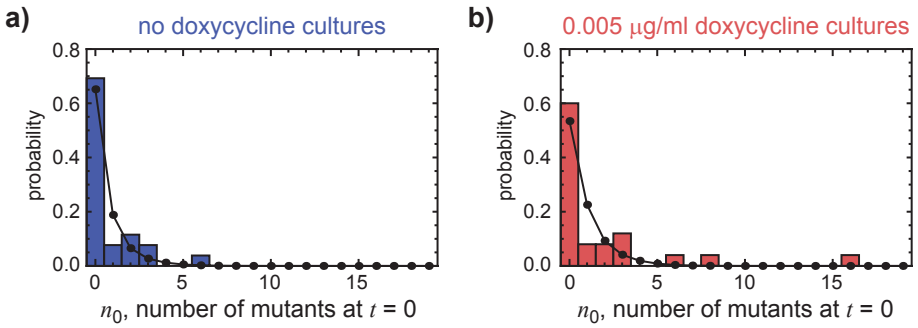

Figure 2.9: **a**, Measured distribution (n = 26) of the estimated initial number of mutants in the experiment performed with no doxycycline. The solid line represents a fit to the Luria-Delbrück distribution considering a subsampling factor of $\alpha = 0.1$. **b**, as in a but for the experiment performed at a concentration of $0.005\,\mu$g/ml doxycycline (n = 25)

sometime during the selection process. In this case we can interpret the time $t^\star$ at which the fraction of mutants is of order $1/N$ as the time of appearance of the first mutant in the population. Some simple algebra leads to the following expression for large N

$$\frac{1}{N} = x(t = t^\star) = \frac{1}{1 + e^{(\gamma_\infty - \gamma_0)(t^\star - t_c)}} \quad \Rightarrow \quad t^\star \simeq t_c - \frac{\ln(N)}{\gamma_\infty - \gamma_0}. \tag{2.14}$$

Neglecting the potential effect of the appearance of successive mutations before the number of mutants reaches a significant level so that the deterministic approximation is valid, the distribution of the times of appearance of fitter mutants is exponential with rate μN, as the mutations have a random chance of appearing at any time, *i.e.* their appearance is a Poisson process. By observing the measured distribution of $t^\star$, or $Nt^\star$, one can validate this hypothesis and obtain a measurement of the effective mutation rate μ.

This simple model neglects stochastic fluctuations and cloning interference effects [44] but it nevertheless provides a first order description of the dynamics based on effective coarse-grained parameters that is useful for relative quantification of features of the experiments performed at different doxycycline concentrations (see for example [45] for a similar approach).

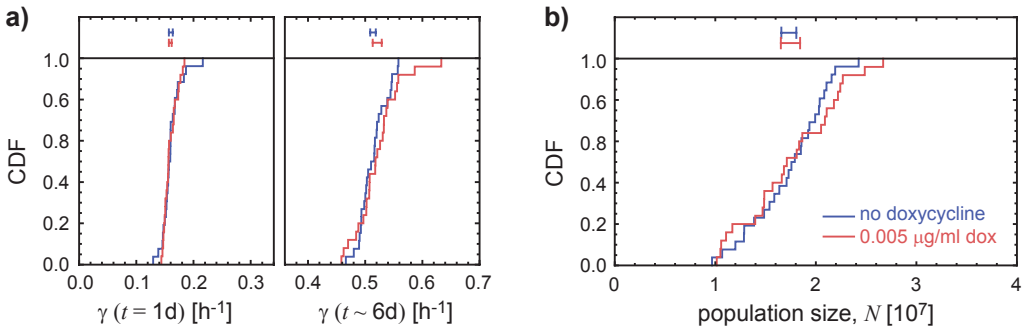

Figure 2.10: **a**, For each set of experiments, cumulative distributions (CDF) of growth rates 1 day into the experiment and at either 6 days or the end of the experiment. **b**, Cumulative distribution (CDF) of population sizes for each set of experiments. On top of the plots, bars indicate the experimental mean and standard deviation.

2.5 Quantification of growth traces

We observed no significant differences in the distribution of population sizes, initial or final growth rates across the two doxycycline concentrations explored (Figure 2.10). In contrast, a clear difference was observed between the $t^\star$-distributions of the two data sets (Figure 2.11). Both $t^\star$-distributions were well approximated by exponentials with different apparent mutation rates: $(0.50 \pm 0.03)\,10^{-8}\mathrm{h}^{-1}$ (best fit $\pm$ 95 % c.i.) in the absence of doxycycline and $(0.91 \pm 0.06)\,10^{-8}\mathrm{h}^{-1}$ in the presence of $0.005\,\mu\mathrm{g/ml}$ doxycycline. These numbers are consistent with the estimates based on fits of the estimated number of mutants at the start of the run to rescaled Luria-Delbrück distributions as described in the previous section. The order of magnitude of these numbers is consistent with recent estimates of per-base-pair mutation rates [46]. So, this analysis suggest that there is indeed a faster rate of generation of fitter phenotypes in the experiment performed at the low, nonzero doxycycline concentration explored.

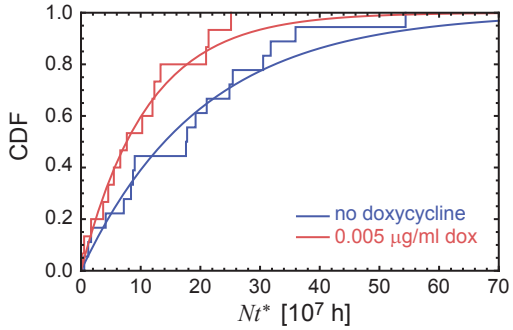

Figure 2.11: Experimental cumulative distributions (CDF) of the population-rescaled times of appearance of the first fitter mutant ($Nt^\star$) as estimated from each growth rate trace. The solid lines are best fits to exponential distributions.

2.6 Molecular characterization of cultures

To reveal the molecular changes in our synthetic gene circuit we characterized the phenotypes of 10 independent mutants in each of the two conditions explored by measuring different traits associated with the underlying network in order to establish if any genetic changes had occurred and to investigate whether the observed increase in the apparent rate of appearance of fitter mutants could be explained by an increase in the number of available beneficial phenotypes. In Figure 2.12 we present a summary of the traits explored and we discuss them in detail in the following subsections.

2.6.1 Quantification of transcript levels and genomic copy numbers

To quantify the level of the relevant transcripts and genomic copy numbers in the adapted cultures each population was first grown overnight in non-selective minimal media and then their RNA was extracted using the RiboPure^{TM}Yeast kit (Applied Biosystems/Ambion). The total RNA concentration of each sample was determined spectroscopically (NanoDrop®ND

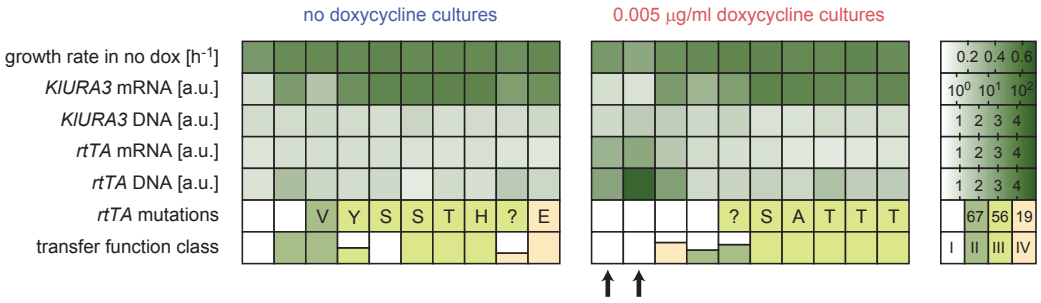

Figure 2.12: Measured traits for 10 adapted cultures in each of the two set of experiments described. Each column represents one evolved population and each row one trait. Reported traits: growth rate in the selective environment after adaptation, mRNA levels relative to the ancestor as measured by q-RTPCR, relative DNA levels as measured by qPCR, mutations found in the *rtTA* gene (the code in each box indicates the new observed aminoacid and the position is indicated by color as referenced on the right; question marks indicate experiments in which the base pair could not be identified with certainty) and distribution of transfer function classes (labeled according to the scheme discussed in subsection 2.6.3; multiple outcomes in one box indicate that independent transformations produced different results). Black arrows indicate cultures in which a significant increase in the content of genomic *rtTA* was detected by qPCR.

1000, NanoDrop) and then samples were diluted to equal concentrations. The relative levels of the different transcripts were measured by quantitative reverse-transcription PCR (q-RTPCR) using the QuantiFast SYBR Green RT-PCR kit (Qiagen) and a MJ Research PCR Machine.

In order to quantify DNA levels a similar procedure was carried on, only that instead of RNA, genomic DNA was extracted following a lyticase based custom protocol and quantitative PCR (qPCR) with reagents from the QuantiFast SYBR Green PCR kit (Qiagen) was used to quantify the relative levels associated with each target.

In all cases calibration curves as a function of mass content were measured and *ACT1* levels were used as a reference. The primers used for each gene targeted a 200 base-pairs stretch of the corresponding coding region and in all

cases the calibration curves indicated a figure consistent with 100% efficiency within experimental uncertainty.

Quantification of transcript levels revealed an increase of *KlURA3* mRNA in most adapted cultures (Figure 2.12, second row), consistent with the idea that they had achieved a fitter metabolic state by recovering wildtype levels of Ura3 protein expression. We observed no correlated increase of the *KlURA3* gene dose (Figure 2.12, third row) indicating that the increase in the level of *KlURA3* transcripts was not due to an increase in the genomic copy number of the corresponding gene.

We did not observe any change in *rtTA* mRNA and *rtTA* gene dose in mutants that were evolved in the absence of doxycycline across 26 independent cultures (Figure 2.12, fourth and fifth rows). However in the presence of doxycycline we found 4 cultures out of 25 that displayed an increase in *rtTA* gene dosage and a correlated increase in the corresponding RNA concentrations (black arrows in Figure 2.12 highlight this effect in two cultures).

These observations indicate that most of the mutants evolved into a fitter phenotypic state by increasing the production of *KlURA3* but that except for 16% of the cases in the experiment performed at $0.005\,\mu$g/ml doxycycline this increase was not correlated to changes in genomic copy numbers.

2.6.2 Sequencing

Genomic DNA from each culture was used as a template for PCR reactions targeted at amplifying the synthetic constructs containing either the *rtTA* or *KlURA3* genes including promoter, coding and terminator regions. For each construct several sequencing primers were chosen so as to produce a complete tiling of the DNA stretch and, for each primer, sequencing was performed at the MIT Biopolymers Laboratory using an Applied Biosystems 3730 capillary DNA sequencer with the Big Dye Terminator Cycle Sequencing Kit. Each obtained sequence was thresholded for base pairs of high quality and alignment was performed manually.

We observed no mutations in the *KlURA3* coding, promoter, and terminator sequences. On the other hand, in most cases we found point muta-

tions in the open reading frame of the *rtTA* gene (Figure 2.12, sixth row) in both data-sets. Interestingly, most of the observed mutations occurred at amino acid position 56, which, among others, had been previously identified in *in-vitro* mutagenesis studies as involved in the functional transformation between the *rtTA* and *tTA* proteins, the latter exhibiting the opposite doxycycline regulation logic [47].

We note that there was a fair amount of variability in the observed mutations, indicating that there are multiple ways in which the circuit adapted to the imposed evolutionary pressure. The fact that all mutations were found in the coding region of the *rtTA* gene could be related to the fact that this gene is exogenous to the host organism and therefore had not being exposed to long periods of evolutionary pressure compared to endogenous components. It could also be a feature of this particular gene, which is interesting in the sense that this system could be prove to be a good tool for engineering genetic networks by experimental evolution as this component would offer a large degree of plasticity, making such experiments feasible in laboratory timescales.

2.6.3 Transfer function analysis

To test whether the circuit regulation had been significantly modified by the observed mutations we chromosomally integrated a construct consisting of the yellow fluorescent protein (*YFP*) driven by the *TET07* promoter into each mutant allowing us to further characterize each adapted population by measuring its response to different inducer levels in what we will refer to as the "transfer function"of the system.

To do this, a construct consisting of the *TET07* promoter driving *YFP* plus the *UTR1* terminator region was cloned into the pRS303 plasmid [32] and then integrated into the *his3* locus of the strain to be characterized. After the transformation was performed several colonies were chosen and after growing them overnight in minimal media with different doxycycline concentrations *YFP* expression levels were quantified by flow cytometry (FACSCalibur HTS, Becton Dickinson). In some cases different colonies out of the same transformation yielded different response curves, which we interpreted as the

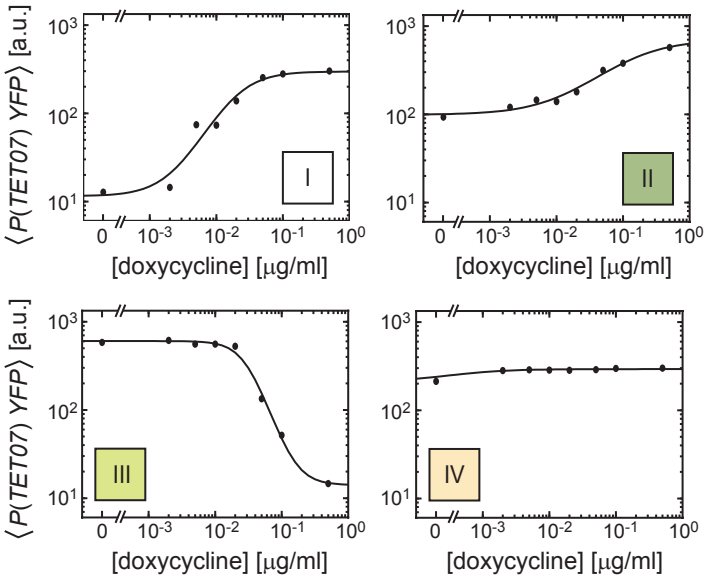

Figure 2.13: Mean fluorescence level as a function of doxycycline in uracil rich media for selected mutants corresponding to different transfer function classes.

original culture not being completely homogeneous at the time of finalization of the adaptation run. Some colonies yielded no signal, which we interpreted as failed transformants. For each culture, at least 4 signal-producing colonies were analyzed and then frozen for future studies.

We observed 4 different transfer function classes, correlated with the observed mutations (Figure 2.13, 2.12 sixth and seventh rows): (I) a wildtype-like response, observed mostly in the cases in which no mutations were detected; (II) positive regulation with a basal level higher than in the wildtype case; (III) cases in which the role of the inducer got inverted, strongly correlated with an amino acid change at position 56, though the substitution was not always the same; and (IV) an unregulated response.

2.7 Validation of observed genetic changes as the source of a fitness increase

To explore if the observed genetic changes were correlated to the observed fitter phenotypic states we decided to translate those changes into an ancestor strain that had never been exposed to the selective pressure. Then we tested whether the growth rate of these hybrid strains decreased when they had to face the selective environment that the ancestor was subjected to during the adaptation run. And furthermore, we analyzed whether their transfer functions had been affected.

2.7.1 Transplants of the *rtTA* construct

In order to perform *rtTA* transplants from each adapted culture back into the ancestor strain we first prepared the ancestor by inserting the *YFP* reporting system described in section 2.6.3 to facilitate the determination of transfer functions after each transplant. Then the original *rtTA* construct was deleted by replacing it by a construct consisting of the *NAT1* gene driven by the *TEF* promoter and finished with the *TEF* termination sequence as PCRed out of the pAG25 plasmid [48]. During the PCR step 50 base-pair DNA pieces homologous to regions flanking the target were added to each end. This PCR product was introduced into the ancestor and transformants were selected in the presence of the antibiotic nourseothricin. Deletions were confirmed by PCR analysis, by checking for the absence of *YFP* signal at different doxycycline levels and by confirming the loss of the ability to synthesize adenine (the marker related to the original *rtTA* construct). Once the ancestor was prepared in this way, the *rtTA* constructs from each adapted culture were PCRed out of corresponding genomic preparations using primers that had 50 base-pair overhangs of homology to the region in the prepared ancestor now containing the *NAT1* construct. These PCR products were introduced into the prepared strain and transformants were selected in media lacking adenine. By performing the transplants in this way, the effect on the genome of the ancestor strain with the reporter system was minimal.

We found a strong correlation between both the distribution of transfer

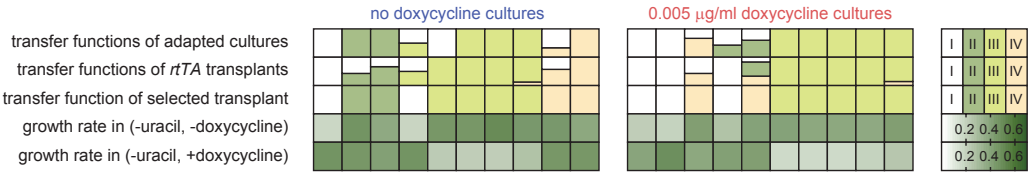

Figure 2.14: For the adapted cultures presented in Figure 2.12 the rows indicate, in order: distribution of observed transfer function classes of the adapted cultures (labeled according to the scheme shown in Figure 2.13); distribution of observed transfer function classes of strains that were constructed by transplanting the *rtTA* construct from each adapted culture into the ancestral strain; the transfer function class of a selected *rtTA* transplant chosen for growth rate characterization; the growth rate of such a clone in media with no uracil and no doxycycline; and the growth rate of the same clone in media with no uracil and 5 μg/ml doxycycline.

function classes and growth rates measured on these transplanted strains and the mutants (Figure 2.14). To further assess the functional implication of the mutations in the *rtTA* gene and their impact on fitness we measured the growth rates of the strains with *rtTA* transplants in media with a concentration of either 0 or 5 μg/ml doxycycline (Figure 2.14, bottom two rows). We found a strong correlation between the growth rate patterns in the different media and the observed transfer function classes, which demonstrates that the *rtTA* transplants are not only responsible for the observed changes in the *YFP* transfer function but that they also correlate with fitness.

This indicates that, in most cases, the observed mutations in the *rtTA* gene are a major cause of the observed adaptation

2.7.2 Addition of an extra copy of the *rtTA* construct

To analyze the effect of having an additional copy of *rtTA* we inserted an extra copy of this construct into the ancestor background and quantified the change this produced on the transfer function associated with the system as well as its effect on fitness.

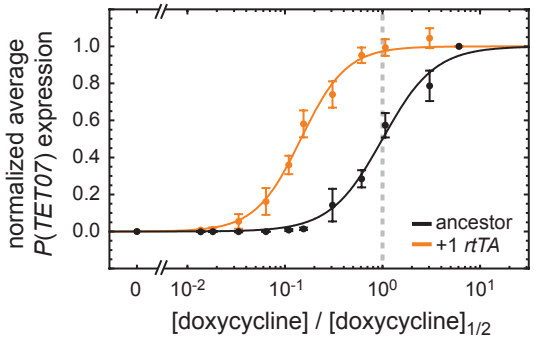

Figure 2.15: Normalized average *P(TET07)YFP* transfer functions (mean ± s.e., $n = 4$) for the ancestor and a strain like the ancestor but with an additional copy of the *rtTA* construct (+1 *rtTA*). Solid lines represent fits to Hill functions as guide to eye. The doxycycline concentration was normalized to the mid-induction point of the transfer function of the ancestor ($[\text{doxycycline}]_{1/2} = 0.33\,\mu\text{g/ml}$).

Transfer function characterization

We digested the plasmid used for introducing the *rtTA* construct into the ancestor (see Section 2.2) with the restriction enzymes KpnI and NotI flanking the construct and ligated it into a pRS304 [32] backbone digested with the same enzymes. The pRS304 plasmid has a functional copy of the *TRP1* gene, which if integrated correctly confers W303 *trp1* cells the ability to grow in media with no tryptophan. We transformed the resulting plasmid linearized with EcoNI into the ancestor and selected for positive transformants in media without tryptophan. We confirmed by qPCR on the *rtTA* gene that the genomic content of *rtTA* had increased after this transformation and the measurements we obtained were consistent with a single integration. After this modification we introduced into the resulting strain the *YFP* reporting system as described in Section 2.6.3.

An increase in the overall concentration of this transcriptional activator implies that less doxycycline is needed to achieve a given rate of protein production. This manifests in a shift of the transfer function curve towards lower doxycycline concentrations (Figure 2.15).

Quantification in terms of a biochemical model of the underlying circuit

To describe the binding response of the system to changes in *rtTA* dosage we considered a biochemical model of the underlying interactions. We took into account that doxycycline can bind to *rtTA* proteins and that in this active configuration, *rtTA* can bind to any of $M = 7$ independent and identical binding sites in the *TET07* promoter region. Finally, we considered the transcription rate to be proportional to the number of *rtTA* proteins bound to the *TET07* promoter. If we denote the concentration of free doxycycline by d, the concentration of free *rtTA* by r, the concentration of active *rtTA* by $r^\star$ and the average concentration of *TET07* sites with k bound *rtTA* proteins across a population of cells by p_k we can write the following equations for describing this biochemical system:

$$\begin{cases} d_{\text{tot}} = d + r + \sum_{k=1}^{M} k p_k, \\ r_{\text{tot}} = r + r^\star + \sum_{k=1}^{M} k p_k, \\ p_{\text{tot}} = \sum_{k=1}^{M} k p_k, \\ dr = K_1 r^\star, \\ r^\star p_k = K_2 \dfrac{k+1}{N-k} p_{k+1} \quad (n = 0, \ldots, N-1). \end{cases} \tag{2.15}$$

In this system the first three equations represent conservation of the total amount of doxycycline (d_{tot}), *rtTA* proteins (r_{tot}) and transcription sites (p_{tot}), the fourth equation describes equilibrium in the binding between doxycycline and *rtTA* and the last set of equations represent the equilibrium binding of active *rtTA* to promoters with free binding sites.

This model can be reduced to the following equations for computing the normalized average transcription f

$$\left(d_{\text{tot}} - r^\star - N p_{\text{tot}} \frac{r^\star}{K_2 + r^\star}\right)\left(r_{\text{tot}} - r^\star - N p_{\text{tot}} \frac{r^\star}{K_2 + r^\star}\right) = K_1 r, \tag{2.16}$$

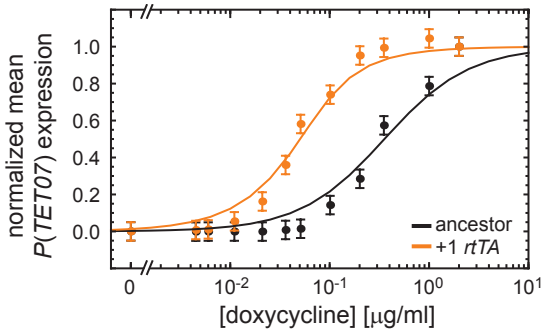

Figure 2.16: Normalized average *P(TET07)YFP* transfer function for the ancestor and a strain like the ancestor but with an additional copy of the *rtTA* construct (+1 *rtTA*) as presented in Figure 2.15 and best fit to the biochemical model described in section 2.7.2. The error bars represent the 10% nominal normal uncertainty considered in the fitting procedure.

$$f = \frac{r^\star}{K_2 + r^\star}. \tag{2.17}$$

Given parameters $\{d_{\text{tot}}, r_{\text{tot}}, Np_{\text{tot}}, K_1, K_2\}$ the first of these equations can be solved numerically for $r^\star$, for instance by applying a bisection method with $r^\star \in [0, r_{\text{tot}}]$. The solution can then be plugged in equation (2.17) to obtain the corresponding normalized average transcription level.

We fitted this model to the data consisting of the normalized average transcription levels as a function of doxycycline considering the datasets corresponding to both the ancestor and the strain that contains an extra copy of *rtTA*. We used as fit parameters a transformation of $\{r^{\text{A}}_{\text{tot}}, r^{\text{A+}}_{\text{tot}}, Np_{\text{tot}}, K_1, K_2\}$ where $r^{\text{A}}_{\text{tot}}$ represents the inferred total concentration of *rtTA* in the ancestor and $r^{\text{A+}}_{\text{tot}}$ is the corresponding concentration in the strain that has an extra copy of the *rtTA* construct. The transformation used consisted in considering the following parameter combinations as independent:

$$\left\{\log_{10}\left(r^{\text{A}}_{\text{tot}}\right), \frac{r^{\text{A+}}_{\text{tot}}}{r^{\text{A}}_{\text{tot}}}, \log_{10}\left(\frac{Np_{\text{tot}}}{r^{\text{A}}_{\text{tot}}}\right), \log_{10}\left(\frac{K_1}{r^{\text{A}}_{\text{tot}}}\right), \log_{10}\left(\frac{K_2}{r^{\text{A}}_{\text{tot}}}\right)\right\} \tag{2.18}$$

In Figure 2.16 we show the results of a fitting procedure based on a

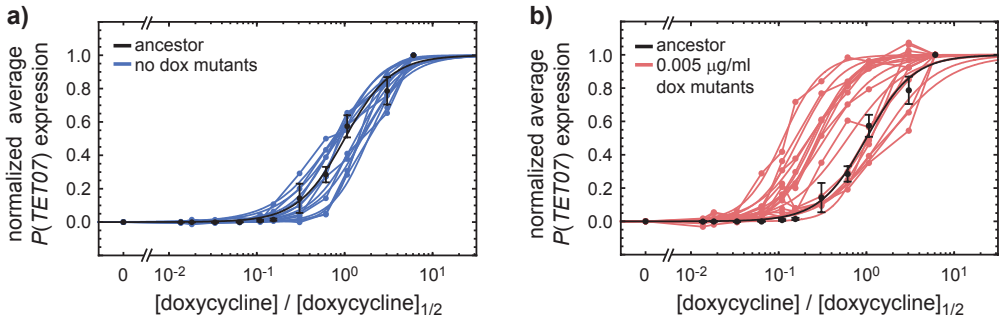

Figure 2.17: **a**, Normalized transfer functions of 8 clones isolated from cultures adapted in the absence of doxycycline and corresponding to transfer functions of class I. **b**, As is b but for 11 clones isolated from cultures that had been evolved in the presence of $0.005\,\mu$g/ml doxycycline. ($[\text{doxycycline}]_{1/2} = 0.33\,\mu$g/ml).

Bayesian inference approach [49] considering a normal likelihood model with a nominal uncertainty of 10% per data point.

The inferred value of the ratio $r_{\text{tot}}^{\text{A+}}/r_{\text{tot}}^{\text{A}}$ was (10 ± 8) (mean $\pm$ 95 % c.i.). This value is consistent with an increase in *rtTA* production, though the number is higher than what would be expected for just a duplication. One potential explanation of this observation is that this increases could originate from differences between transcriptional efficiencies at different chromosomal locations.

Alternative models that consider different interaction structures at the promoter sites (*v.g.* a model with independent sites but in which transcription in driven at the same rate as long as one *rtTA* protein is bound, a model that considers cooperativity through an effective Hill coefficient, etc.) yield similar results.

Note that this analysis is essentially equivalent to considering that an increase in the *rtTA* copy number produces a shift of the mid-point of the normalized transfer function vs. doxycycline curve towards smaller values and that a measurement of such a shift represents the effective increase in the *rtTA* concentration.

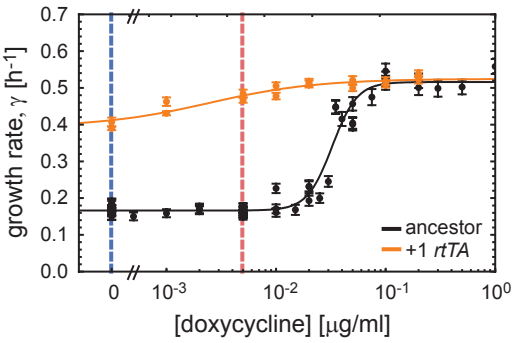

Figure 2.18: Growth rate profile over a range of doxycycline concentrations for both the ancestor and the strain with an extra copy of the *rtTA* construct (+1 *rtTA*).

Comparison with adapted cultures

We found that measurements of class I transfer functions associated with cultures that adapted in the absence of doxycycline were consistent with the curves measured in the ancestor (Figure 2.17a). On the other hand, class I transfer functions of populations evolved in 0.005 μg/ml doxycycline showed profiles ranging from that of the ancestor to that of the ancestor containing an extra copy of the *rtTA* construct (Figure 2.17b), providing further evidence that some of these cultures include *rtTA* duplications.

Effect on fitness

Characterization of the growth rate of cultures as a function of doxycycline for both the ancestor and the strain with an additional copy of *rtTA* evidenced that the benefit of duplicating this gene is more pronounced at a concentration of 0.005 μg/ml doxycycline (Figure 2.18), providing a partial explanation for why this evolutionary solution was preferentially observed in the adaptation experiment performed at 0.005 μg/ml doxycycline.

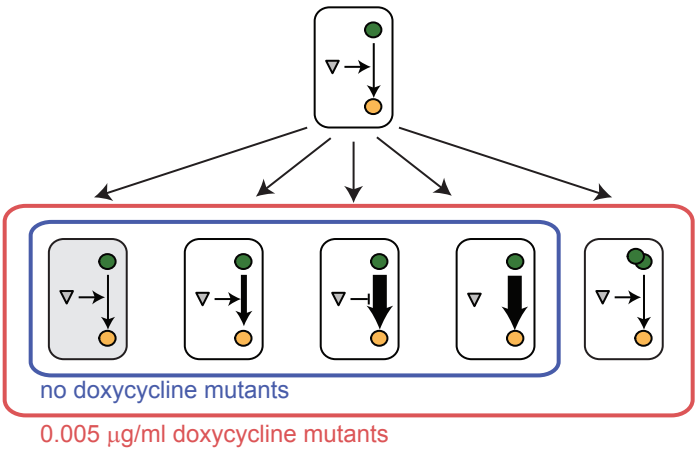

Figure 2.19: Schematic of the observed changes in the genetic circuit after adaptation. The diagram on the top represents the initial system and on the bottom, from left to right, the different boxes represent: no observed changes in the circuit (hypothetical modifications somewhere else), increase of basal production of Ura3, logic reversal of doxycycline regulation, a system that has become independent of doxycycline and a system in which the genetic dosage of *rtTA* had increased. The blue and red surrounding boxes indicate in which set of experiments these circuit rewirings were observed.

2.8 Conclusions

Taken together, these experiments suggest that mutations in the *rtTA* gene change its function in the circuit and improve growth through its effect on the production of Ura3 transcripts and then proteins. These beneficial mutations are observed in both datasets. However we only observed duplications of the *rtTA* gene in mutants evolved in the presence of 0.005 μg/ml doxycycline and never in the absence of inducer. This suggests that the *rtTA* gene duplication has a positive effect on fitness and becomes a viable additional evolutionary path only in the presence of a low doxycycline concentration. If the duplication process occurs at a rate similar to that of point mutations [50] the availability of an extra evolutionary path would result in the faster apparent effective mutation rate compared to the evolution of mutants in media without doxycycline in which duplications were never observed. This might pro-

vide a partial explanation for the observed increase in the measured apparent mutation rate across the two environments studied (Figure 2.11).

Our work illustrates how even simple genetic networks can evolve in diverse ways in response to a simple selective environment (Figure 2.19). We discussed how alternative evolutionary solutions might become available as the environment makes the underlying circuit work at different operating points. In the case at hand the addition of a small amount of an extracellular inducer puts the network close to its inducibility threshold and triggers the appearance of gene duplications of the upstream activator as feasible evolutionary solutions to the imposed fitness constraint. Finally, we also discussed how the appearance of these alternative solutions affects the apparent effective mutation rate of a culture.

These observations and approaches provide a framework for exploring how the multitude of adaptive solutions that endogenous gene networks might explore in the presence of selective pressure are constrained by the operating state of the underlying circuit.

Chapter 3

The effect of gene dosage in a complex network of genes

3.1 Introduction

Having seen the effect of gene duplications on a small genetic system with minimal interactions the question of what will be the effect of gene duplications in more complex system naturally arises. In this chapter I present an exploration of this question in the context of the galactose uptake network in the yeast *Saccharomyces cerevisiae*.

In a diploid background we combinatorially deleted one of the two copies of the regulatory genes of this network (*GAL2*, *GAL3*, *GAL4*, *GAL80*) obtaining 16 diploid strains that allowed us to measure and characterize the effect of the dosage of all these genes around the diploid phenotype. Interestingly we found that only two of these genes had a significant effect in a phenotype of choice, strengthening the notion presented in the previous chapter that the effect of a gene duplication is strongly dependent on the operating point of the underlying genetic circuit. Furthermore, we found that the activity of the network was invariant to a change in network dosage, *i.e.* the phenotype did not change if the copy numbers of all genes were modified in a proportional manner.

This last observation is interesting not only in the context of evolutionary phenomena but it is also important as a potential way for cells to cope with undesirable variations in network-dosage and therefore maintain optimal activity levels in gene networks. The number of copies of a gene network in a cell, or network dosage, has a direct effect on cellular phenotypes [51,52]. Network dosage is altered in situations such as the switching of some organisms between haploid and diploid life forms [53,54], doubling of chromosomes during cell cycle [55,56], genome-wide duplication of genetic content [16,57], and global variation [58] in gene expression. Different phenotypes have different levels of sensitivity to such variations and the need for effective compensation mechanisms arises when cells cannot tolerate these alterations.

It is believed that in the transition between haploid and diploid forms of life cells utilize a volume-mediated compensation mechanism to keep the concentrations of transcription factors constant as cell volume increases with ploidy [53]. However, this mechanism cannot subdue the effects of global expression variation and genome duplication or loss events as they affect cellular phenotypes independently of cell volume. These observations raise the question of whether there are alternative layers of dosage compensation mechanisms independent of external factors such as cell volume. To what extent would network activity be robust to alterations in network dosage if we fixed cell volume and therefore excluded its compensatory effect? Could there be a molecular mechanism intrinsic to the network structure that helps cells diminish the effects of dosage variations? Despite the fundamental nature of these questions, what these mechanisms are and how they can be implemented has remained unclear.

3.2 The galactose uptake network

The galactose upkate network (*GAL* network in short, Figure 3.1) is an ideal platform to experimentally investigate the question of the effect of gene and network-dosage. It has a well-characterized [59] bistable expression profile due to nested positive and negative feedback loops, ubiquitous regulatory elements in eukaryotic gene networks. Bistability [59–61] is a dynamical system property giving rise to two distinct gene expression states (OFF and ON) in a population of isogenic cells grown in the same environment. In

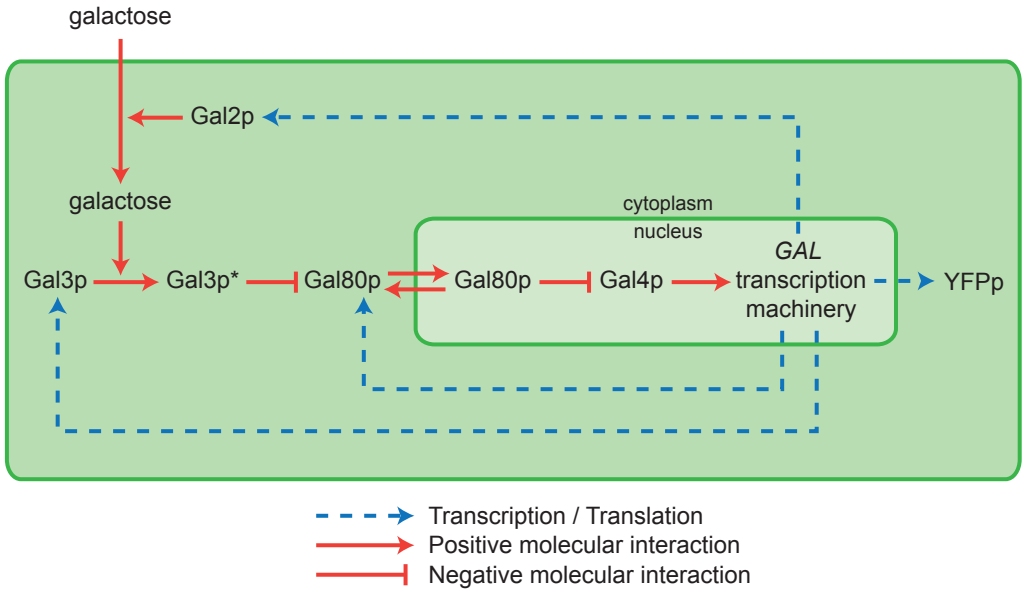

Figure 3.1: Regulatory components of the galactose uptake network and their interaction topology. Red arrows with pointed and blunted ends denote activating and inhibiting molecular interactions, respectively. Gal3p* represents the galactose-bound, active form of Gal3p. The shuttling of Gal80p between the cytoplasm and the nucleus is denoted by the bidirectional red arrows. The dotted blue arrows show how the transcriptional feedback loops are established through Gal2p, Gal3p, and Gal80p.

a bistable gene network, the fraction of cells occupying the ON-state can be defined as the inducibility of the system and serves as a quantitative phenotypic trait.

In the *GAL* network, four genes (*GAL2*, *GAL3*, *GAL4*, and *GAL80*) play key roles in regulating gene expression. The constitutively expressed Gal4p protein is a transcriptional activator that regulates expression of the other *GAL* pathway genes by binding to DNA sites upstream of the corresponding open reading frames [62]. Gal80p binds [63] to this protein and prevents Gal4p-mediated transcriptional activation, establishing a negative feedback loop. The protein Gal3p is activated [64] by galactose molecules that are imported into the cell by the galactose permease Gal2p. In its active form,

Gal3p sequesters the Gal80p repressor to the cytoplasm, indirectly promoting transcription [65, 66]. As a result, Gal2p and Gal3p contribute to the network architecture by forming two positive feedback loops. Excepting the constitutive *GAL4* promoter, the activities of the different *GAL* pathway promoters are similar to each other [59] as they are regulated by the same transcriptional machinery.

3.3 Inducibility curve as a quantitative phenotype

To quantify the activity of the *GAL* pathway at the single-cell level, we used the yellow fluorescent protein (*YFP*) driven by the *GAL1* promoter as our reporter system and measured expression profiles at different galactose concentrations using flow cytometry. In order to do this KpnI-P_{GAL1}-BamHI and BamHI-*YFP*-EcoRI fragments were cloned into the pRS402 backbone [31,32] upstream of *CYC1* transcriptional terminator and then integrated into a W303 mat α strain, selecting for positive transformants in plates lacking adenine. A W303 mat a strain was tagged with a *HIS3* marker by transforming the pRS303 plasmid into it [31, 32]. Finally a diploid strain, that we will refer to as wildtype, was created by mating these two. The P_{GAL1} promoter sequence corresponds to the 669 base-pair region directly upstream of the start codon of the *GAL1* gene.

Cultures were grown in synthetic dropout media with the appropriate amino-acid supplements. Cells were first grown overnight during 20 hours in a 30 °C shaker, using 2 % $^{w}/_{v}$ raffinose as the carbon source. This overnight growth period was followed by an induction stage of 20 hours in a 30 °C shaker, with cultures now containing 0.1 % $^{w}/_{v}$ glucose and $0 - 0.4$ % $^{w}/_{v}$ galactose as carbon sources. After the induction period, the expression distributions were determined by flow cytometry (FACScan; Becton Dickinson). The densities of the cultures were kept low throughout the experiment ($OD_{600} < 0.33$ at the end of the induction period) to prevent nutrient depletion. The volume of all cultures was 10 ml during both the overnight growth and induction periods.

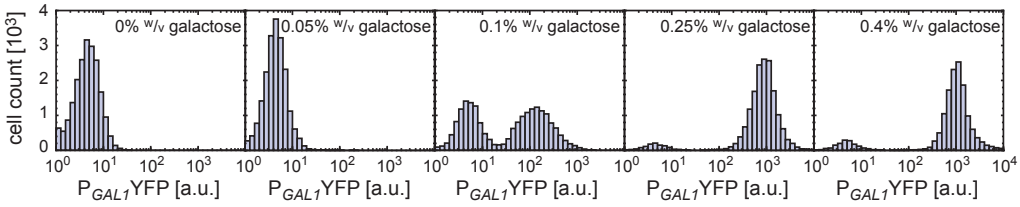

Figure 3.2: Histograms show the induction profile of the wildtype galactose pathway when the galactose concentration varies between 0 and 0.4 $\%^{w}/_{v}$. The activity of the pathway was read-out at the single-cell level using *YFP* driven by the *GAL1* promoter. At least 10,000 cells were analyzed by flow cytometry.

In Figure 3.2 we present histograms of the expression levels of the *YFP* gene across populations of cells grown in media with different levels of galactose, the inducer of the system. For intermediate concentrations it can be seen that only a fraction of the cells expresses this gene significantly whereas the remaining cells fail to express it beyond basal values. Measurements of the same distributions after 6 extra hours of culturing in the inducing conditions did not exhibit significant changes in the distributions. These observations indicate that the system is bistable [59].

Given that the expression peaks corresponding to the ON and OFF states are well separated in these histograms, it is possible to quantify the fraction of ON cells, *i.e.* the fraction of cells that are expressing the *YFP* gene off the *GAL1* promoter in a manner significantly different than that observed in cells grown in the absence of galactose. This was done manually by setting a threshold in between the two observed expression peaks.

We will refer to a sequence of the measured fractions of actively expressing cells for a series of increasing galactose concentrations as the inducibility curve of the system (Figure 3.3) and we will use this notion as our phenotype of interest. This phenotype describes the range of galactose concentrations at which the organism is able to turn on the galactose uptake system and therefore take advantage of this additional source of energy.

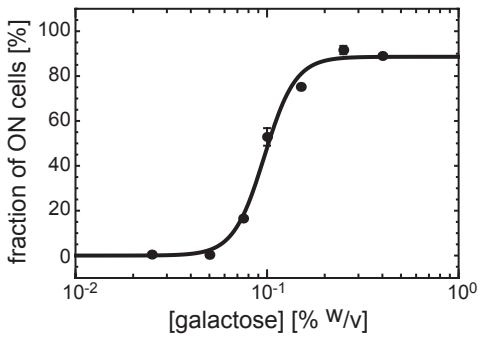

Figure 3.3: Inducibility curve: fraction of ON cells as a function of galactose concentration. The solid line is a guide to the eye constructed by fitting a Hill function to the data.

3.4 Effective model for describing the observations

We interpreted these experimental results in the context of an effective model in which the total concentrations of the different regulatory *GAL* proteins affect slow transitions between transcriptionally active (ON) and inactive (OFF) states [67]. We chose the functional dependence of these rates on the concentration of the regulatory proteins by using simple functional relationships reflecting the cascade of interactions among the network components. We restricted the corresponding transcription rates to obtain steady-state concentrations compatible with previous measurements of transcript levels in different environments [68, 69]. Finally, as an approximation to the dynamics of this model, we used a set of differential equations for describing the fraction of ON-cells under a given condition.

3.4.1 Model specification

We consider an effective stochastic model in which a given promoter site of each of the *GAL* regulated genes (*GAL2*, *GAL3* and *GAL80*) can be in either a state of active transcription (ON-state) or in a state in which transcription

occurs less often (OFF-state) [67,70]. Each of these states is characterized by its typical transcription rate. We chose to parameterize the system so that one copy of the *GALi* gene will produce the corresponding proteins at rate θ_i (which coarse-grains the processes of transcription, translation and protein folding) when in the ON-state and at rate $\lambda\theta_i$, when in the OFF-state. So λ represents the relative transcriptional strength of the OFF-state compared to the ON-state.

We consider that slow stochastic transitions between these transcriptional states are possible and that the total concentration of the different regulatory proteins affects the rate at which the OFF $\rightarrow$ ON transitions takes place.

$$\mathrm{OFF} \underset{h}{\overset{h\rho(x_2,x_3,x_4,x_{80})}{\rightleftarrows}} \mathrm{ON} \tag{3.1}$$

In this scheme the parameter h represents a typical timescale at which these transitions take place and ρ is a dimensionless function that quantifies how the total concentrations of the different *GAL* proteins (x_2, x_3, x_4, x_{80}) affect the rate of OFF $\rightarrow$ ON transitions. This description is valid as long as the molecular interactions that shape the regulating function ρ occur much more rapidly than the typical timescale at which protein concentrations change due to the processes of transcription, translation, and protein dilution/degradation.

We parameterized ρ by taking into account what is known about the way the different *GAL* proteins interact with each other and affect transcription. First, it is known that the *GAL4* protein is the main transcriptional activator when it is not bound by *GAL80* proteins, so we proposed the form

$$\rho = \left(\frac{x_4^\star}{K_4}\right)^\eta \tag{3.2}$$

where K_4 represents the effective typical concentration scale of the interaction, $\eta > 0$ is its typical effective nonlinearity, and $x_4^\star$ is the concentration of *GAL4* that is not bound by *GAL80* and can therefore freely activate transcription. Instead of writing a set of reactions for describing how $x_4^\star$ depends on the total concentrations of the GAL network proteins, we propose to use simple functional forms that effectively describe the main nature of the interactions (Figure 3.4).

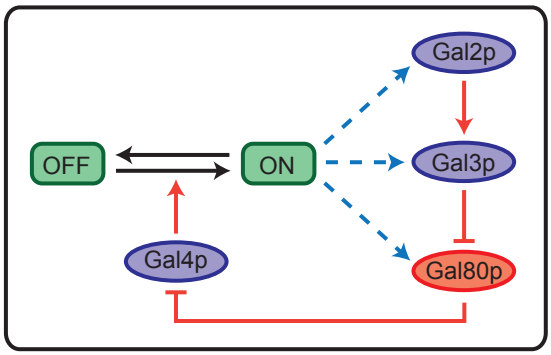

Figure 3.4: Simplified schematic of the GAL Network: ON-cells express (blue dotted arrows) both the positive (Gal2p and Gal3p) and the negative (Gal80p) regulators of the network while the OFF-cells gene expression does not exceed basal levels. Gal4p is a constitutively expressed protein. Red arrows with pointed and blunted ends reflect the rapid positive and negative effects of one network component on another, respectively.

In these functional forms, the molecular interactions are characterized by typical concentration scales of action for each protein and by typical degrees of nonlinearity quantified by positive exponents. In this case, we know that the amount of free *GAL4* proteins will be a decreasing function of the concentration of GAL80 proteins in the nucleus and an increasing fraction of the total concentration of *GAL4* proteins. Therefore, we propose to use the form

$$x_4^\star = \frac{x_4}{1 + \left(\frac{x_{80}^\star}{K_{80}}\right)^\beta} \tag{3.3}$$

where $x_{80}^\star$ is the concentration of *GAL80* proteins in the nucleus. This quantity, in turn, is regulated by the active *GAL3* proteins due to sequestration

$$x_{80}^\star = \frac{x_{80}}{1 + \left(\frac{x_3^\star}{K_3}\right)^\alpha} \tag{3.4}$$

where $x_3^\star$ is the concentration of active *GAL3* proteins. The internal galactose concentration, $g^\star$, regulates the activation of *GAL3* proteins and therefore

we propose to write

$$x_3^\star = \frac{x_3}{1 + \left(\frac{g^\star}{K_g}\right)^{-\nu}} \tag{3.5}$$

Note that in this case the number of active *GAL3* proteins is an increasing function of the concentration of internal galactose because we assume $\nu > 0$. Finally, the concentration of internal galactose is regulated by the concentration of galactose with which the cells were grown, g, and the amount of *GAL2* proteins (the galactose permease) and so we write

$$g^\star = \frac{g}{1 + \left(\frac{x_2}{K_2}\right)^{-\mu}} \tag{3.6}$$

Equations (3.2) to (3.6) describe how the rate of the OFF $\rightarrow$ ON transitions is regulated by the total concentrations of the different proteins involved as well as the concentration of external galactose.

To finalize the specification of the model, we also assumed that protein degradation rates were slow compared to the growth rate of the organism and so we only included the effect of the dilution of proteins at rate γ, the average growth rate of yeast in the laboratory.

We simulated this system by using a custom-written C++ implementation of the Gillespie algorithm [71], which considered the production of proteins, their dilution due to cell growth, and the transitions between the transcriptional states as first-order stochastic reactions. In Figure 3.5 we show sample trajectories for the different variables in a simulation corresponding to the parameters reported in Sections 3.4.2 and 3.5.2.

3.4.2 Constraints on model parameters

To keep our model realistic, we constrained the values of several parameters to previously measured quantities. However, some quantities (especially the effective parameters we introduced) were difficult to estimate based on published work and therefore we extracted them out by fitting the model to some of our data.

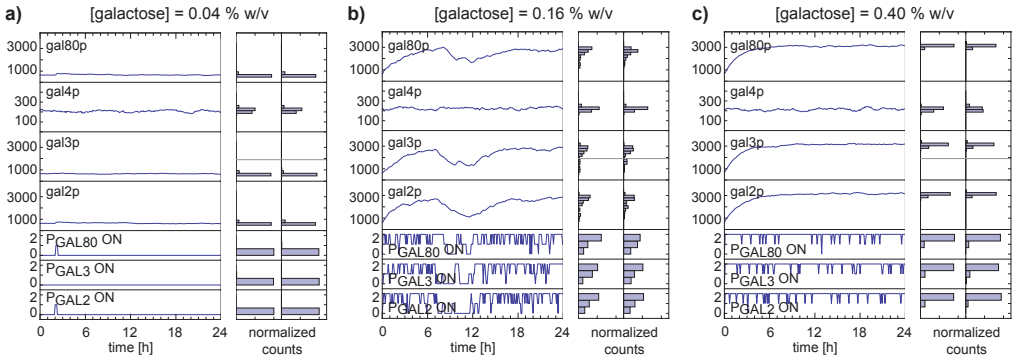

Figure 3.5: Stochastic simulations of the proposed model. Left sub-panels: traces of the different variables (P_{GALi} ON: number of promoters in an ON state; gal$i$p: protein concentrations) corresponding to one realization of the stochastic model proposed in section 3.4.1 for a time of 24 h. Right sub-panels: Distribution of the different variables after 24 h (left) and 48 h (right) across 100 independent realizations. The fact that these distributions are similar to each other indicates that the process has reached a steady state. In all cases, simulations were started from initial conditions corresponding to an OFF state, namely: all promoters were OFF and the initial protein concentrations were chosen as $\theta_i\lambda/\gamma$ for *GAL2*, *GAL3* and *GAL80* and as θ_4/γ for *GAL4*. The parameters used were those indicated in Sections 3.4.2 and 3.5.2; the value of h used in this set of simulations was $h = 2.5\,\mathrm{h}^{-1}$.

On one hand, the doubling time of yeast in the environments used in this study is about 90 minutes, which imposes the constraint $\gamma \simeq 0.46\,\mathrm{h}^{-1}$.

Previous high-throughput studies identified fold-differences in transcript levels for several yeast genes under different growth conditions [69]. More specifically, yeast cells were grown in two separate environments, one promoting the expression of the *GAL* genes and one repressing it. Average differences of about 5.5-fold and 3.7-fold were reported for *GAL3* and *GAL80*, respectively [69]. A high-throughput study that quantified the amount and localization of different yeast proteins reported that there were about 800 *GAL3* and 700 *GAL80* proteins per cell when the *GAL* genes were repressed [68], which in the context of the proposed model would correspond

Parameter	Value
θ_2	1500 proteins/h
θ_3	1500 proteins/h
θ_4	100 proteins/h
θ_{80}	1500 proteins/h
λ	0.2
γ	$0.46\,\mathrm{h}^{-1}$

Table 3.1: Parameters fixed based on previous observations.

to the situation in which these genes are in the OFF transcriptional state. Considering these observations and with the aim of simplifying the description further, we assumed that all *GAL*-regulated genes in the network follow a similar regulation scheme and fixed the values of the parameters θ_3, θ_{80} and λ so as to obtain basal expression levels of about 750 proteins per cell and a 5-fold increase in protein levels when the network is fully induced. So, taking the dilution rate into account, this implies the constraints $\theta_3 \simeq \theta_{80} \simeq 1725\,\mathrm{proteins/h}$ and $\lambda \simeq 0.2$. Related experimental evidence for *GAL2* is more elusive and for the sake of simplicity we assumed the same transcription rate as for *GAL3* and *GAL80*. The same studies also reported that the level of *GAL4* transcripts does not change significantly between galactose-free and galactose-rich media [69] and that *GAL4* proteins are present at a concentration of about 200 proteins per cell [68], which implies the constraint $\theta_4 \simeq 92\,\mathrm{proteins/h}$.

We note that in the model at hand the stochasticity in the change of the transcriptional plan is described by slow transitions between two different states. In this framework, the fluctuations in protein expression levels play a secondary role in establishing the fraction of active cells under a given condition, though they still play a major role in shaping the distributions associated with OFF and ON expression states. Furthermore, the way we set up the regulation scheme indicated by equations (3.2)-(3.6) allows us to interpret the constants K_i as being expressed in terms of the typical concentrations of the associated proteins. Therefore, we don't lose generality by using parameter values θ_i that might deviate slightly from experimental measurements.

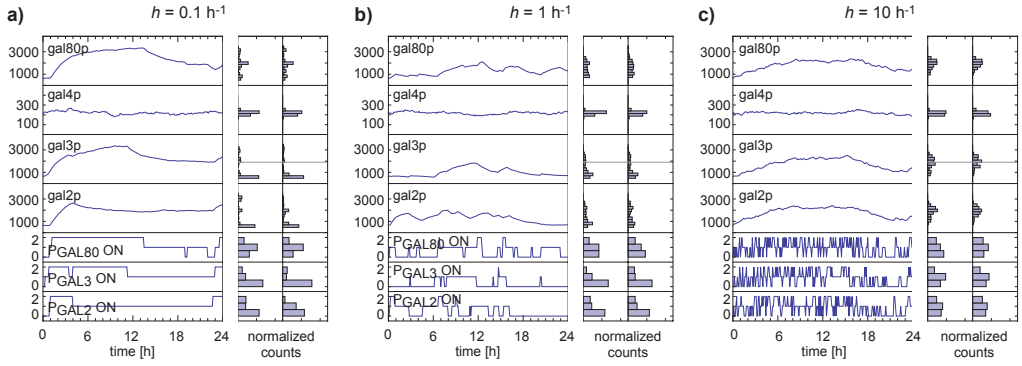

Figure 3.6: Simulations similar to those described in Figure 3.5 but for a fixed galactose concentration of $0.1\,\%^{w}/_{v}$ and three different values of the parameter h, which quantifies the timescale of the stochastic transitions.

The values of the parameters used in simulations and fits presented in this study and constrained along these guidelines can be found in Table 3.1.

In order to obtain OFF and ON expression states that are well-separated from each other, the time that it takes for protein levels to equilibrate has to be shorter than the typical timescale of the transitions between the two transcriptional states. We quantified the timescale of transitions through the parameter h in equation (3.1). Exploring a range of values for this quantity, we found reasonable agreement with the experimental results for $h = 2.5\,\mathrm{h}^{-1}$. If the value of h is too high (Figure 3.6c) the distribution of protein numbers becomes monomodal; on the other hand if this parameter is too low (Figure 3.6a) the dynamics of establishment of fractions would be too slow compared to the experimental observations and in the case of multiple promoters it would lead to the appearance of three distinct expression states, which is something that is not observed experimentally. We also note that all inferences presented in this work are based on the analytical approximation described in the next section, where the exact value of the parameter h becomes immaterial.

3.4.3 Analytical approximation

Stochastic simulations of the model described above are computationally time consuming. To simplify the exploration of parameters and/or alternative models, we developed an approximation for the steady-state fraction of actively transcribing cells in a macroscopic population. We note that the presentation of the approximation proposed here is not a rigorous derivation. We based it on intuition and heuristic observations, and we eventually confirmed its power by comparing inferences drawn from it to those obtained from detailed stochastic simulations.

For one cell, we can approximate the time-evolution of the number of proteins associated with each GAL-network-regulated gene with a set of Langevin equations [38, 39, 58] of the form

$$\dot{x}_i = \theta_i \left[\phi + \lambda(1-\phi)\right] - \gamma x_i + \xi_i \tag{3.7}$$

where x_i represents the concentration of the protein associated with the GALi gene, ϕ is a random binary variable that indicates whether the cell is transcribing or not, and ξ_i is a random variable that approximates the intrinsic stochasticity associated with the processes of protein production and dilution.

An equation for the evolution of mean protein numbers across a population of cells, $\langle x_i \rangle$, can be obtained by averaging equation (3.7) above. If we assume that the intrinsic noise in protein expression can be neglected and we consider a steady state situation, we obtain a set of equations of the form

$$0 = \theta_i \left[\lambda + (1-\lambda)\langle\phi\rangle\right] - \gamma\langle x_i \rangle \tag{3.8}$$

where $\langle\phi\rangle$ represents the fraction of cells that are actively transcribing. Following a mean-field approximation approach, we estimate the fraction $\langle\phi\rangle$ with the value that we would infer from assuming a constant background of

protein concentrations equal to their average values, namely

$$\begin{aligned}\langle\phi\rangle &\simeq \left\langle \frac{k_{\mathrm{OFF}\to\mathrm{ON}}(x_2,x_3,x_4,x_{80})}{k_{\mathrm{OFF}\to\mathrm{ON}}(x_2,x_3,x_4,x_{80})+k_{\mathrm{ON}\to\mathrm{OFF}}}\right\rangle \\ &\simeq \frac{1}{1+\frac{k_{\mathrm{ON}\to\mathrm{OFF}}}{k_{\mathrm{OFF}\to\mathrm{ON}}(\langle x_2\rangle,\langle x_3\rangle,\langle x_4\rangle,\langle x_{80}\rangle)}} \\ &= \frac{1}{1+\left[\rho\left(\langle x_2\rangle,\langle x_3\rangle,\langle x_4\rangle,\langle x_{80}\rangle\right)\right]^{-1}} \\ &\equiv f\left(\langle x_2\rangle,\langle x_3\rangle,\langle x_4\rangle,\langle x_{80}\rangle\right)\end{aligned} \tag{3.9}$$

where we have explicitly incorporated the parameterization proposed in equation (3.1) and where we have assumed that the different copies of each promoter act in a correlated way due to the effect of the different proteins involved.

Taking into account that *GAL4* is not subject to regulation, the argument above implies that, in order to obtain a self-consistent solution, the following set of algebraic equations must be satisfied:

$$\begin{cases} 0 = \theta_2\left[\lambda+(1-\lambda)f(x_2,x_3,x_4,x_{80})\right]-\gamma x_2 \\ 0 = \theta_3\left[\lambda+(1-\lambda)f(x_2,x_3,x_4,x_{80})\right]-\gamma x_3 \\ 0 = \theta_4-\gamma x_4 \\ 0 = \theta_{80}\left[\lambda+(1-\lambda)f(x_2,x_3,x_4,x_{80})\right]-\gamma x_{80}\end{cases} \tag{3.10}$$

where we have dropped the angled brackets to simplify notation. We note that these equations imply that in equilibrium the concentrations of *GAL2*, *GAL3*, and *GAL80* will be proportional to each other through their relative transcriptional strengths,

$$\frac{x_2}{\theta_2}=\frac{x_3}{\theta_3}=\frac{x_{80}}{\theta_{80}} \tag{3.11}$$

which allows us to reduce this set of relations to just one equation:

$$0=\theta_3\left[\lambda+(1-\lambda)f\left(\tfrac{\theta_2}{\theta_3}x_3,x_3,\tfrac{\theta_4}{\gamma},\tfrac{\theta_{80}}{\theta_3}x_3\right)\right]-\gamma x_3 \tag{3.12}$$

Solving this equation for x_3 and then computing $f\left(\frac{\theta_2}{\theta_3}x_3,x_3,\frac{\theta_4}{\gamma},\frac{\theta_{80}}{\theta_3}x_3\right)$ allows us to obtain an approximation for the fraction of actively transcribing cells in a given population. For instance, we solved this equation by bisection search considering as extrema the minimum and maximum possible

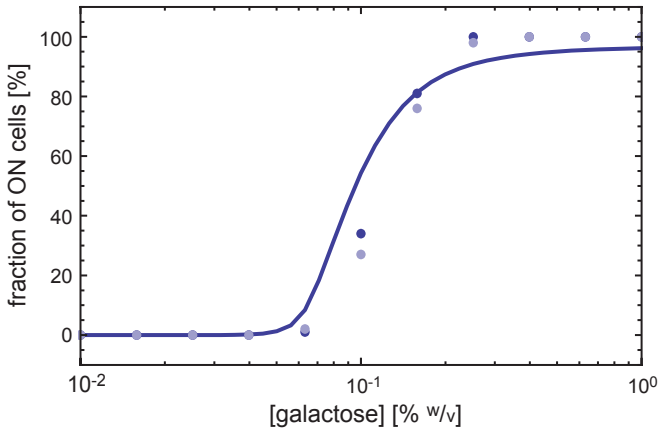

Figure 3.7: Fraction of ON cells according to the analytical approximation described in Section 3.4.3 (solid line) and the stochastic simulations described in Section 3.4.1 (dots). In the case of the stochastic simulations, the fraction of ON cells was determined by dividing the *GAL3* expression profile in two regions separated by the gray line shown in the right sub-panels of Figures 3.5 and 3.6 and counting the fraction of cells in the region that corresponds to higher expression levels. That gray line lies exactly in between the maximal and basal expression levels. Blue dots are the fractions obtained from 100 simulations at 24 h and the gray dots correspond to the results at 48 h.

GAL3 concentrations that can be achieved under this scheme ($\lambda\theta_3/\gamma$ and θ_3/γ respectively).

In Figure 3.7 we compare the results obtained from detailed simulations of the stochastic process specified in Section II to the deterministic approximation proposed in this section for a range of galactose values that includes those used in the experiments presented in this article. We observe reasonable agreement, which supports the usefulness of the approximation proposed as a proxy for studying the behavior of the system in a manner that is less taxing from the computational point of view.

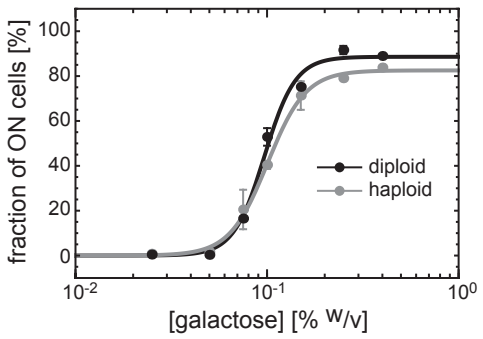

Figure 3.8: Fraction of ON cells as a function of galactose concentration for both diploid and haploid strains. The strains display similar induction profiles. The solid lines are guides to the eye constructed by fitting a sigmoidal function to the data.

3.5 Effect of gene dosage on inducibility profiles and network-dosage invariance

We observed similar inducibility profiles between haploid and diploid strains that contain the same reporter system (Figure 3.8), demonstrating that the system is invariant to ploidy changes. This indicates that the system might be built in such a way that parallel changes in copy numbers are intrinsically compensated or, in this case, it could be possible that the compensation is due to a volume-mediated effect. As the ploidy of yeast increases the rate at which genes are transcribed goes up as there are more gene copies available for transcription but, at the same time, the volume of the cells also goes up, diluting transcripts and proteins. One hypothesis is that these two mechanisms, having opposing effects on the steady state concentrations of regulatory agents, might be what underlies the observed network-dosage compensation.

To dissect how network-dosage variations affect the inducibility of the network in the absence of volume effects, we systematically reduced the number of copies of the 4 regulatory genes in the *GAL* network from 2 to 1 in diploid backgrounds by using *KanMX4* and *NatMX4* gene deletion cassettes [48, 72]

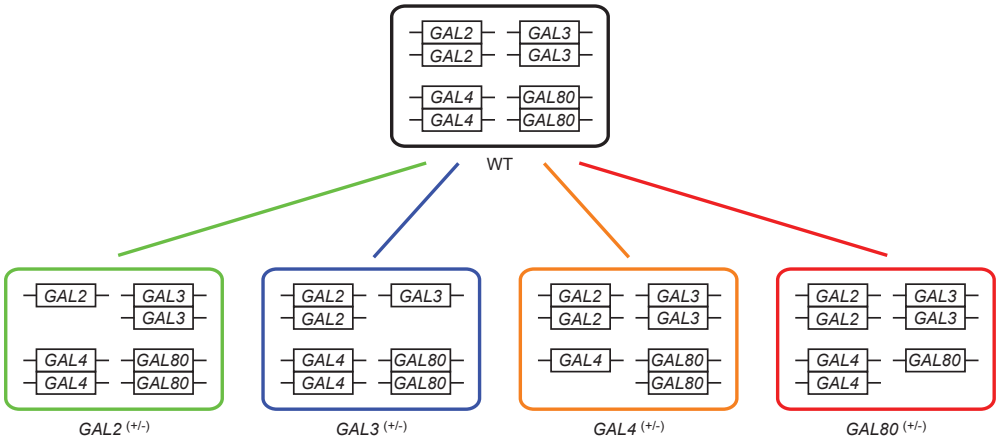

Figure 3.9: Construction scheme of the first-order dosage-varied yeast strains. Each rectangle denotes a diploid strain that was dosage-halved in one of the four regulatory genes of the network.

(Figure 3.9), obtaining 16 different diploid yeast strains including the hemizygous and the wildtype strains that have all 4 genes at one and two copies, respectively.

3.5.1 Effect of removing one copy of each gene

Figures 3.10a,b show how the wildtype inducibility levels are affected by the dosage of each regulatory gene in the network. Halving the dosage of *GAL3* dramatically reduces the inducibility of the system whereas halving the dosage of *GAL80* makes the cells need less galactose to reach full induction (Figure 3.10a). Interestingly, varying *GAL2* or *GAL4* dosage levels turned out not to have a large effect on network activity (Figure 3.10b).

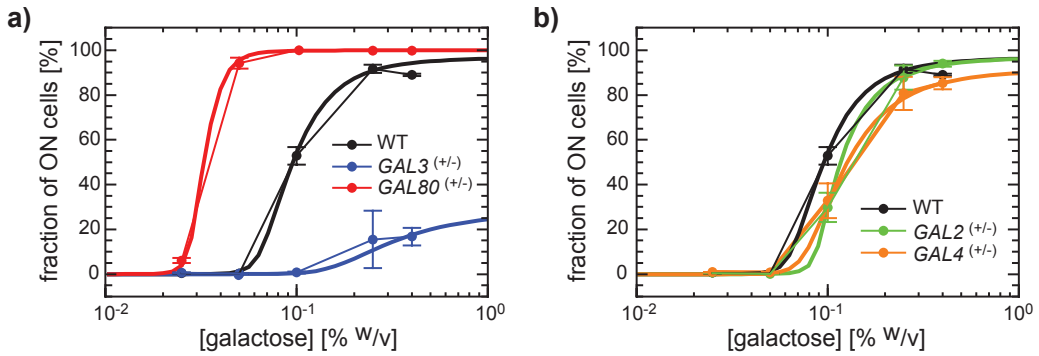

Figure 3.10: **a**, Inducibility profiles of the *GAL* network hemizygous in *GAL3* (blue) or *GAL80* (red) relative to the wildtype profile (black). Network inducibility is highly sensitive to *GAL3* and *GAL80* dosage. **b**, Inducibility profile of the *GAL* network hemizygous in *GAL2* (green) or *GAL4* (orange) relative to the wildtype profile (black). Network inducibility is almost neutral to *GAL2* and *GAL4* dosage. In both a and b, the thick solid lines represent the best fit of the model described in section 3.4 to the 5 different inducibility profiles shown in these two figures.

3.5.2 Fitting procedure to the analytical model and best fit results

We determined the set of parameters that best describes the data by confronting the measurements with the predictions of the model described in section 3.4 using a Bayesian inference approach [49]. Briefly, we assumed that for a given set of parameters, the likelihood of observing each measurement follows a normal distribution centered on the value indicated by the model and with an estimated uncertainty of 10% which is representative of the repeatability of the experiments. Applying Bayes theorem, this defines a distribution over parameter space where each parameter set gets weighed according to its likelihood of representing the data. We sampled this distribution using a Metropolis-Hastings algorithm, which allowed us to obtain estimates of the parameter set that has the highest likelihood of being a good description of the data as well as corresponding uncertainties.

Parameter		Sampling parameters			
symbol	unit	starting point	jump width	min	max
K_2	proteins	1000	200	0.1	4000
K_g	proteins	0.03	0.01	0.002	1
K_3	proteins	2.0	0.5	0.1	10
K_{80}	proteins	8.0	0.5	0.3	200
K_4	proteins	1.0	2.5	0.1	200
μ	-	0.50	0.25	0.05	20
ν	-	1.0	0.20	0.05	20
α	-	0.85	0.02	0.04	20
β	-	5.0	0.5	0.05	50
η	-	1.5	0.2	0.05	20

Table 3.2: Summary of the parameters used in the fitting procedure. Starting point indicates the initial value used in the Metropolis Hastings procedure [49]. "Jump width" is the standard deviation of the normal distribution used to create new targets and "min" and "max" represent imposed hard boundaries beyond which sampling was not allowed.

The sampling algorithm was run by following 10 independent Markov chains starting from the point indicated in Table 3.2. For each parameter, normal distributions with widths as indicated in the same table and centered in the previous point were used as jump distributions. We also imposed lower and upper bounds as indicated in that table but the chains stayed away from the boundaries except in the case of K_2, which we relate to the fact that the experimental system does not exhibit much sensitivity to changes in the dosage of *GAL2* in the conditions explored. Each chain was followed for 10,000 iterations and only the second half of the simulations was used to draw inferences.

In Figure 3.11, we show the inferred distributions for each parameter and in Table 3.3 we report first order statistics that describe the inferred values for each fit parameter.

Parameter		Inferences		
symbol	unit	mode	mean	standard deviation
K_2	proteins	600	1700	1000
K_g	proteins	0.052	0.040	0.008
K_3	proteins	4.1	2.9	0.7
K_{80}	proteins	8	8	2
K_4	proteins	2	8	6
μ	-	0.9	0.6	0.2
ν	-	1.3	1.5	0.2
α	-	0.85	0.82	0.02
β	-	6.1	6.6	0.7
η	-	1.6	1.8	0.2

Table 3.3: Summary of the inferred parameters statistics.

3.5.3 Combinatorial exploration of gene dosage variations

To explore the degree of dosage compensation in the *GAL* network, we measured the inducibility profiles of all 16 different strains, grouped the measurements in 4 different dosage-perturbation orders depending on how many genes had their dosage halved, and compared the profiles to one another (Figures 3.12). The strongest second-order compensation was observed when both *GAL3* and *GAL80* were dosage-perturbed in the same strain. At the third-order, irrespective of the second-order genetic background on which the dosage of a third gene was reduced, halving the dosage of *GAL3* (*GAL80*) always decreased (increased) the inducibility levels. On the other hand, varying the dosage of *GAL2* had a neutral effect on inducibility and halving the dosage of *GAL4* decreased it slightly. The fit to our model based on the measurements on the wildtype and first-order dosage-perturbed strains (solid lines in Figure 3.10) reasonably predicts how most higher order dosage perturbations to the *GAL* network affect the cellular inducibility profiles (Figure 3.12) indicating the usefulness of the chosen modeling approach as a framework for describing the observations.

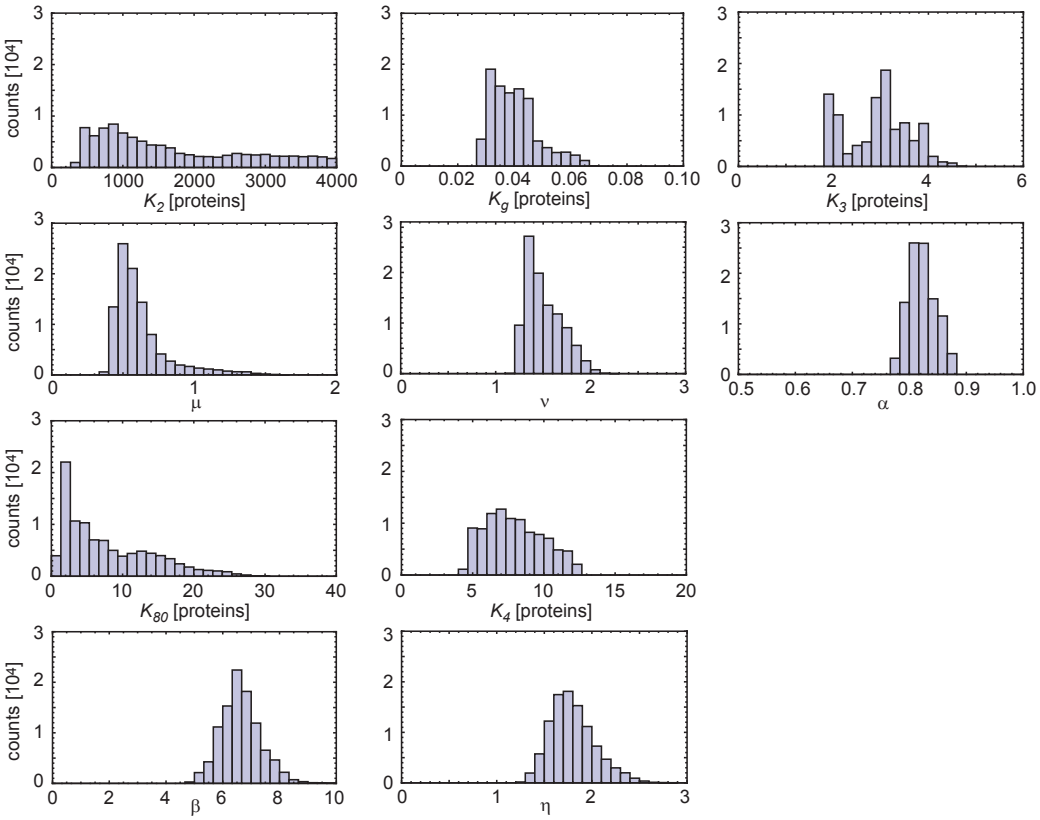

Figure 3.11: Metropolis-Hastings samples of parameter values according to their likelihood of describing the experimental data, following a Bayesian-inference approach [49].

3.5.4 Network-dosage invariance

We uncovered the level of network-dosage compensation in the system when we analyzed the inducibility profile of the fourth-order hemizygous strain in comparison to the one of the wildtype strain. We observed similar inducibility levels for the two strains, implying the presence of network-dosage invariance in the *GAL* network even in the absence of volume-mediated compensation effects (Figure 3.13).

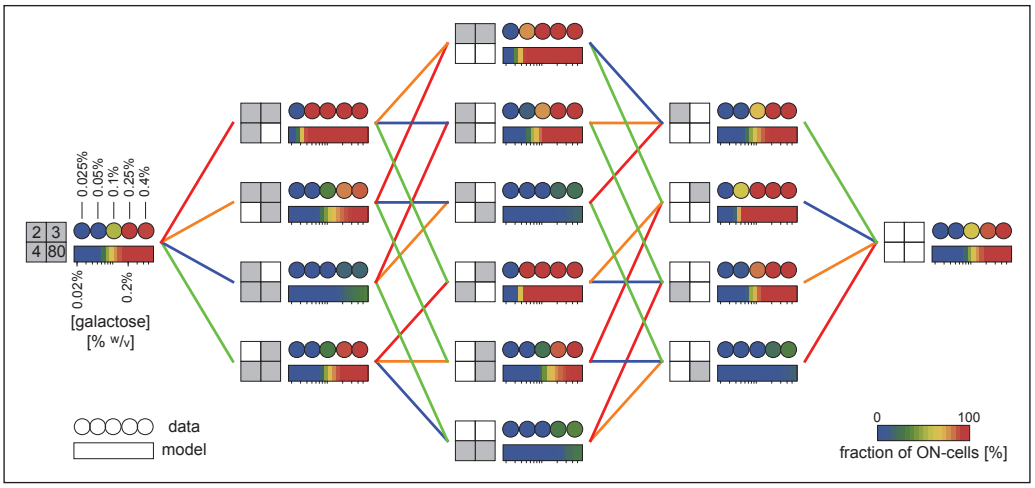

Figure 3.12: Systematic dosage variations and network-dosage compensation. The color of each filled circle represents the network inducibility level for a specific galactose concentration ($0.025 - 0.4\,\%^{w}/_{v}$). Inducibility is quantified by measuring the fraction of ON-cells in a bimodal population. The rectangular, color-coded bars reflect the predictions of the model based on the best fit to the data presented in Figure 3.10. The genetic background of each strain is specified by a square at its immediate left. Each big square contains four subsections that represent the four regulatory genes of the *GAL* network (top-left: *GAL2*, top-right: *GAL3*, bottom-left: *GAL4*, bottom-right: *GAL80*). Grey (white) color marks the presence of two (one) copies of a specific gene. A line between two strains indicates that the two genetic backgrounds differ by a single copy of a specific gene and the color of the line codifies that gene (blue for *GAL3*, red for *GAL80*, green for *GAL2*, and orange for *GAL4*).

3.5.5 Contribution of each gene

To show, in a concise fashion, the contribution of each regulatory gene in affecting wildtype inducibility levels, we quantified the average contribution of the second copy of each regulatory gene to network inducibility by calculating two quantities associated with each gene. We computed both the mean deviation (Δ) and mean squared deviation (χ^2) of the fraction of ON

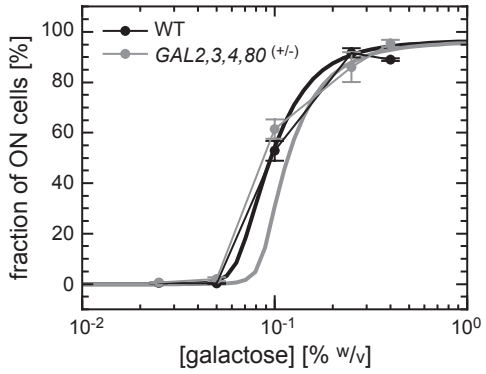

Figure 3.13: Similarity between the inducibility profiles of the wildtype strain (black) and the strain containing one copy of each regulatory gene (gray): the inducibility curve is robust to the variation in network dosage. The thick solid lines represent model predictions.

cells at a given galactose concentration after having halved the copy number of a given gene, averaging over all the galactose concentrations and genetic backgrounds. Figure 3.14 depicts the importance of *GAL3* as an activator and *GAL80* as an inhibitor over the relatively smaller contributions of *GAL2* and *GAL4* on the inducibility profiles.

The fact that only two of the regulatory genes produce a significant shift in phenotype when their dosage is altered is reminiscent of what was observed in the simpler system discussed in Chapter 2 in the sense that it is the operating point of the network what determines whether a quantitative change in the rate of production of a gene will have a significant impact on phenotype or not. In the case at hand one could say that, in some sense, *GAL2* and *GAL4* are operating in a regime of saturation and therefore the network output is not affected if their dosage is modified. Interestingly, these genes are known to be required for the network to be functional: if *GAL4* is knocked out the galactose system cannot turn on [73] and a full deletion of *GAL2* leads to a significant decrease in network inducibility [74]. This indicates that the property we observed is only local to the state around which the network is operating.

Regarding the notion of network-dosage invariance these experimental ob-

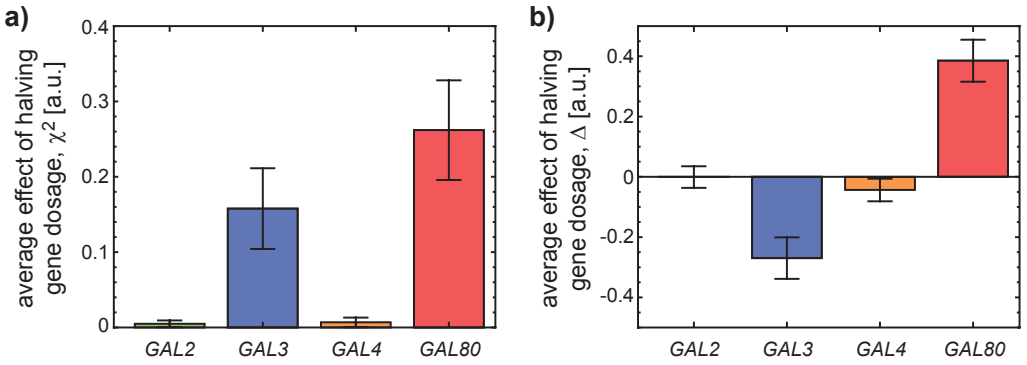

Figure 3.14: **a**, Contribution of the second copy of each regulatory gene to network inducibility quantified as the squared difference (χ^2) between inducibility levels in strains that differ in one copy of the corresponding gene and averaged across different galactose concentrations and genetic backgrounds. **b**, Signed contribution of the second copy of each regulatory gene to network inducibility quantified as the signed difference (Δ) between inducibility levels in strains that differ in one copy of the corresponding gene and averaged across different galactose concentrations and genetic backgrounds.

servations suggest that it may be possible to build a network dosage invariant phenotype into a gene network by using only 2 components. However, up to this point, it is not clear whether the specific wiring topology of the network components would also play a role in this or not.

3.6 Minimal conditions required for network-dosage invariance

In order to investigate the necessary features that can make natural gene networks display dosage invariance, we considered a set of genes subject to a common regulation scheme and set to address the general question of what conditions on the regulation scheme guarantee would the activity of the transcriptional center to be invariant to proportional changes in the transcription

rates of all the genes involved. Such changes would be produced as an organism undergoes a change in ploidy, as chromosomes are replicated throughout cell cycle, in genome-wide duplication or loss events and/or by global noise in the expression of transcription factors.

3.6.1 Analysis of generic systems

We consider that each gene is transcribed and then translated proportionally to the activity of its transcription center (a number between 0 and 1 that might represent, for instance, the fractional occupancy of active promoter sites), the proportionality constant being the maximal transcriptional rate associated with the gene. We assume that all proteins are effectively degraded at the cell-division rate (γ), thinking about a situation in which the lifetime of the proteins is much longer than the cell-division time. We further consider that proteins generated off each gene interact with each other on fast timescales and that this interaction defines the state of the transcriptional center. Finally, we consider a mean field approximation in the sense defined in Section 3.4.3.

Under these conditions, we describe the time evolution of the concentrations of the relevant proteins by the following set of differential equations:

$$\begin{cases} \dfrac{\mathrm{d}x_1}{\mathrm{d}t} = \theta_1 f(\rho, x_1, \ldots, x_N) - \gamma x_1 \\ \qquad \vdots \\ \dfrac{\mathrm{d}x_N}{\mathrm{d}t} = \theta_N f(\rho, x_1, \ldots, x_N) - \gamma x_N \end{cases} \tag{3.13}$$

In these equations, x_i represents the average total concentration of the protein coded by the i-th gene, θ_i is the transcriptional strength associated with it, γ is the dilution rate, ρ is some external control parameter (*v.g.*, galactose in the case of *GAL* network) and $f(\rho, x_1, \ldots, x_N)$ is a dimensionless quantity that takes values in [0, 1] and represents the activity level of the transcriptional system under consideration given a background of protein concentrations $x_1, \ldots, x_N$. We can think that f represents the fraction of active promoter sites. Under the context of this framework, one approach to

formalize the question we are interested in addressing is to ask: what family of functions f representative of biochemical interactions describe systems in which the steady state value of f is invariant to proportional changes in all the θs for a wide range of the control parameter ρ (*i.e.*, exploring the full range of values of f, so as to get an inducible system).

One-dimensional case

Let's first consider the simplest possible case: a network with just one gene. At steady state, we have

$$\theta f(\rho, x) = \gamma x \tag{3.14}$$

Mathematically, in order for a function to be invariant with respect to a variable, its derivative with respect to that variable has to be zero. Taking the derivative of the above expression with respect to θ we get

$$f + \theta \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\theta} = \gamma \frac{\mathrm{d}x}{\mathrm{d}\theta} \tag{3.15}$$

from where we obtain that

$$\frac{\mathrm{d}f}{\mathrm{d}\theta} = \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\theta} = \frac{f\frac{\partial f}{\partial x}}{\gamma - \theta\frac{\partial f}{\partial x}}. \tag{3.16}$$

We conclude that for the system to be invariant for nontrivial fractions ($f \neq 0$) we need $\frac{\partial f}{\partial x} = 0$ at the value of x that solves the steady state equation. But if the system is to be inducible, we should assume that the value of x will change as we change ρ and therefore in order to get invariance across a range of induction conditions we need f to be independent of x, *i.e.* we cannot have feedback at all.

This means that the only possible way of getting an invariant system with just one species is if the system is not auto-regulated, which makes the situation trivial: if the state of a promoter is not affected by the proteins it codes for, its fractional occupancy will be invariant to changes in its transcriptional strength. This situation corresponds to the case of a constitutively regulated gene. Having more copies of that gene in the cell is not expected to impose any change in the state of its constitutive promoter.

Two-dimensional case

Now we consider a network composed of two genes. The system under consideration is represented by the following set of differential equations:

$$\begin{cases} \dfrac{\mathrm{d}x_1}{\mathrm{d}t} = \theta_1 f(\rho, x_1, x_2) - \gamma x_1, \\ \dfrac{\mathrm{d}x_2}{\mathrm{d}t} = \theta_2 f(\rho, x_1, x_2) - \gamma x_2 \end{cases} \tag{3.17}$$

and we are interested in studying how the value of f in steady state will be affected by proportional changes in θ_1 and θ_2.

Let's first note that using the same regulation scheme for the two genes imposes the condition that at steady state we must have (all variables represent steady state values from now on)

$$\frac{\theta_1}{\theta_2} = \frac{x_1}{x_2} \tag{3.18}$$

which implies that whatever change x_1 might undergo, x_2 is going to suffer a proportional modification as well. To study system behavior with respect to proportional changes in θ_1 and θ_2, we introduce an additional parameter δ in the following way:

$$\begin{cases} (1+\delta)\theta_1 f(\rho, x_1, x_2) = \gamma x_1, \\ (1+\delta)\theta_2 f(\rho, x_1, x_2) = \gamma x_2, \end{cases} \tag{3.19}$$

which allows us to vary the transcriptional rates in a proportional manner and to explore how the value of f is affected by such changes.

Taking derivatives of both sides of (3.19) with respect to δ, we obtain

$$\theta_1 f + (1+\delta)\theta_1 \left[\frac{\partial f}{\partial x_1}\frac{\mathrm{d}x_1}{\mathrm{d}\delta} + \frac{\partial f}{\partial x_2}\frac{\mathrm{d}x_2}{\mathrm{d}\delta} \right] = \gamma \frac{\mathrm{d}x_1}{\mathrm{d}\delta}. \tag{3.20}$$

Using equation (3.18) relating x_1 to x_2 at steady state, we can write $\frac{\mathrm{d}x_2}{\mathrm{d}\delta} = \frac{\theta_2}{\theta_1}\frac{\mathrm{d}x_1}{\mathrm{d}\delta}$ and plugging this expression into (3.20) we can solve the resulting equation for $\frac{\mathrm{d}x_1}{\mathrm{d}\delta}$:

$$\frac{\mathrm{d}x_1}{\mathrm{d}\delta} = \frac{\theta_1 f}{\gamma - (1+\delta)\left(\theta_1 \frac{\partial f}{\partial x_1} + \theta_2 \frac{\partial f}{\partial x_2}\right)} \tag{3.21}$$

where everything is evaluated at steady state. This implies that the change in f due to some small change in δ is proportional to

$$\frac{\mathrm{d}f}{\mathrm{d}\delta} = \frac{\partial f}{\partial x_1}\frac{\mathrm{d}x_1}{\mathrm{d}\delta} + \frac{\partial f}{\partial x_2}\frac{\mathrm{d}x_2}{\mathrm{d}\delta} = \frac{\left(\theta_1 \frac{\partial f}{\partial x_1} + \theta_2 \frac{\partial f}{\partial x_2}\right) f}{\gamma - (1+\delta)\left(\theta_1 \frac{\partial f}{\partial x_1} + \theta_2 \frac{\partial f}{\partial x_2}\right)}. \tag{3.22}$$

We conclude that for the system to be invariant with generality we need to satisfy

$$\theta_1 \frac{\partial f}{\partial x_1} + \theta_2 \frac{\partial f}{\partial x_2} = 0 \tag{3.23}$$

at steady state, but this implies that the signs of $\frac{\partial f}{\partial x_1}$ and $\frac{\partial f}{\partial x_2}$ have to be different; *i.e.* we need one activator and one inhibitor.

Therefore, a gene circuit with two components that are regulated by the same transcriptional machinery requires components of opposite sign for the activity of the system to be invariant to network dosage. Contrary to the one-dimensional case, the genes here do not have to give up their feedback regulation schemes. This describes a minimal condition necessary to build dosage-invariant phenotypes into gene networks.

3.6.2 Topology requirements on two-dimensional systems

To further explore if certain wiring topologies of 2-component generic network configurations would make it easier or harder for the cells to display dosage invariance, we performed numerical investigations on the possible network topologies in which an activator and an inhibitor are controlled by similar transcriptional machineries and analyzed their inducibility properties (Figure 3.15).

In the context of the proposed modeling framework, each interaction topology is represented by a 4-parameter functional form defining the relationship between the fraction of transcriptionally active cells (f) and the total concentrations of the activating (a) and inhibiting (i) agents. We consider that each one of these agents has a typical scale of action related to

parameters S_a and S_i and an effective nonlinearity around that point quantified by parameters α and β following the functional forms described below. We also consider the presence of an external inducing agent, g, that essentially affects the scale of action of the activator.

In the network configuration at the left of Figure 3.15a the activator indirectly activates transcription by regulating the effective activity of the repressor, which directly inhibits transcription, giving rise to the functional form

$$f = \frac{1}{1 + \left[\frac{S_i i}{1+(S_a g a)^{\alpha}}\right]^{\beta}}. \tag{3.24}$$

The activator could in principle directly enhance transcription, giving rise to an alternative network topology as depicted in the middle panel of Figure 3.15a. In this case, the positive and the negative feedback loops run in parallel to each other and there is no interaction between the activator and inhibitor. The corresponding functional relationship could be parameterized as

$$f = \left[\frac{1}{1 + (S_a g a)^{-\alpha}}\right] \left[\frac{1}{1 + (S_i i)^{\beta}}\right]. \tag{3.25}$$

Yet another network configuration would be achieved if the inhibitor gave up its direct repressor role and the activator assumed a direct activator function (Figure 3.15a, right panel), leading to

$$f = \frac{1}{1 + \left[\frac{S_a g a}{1+(S_i i)^{\beta}}\right]^{-\alpha}}. \tag{3.26}$$

We randomly sampled the parameters characterizing these forms over large ranges to obtain numerical inducibility curves corresponding to the networks carrying one or two copies of the network genes. In order to do this we computed numerical approximations to the dynamical system representing the evolution of the overall concentrations of activator (a) and inhibiting (i) agents according to a modeling scheme analogous to that described in section 3.4.3, *i.e.* we numerically solved the equations

$$\begin{cases} \dot{a} = \theta_a \left[\lambda + (1-\lambda) f(\rho, a, i)\right] - \gamma a \\ \dot{i} = \theta_i \left[\lambda + (1-\lambda) f(\rho, a, i)\right] - \gamma i \end{cases} \tag{3.27}$$

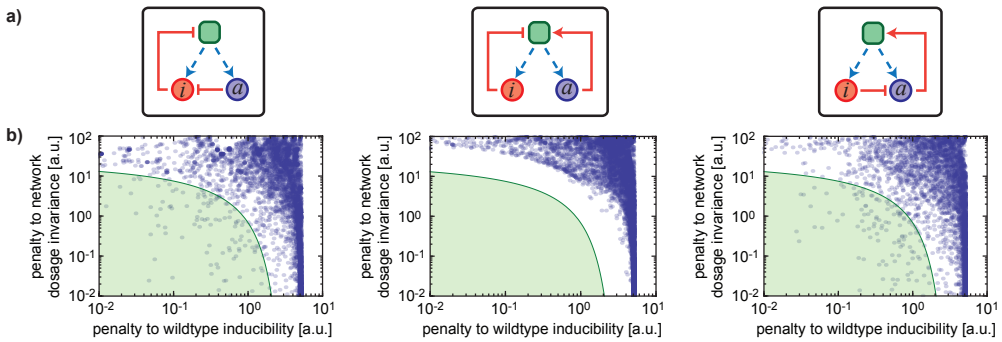

Figure 3.15: **a**, Schematics of the different generic network topologies explored. Blue and red circles represent activating (a) and inhibiting (i) agents, respectively, in all 3 networks. Dashed blue arrows denote the transcriptional production of the network components. The green square represents a transcriptional center. Pointing red arrows show direct activation while the blunt red arrows represent inhibition. **b**, For each configuration depicted in a, parameters were randomly sampled from a large range of values and fed into the proposed quantitative model to obtain numerical induction curves corresponding to the network configuration with one or two copies of the network genes. A proxy for the area between the two curves was quantified and plotted on the y-axis to represent the degree of dosage-invariance in the system. The difference between these numerical curves and a reference induction curve was also calculated and plotted on the x-axis to represent the ability of the network to be induced in a similar way as the experiment galactose system described in this Chapter.

over a time interval of 24 h and computed the value of f corresponding to the values of a and i achieved at that point. Under some conditions this procedure yielded multiple solutions in the case of the network topology presented in the center column of Figure 3.15; to obtain an average inducibility level for each galactose concentration we averaged the results across ten random initial conditions (a_0, i_0) distributed uniformly across a range of values enclosing the possible physiological steady states that a and i could attain ($a_0 \in [0, \theta_a/\gamma]$, $i_0 \in [0, \theta_i/\gamma]$).

For each pair of these numerical curves, we calculated the level of dosage invariance by integrating the area between the two curves, large areas co-

rresponding to large penalties to network-dosage invariance, and vice versa (Figure 3.15b, y-axis). The penalty figures presented in Figure 3.15 are numbers proportional to discrete estimates of this area based on finite sampling.

In principle, a high degree of dosage invariance can be observed at several different inducibility levels. For example, a biological network always staying in its OFF state is network-dosage invariant, but it lacks the ability to respond to signals of any kind. Therefore, it is important to determine if a dosage-compensated system is also inducible or not. In a similar way as before, we quantified the relative inducibility levels of our numerical curves relative to a reference induction profile why measuring the area between the curves with large differences with respect to the reference curve corresponding to large penalties to wildtype inducibility (Figure 3.15b, x-axis).

A comparative examination of the dot-plots corresponding to each network configuration reveals that the topologies at left and right allow their host networks to be both dosage-invariant and inducible. The specific interaction scheme in the two networks is essential for the systems to display such a behavior (Figure 3.15, left and right panels). However, the choice between activator and inhibitor in directly influencing transcription is not essential as long as the effect of the other is indirect.

The green regions in Figure 3.15b enclose networks that are both dosage-invariant and inducible (low penalties in both axes). To observe whether there were further restrictions on these systems we analyzed the distribution of parameters in these regions (Figure 3.16). The parameter that quantifies the nonlinearity of the interaction between the inhibiting and activating agents (α in Figure 3.16a and β in Figure 3.16b) was the only one severely restricted in the values it took, with values following a narrow distribution centered around 1. This finding suggests a further requirement on the network architecture: the effective stoichiometry of the interaction between the activating and inhibiting agents has to be 1-to-1 in order to produce a system that is both inducible and network-dosage invariant. This can be also understood by noting that any function of the form $f = f(x_1/x_2)$ will satisfy equation (3.23).

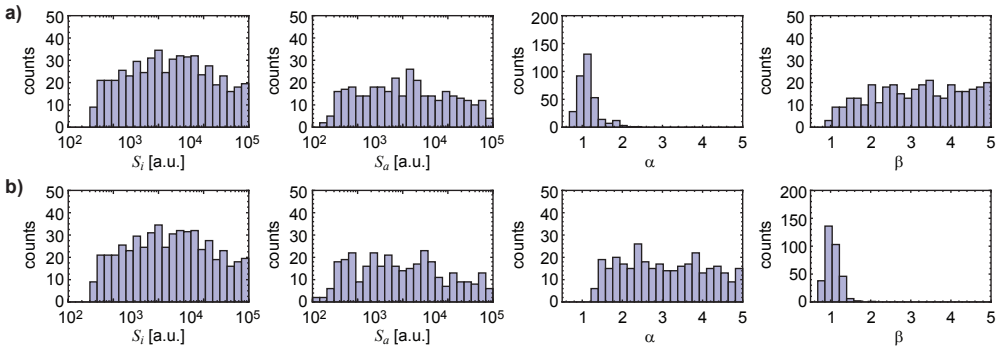

Figure 3.16: **a**, For the left configuration in Figure 3.15, histograms of the parameter values corresponding to the green region shown in that figure, where the system is both inducible and network-dosage invariant. **b**, As in a but for the right configuration shown in Figure 3.15. In addition to the topology requirement, a 1-to-1 stoichiometric interaction between the activating and inhibiting agents is essential to display a system that is both inducible and dosage-compensated at the network level.

3.6.3 Relationship to the *GAL* network

We note that the *GAL* system satisfies the requirements described in this section: the interaction circuitry (Figure 3.1) between its activator (*GAL3*) and inhibitor (*GAL80*) indeed accommodates the topology depicted in the left panel of Figure 3.15. Regarding the stoichiometry requirement, it has been experimentally shown that *GAL3* and *GAL80* share a 1-to-1 interaction stoichiometry [75]. These observations further validate our findings about the minimal set of essential elements that are necessary to build a network-dosage invariant phenotype into a natural gene network.

3.7 Conclusions

We observed the effect of gene dosage in a phenotype controlled by a complex genetic network and noted how the dosage of some of the genes had no effect around the operating point of the system.

Furthermore, we observed that the system at hand had the special property of network dosage invariance, *i.e.* the phenotype was conserved if the copy number of all the genes involved was modified in the same way.

We developed a framework for rationalizing these observations and by analyzing it we identified a volume-independent mechanism responsible from network-dosage invariance in gene networks. In order for a natural gene network to display such a behavior, it has to have at least two network components: one positive and one negative regulator. These components have to interact with a 1-to-1 effective stoichiometry and the topology of the underlying circuit has to be such that only one of them would directly affect transcription. Even though it is not necessarily true that these requirements will apply to different quantitative phenotypes, the developed framework could be applied to the analysis of different situations with minimal modifications.

This type of interaction topology is frequently observed [76–79] in natural gene circuits due to the abundant nature of sequestration-based signal transduction schemes in which an active protein is sequestered into an inactive complex by another protein in the network. Cells might implement this compensation mechanism when they need to exert extra layers of control over their phenotypes. Changes in ploidy, fluctuations in the number of chromosomes during cell cycle progression, global variation in gene expression, and genome duplication or loss are among the situations that can cause variations in network-dosage and raise the need for compensation.

In the evolutionary front, this analysis allows one to understand what network topologies are more likely to maintain or change phenotypes in the face of genomic duplication events.

Bibliography

[1] J. Hanna, K. Saha *et al.*, "Direct cell reprogramming is a stochastic process amenable to acceleration," *Nature*, vol. 462, pp. 595–601, 2009.

[2] Q. Yang, B. Pando *et al.*, "Circadian gating of the cell cycle revealed in single cells," *Science*, vol. 327, pp. 1522–1526, 2010.

[3] U. Alon, "Biological networks: the tinkerer as an engineer," *Science*, vol. 301, pp. 1866–1867, (2003.

[4] A. L. Barabasi and Z. N. Oltvai, "Network biology: understanding the cell's functional organization," *Nature Reviews Genetics*, vol. 5, pp. 101–113, 2004.

[5] T. Yamada and P. Bork, "Evolution of biomolecular networks: lessons from metabolic and protein interactions," *Nature Reviews Molecular Cell Biology*, vol. 10, pp. 791–803, 2009.

[6] M. Madan Babu, S. A. Teichmann and L. Aravind, "Evolutionary dynamics of prokaryotic transcriptional regulatory networks," *Journal of Molecular Biology*, vol. 358, pp. 614–633, 2006.

[7] A. E. Tsong, B. B. Tuch, H. Li and A. D. Johnson, "Evolution of alternative transcriptional circuits with identical logic," *Nature*, vol. 443, pp. 415–420, 2006.

[8] J. C. Venter *et al.*, "The sequence of the human genome," *Science*, vol. 291, pp. 1304–1351, 2001.

[9] K. C. Atwood, L. K. Schneider and F. J. Ryan, "Periodic selection in Escherichia coli," *Proceedings of the National Academy of Sciences USA*, vol. 37, pp. 146–155, 1951.

[10] J. Adams, C. Paquin, P. W. Oeller and L. W. Lee, "Physiological characterization of adaptive clones in evolving populations of the yeast Saccharomyces cerevisiae," *Genetics*, vol. 110, pp. 173–185, 1985.

[11] M. J. Dunham *et al.*, "Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae," *Proceedings of the National Academy of Sciences USA*, vol. 99, pp. 16144–16149, 2002.

[12] J. E. Barrick, D. S. Yu *et al.*, "Genome evolution and adaptation in a long-term experiment with Escherichia coli," *Nature*, vol. 461, pp. 1243–1247, 2009.

[13] D. Hartl and A. Clark, *Principles of population genetics*. Sinauer Associates, third ed., 1997.

[14] D. Gresham *et al.*, "Genome-wide detection of polymorphisms at nucleotide resolution with a single DNA microarray," *Science*, vol. 311, pp. 1932–1936, 2006.

[15] A. Segrè, A. Murray and J. Leu, "High-resolution mutation mapping reveals parallel experimental evolution in yeast," *PLoS Biology*, vol. 4, p. e256, 2006.

[16] G. Rancati, N. Pavelka *et al.*, "Aneuploidy underlies rapid adaptive evolution of yeast cells deprived of a conserved cytokinesis motor," *Cell*, vol. 135, pp. 879–893, 2008.

[17] C. J. Brown, K. M. Todd and R. F. Rosenzweig, "Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment," *Molecular Biology and Evolution*, vol. 15, pp. 931–942, 1998.

[18] R. Koszul, S. Caburet, B. Dujon and G. Fischer, "Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments," *EMBO Journal*, vol. 23, pp. 234–243, 2004.

[19] I. Wapinski, A. Pfeffer, N. Friedman and A. Regev, "Natural history and evolutionary principles of gene duplication in fungi," *Nature*, vol. 449, pp. 54–61, 2007.

[20] Y. Guan, M. J. Dunham and O. G. Troyanskaya, "Functional analysis of gene duplications in Saccharomyces cerevisiae," *Genetics*, vol. 175, pp. 933–943, 2007.

[21] A. C. Gerstein, H. J. Chun, A. Grant and S. P. Otto, "Genomic convergence toward diploidy in Saccharomyces cerevisiae," *PLoS Genetics*, vol. 2, p. e145, 2006.

[22] C. Zeyl, T. Vanderford and M. Carter, "An evolutionary advantage of haploidy in large yeast populations," *Science*, vol. 299, pp. 555–558, 2003.

[23] K. N. Smith and A. Nicolas, "Recombination at work for meiosis," *Current Opinion in Genetics & Development*, vol. 8, pp. 200–211, 1998.

[24] J. P. Gogarten and J. P. Townsend, "Horizontal gene transfer, genome innovation and evolution," *Nature Reviews Microbiology*, vol. 3, pp. 679–687, 2005.

[25] S. F. Elena and R. E. Lenski, "Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation," *Nature Reviews Genetics*, vol. 4, pp. 457–469, 2003.

[26] R. E. Lenski, M. R. Rose, S. C. Simpson and S. C. Tadler, "Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2000 generations," *American Naturalist*, vol. 138, pp. 1315–1341, 1991.

[27] T. L. Ferea, D. Botstein, P. O. Brown and R. F. Rosenzweig, "Systematic changes in gene expression patterns following adaptive evolution in yeast," *Proceedings of the National Academy of Sciences USA*, vol. 96, pp. 9721–9726, 1999.

[28] E. Stolovicki, T. Dror, N. Brenner and E. Braun, "Synthetic gene recruitment reveals adaptive reprogramming of gene regulation in yeast," *Genetics*, vol. 173, pp. 75–85, 2006.

[29] J. Y. Leu and A. W. Murray, "Experimental evolution of mating discrimination in budding yeast," *Current Biology*, vol. 16, pp. 280–286, 2006.

[30] E. Dekel and U. Alon, "Optimality and evolutionary tuning of the expression level of a protein," *Nature*, vol. 436, pp. 588–592, 2005.

[31] C. B. Brachmann *et al.*, "Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications," *Yeast*, vol. 14, pp. 115–132, 1998.

[32] R. S. Sikorski and P. Hieter, "A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae," *Genetics*, vol. 122, pp. 19–27, 1989.

[33] A. Becskei, B. B. Kaufmann and A. van Oudenaarden, "Contributions of low molecule number and chromosomal positioning to stochastic gene expression," *Nature Genetics*, vol. 37, pp. 937–944, 2005.

[34] E. Gari, L. Piedrafita, M. Aldea and E. Herrero, "A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in Saccharomyces cerevisiae," *Yeast*, vol. 13, pp. 837–848, 1997.

[35] A. Radzicka and R. Wolfenden, "A proficient enzyme," *Science*, vol. 267, pp. 90–93, 1995.

[36] V. Bryson and W. Szybalski, "Microbial selection," *Science*, vol. 116, pp. 45–51, 1952.

[37] M. Acar, J. T. Mettetal and A. van Oudenaarden, "Stochastic switching as a survival strategy in fluctuating environments," *Nature Genetics*, vol. 40, pp. 471–475, 2008.

[38] G. Gardiner, *Handbook of Stochastic Methods*. New York: Springer-Verlag, 1985.

[39] N. G. van Kampen, *Stochastic processes in physics and chemistry*. North-Holland, second ed., 1992.

[40] S. E. Luria and M. Delbrück, "Mutations of bacteria from virus sensitivity to virus resistance," *Genetics*, vol. 28, pp. 491–511, 1943.

[41] D. E. Lea and E. A. Coulson, "The distribution of the numbers of mutants in bacterial populations," *Journal of Genetics*, vol. 49, pp. 264–285, 1949.

[42] B. Mandelbrot, "A population birth-and-mutation process, I: Explicit distributions for the number of mutants in an old culture of bacteria," *Journal of Applied Probability*, vol. 11, pp. 437–444, 1974.

[43] S. Sarkar, "Haldane's solution of the Luria-Delbrück distribution," *Genetics*, vol. 127, pp. 257–261, 1991.

[44] M. M. Desai, D. S. Fisher and A. W. Murray, "The speed of evolution and maintenance of variation in asexual populations," *Current Biology*, vol. 17, pp. 385–394, 2007.

[45] M. Hegreness, N. Shoresh, D. Hartl and R. Kishony, "An equivalence principle for the incorporation of favorable mutations in asexual populations," *Science*, vol. 311, pp. 1615–1617, 2006.

[46] G. I. Lang and A. W. Murray, "Estimating the per-base-pair mutation rate in the yeast Saccharomyces cerevisiae," *Genetics*, vol. 178, pp. 67–82, 2008.

[47] S. Urlinger *et al.*, "Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity," *Proceedings of the National Academy of Sciences USA*, vol. 97, pp. 7963–7968, 2000.

[48] A. L. Goldstein and J. H. McCusker, "Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae," *Yeast*, vol. 15, pp. 1541–1553, 1999.

[49] A. Gelman, J. B. Carlin, H. S. Stern and D. S. Rubin, *Bayesian data analysis*. New York: Chapman & Hall/CRC, second ed., 2004.

[50] A. Motegi and K. Myung, "Measuring the rate of gross chromosomal rearrangements in Saccharomyces cerevisiae: A practical approach to study genomic rearrangements observed in cancer," *Methods*, vol. 41, pp. 168–176, 2007.

[51] A. Lee and J. R. Lupski, "Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders," *Neuron*, vol. 52, pp. 103–121, 2006.

[52] J. R. Korenberg *et al.*, "Down syndrome phenotypes: the consequences of chromosomal imbalance," *Proceedings of the National Academy of Sciences USA*, vol. 91, pp. 4997–5001, 1994.

[53] T. Galitski *et al.*, "Ploidy regulation of gene expression," *Science*, vol. 285, pp. 251–254, 1999.

[54] S. Di Talia *et al.*, "The effects of molecular noise and size control on variability in the budding yeast cell cycle," *Nature*, vol. 448, pp. 947–951, 2007.

[55] S. D. M. Santos and J. E. Ferrell, "On the cell cycle and its switches," *Nature*, vol. 454, pp. 288–289, 2008.

[56] S. Di Talia *et al.*, "Daughter-specific transcription factors regulate cell size control in budding yeast," *PLoS Biology*, vol. 7, p. e1000221, 2009.

[57] M. Kellis, B. W. Birren and E. S. Lander, "Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae," *Nature*, vol. 428, pp. 617–624, 2004.

[58] J. M. Pedraza and A. van Oudenaarden, "Noise propagation in gene networks," *Science*, vol. 307, pp. 1965–1969, 2005.

[59] M. Acar, A. Becskei and A. van Oudenaarden, "Enhancement of cellular memory by reducing stochastic transitions," *Nature*, vol. 435, pp. 228–232, 2005.

[60] T. S. Gardner, C. R. Cantor and J. J. Collins, "Construction of a genetic toggle switch in Escherichia coli," *Nature*, vol. 403, pp. 339–342, 2000.

[61] W. Xiong and J. E. Ferrell Jr., "A positive-feedback-based bistable 'memory module' that governs a cell fate decision," *Nature*, vol. 426, pp. 460–465, 2003.

[62] A. Mizutani and M. Tanaka, "Regions of GAL4 critical for binding to a promoter in vivo revealed by a visual DNA-binding analysis," *EMBO Journal*, vol. 22, pp. 2178–2187, 2003.

[63] K. Melcher and H. E. Xu, "Gal80-Gal80 interaction on adjacent Gal4p binding sites is required for complete GAL gene repression," *EMBO Journal*, vol. 20, pp. 841–851, 2001.

[64] T. Suzuki-Fujimoto *et al.*, "Analysis of the galactose signal transduction pathway in Saccharomyces cerevisiae: interaction between Gal3p and Gal80p," *Molecular and Cellular Biology*, vol. 16, pp. 2504–2508, 1996.

[65] G. Peng and J. E. Hopper, "Evidence for Gal3p's cytoplasmic location and Gal80p's dual cytoplasmic-nuclear location implicates new mechanisms for controlling Gal4p activity in Saccharomyces cerevisiae," *Molecular and Cellular Biology*, vol. 20, pp. 5140–5148, 2000.

[66] G. Peng and J. E. Hopper, "Gene activation by interaction of an inhibitor with a cytoplasmic signaling protein," *Proceedings of the National Academy of Sciences USA*, vol. 99, pp. 8548–8553, 2002.

[67] B. B. Kaufmann, Q. Yang, J. T. Mettetal and A. van Oudenaarden, "Heritable stochastic switching revealed by single-cell genealogy," *PLoS Biology*, vol. 5, p. e239, 2007.

[68] S. Ghaemmaghami *et al.*, "Global analysis of protein expression in yeast," *Nature*, vol. 425, pp. 737–741, 2003.

[69] A. P. Gasch *et al.*, "Genomic expression programs in the response of yeast cells to environmental changes," *Molecular Biology of the Cell*, vol. 11, pp. 4241–4257, 2000.

[70] A. Raj *et al.*, "Stochastic mRNA synthesis in mammalian cells," *PLoS Biology*, vol. 4, p. e309, 2006.

[71] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics*, vol. 22, pp. 403–434, 1976.

[72] U. Güldener *et al.*, "A new efficient gene disruption cassette for repeated use in budding yeast," *Nucleic Acids Research*, vol. 24, pp. 2519–2524, 1996.

[73] J. Ma and M. Ptashne, "Deletion analysis of GAL4 defines two transcriptional activating segments," *Cell*, vol. 48, pp. 847–853, 1987.

[74] K. M. Hawkins and C. D. Smolke, "The regulatory roles of the galactose permease and kinase in the induction response of the GAL network in Saccharomyces cerevisiae," *Journal of Biological Chemistry*, vol. 281, pp. 13485–13492, 2006.

[75] D. J. Timson, H. C. Ross and R. J. Reece, "Gal3p and Gal1p interact with the transcriptional repressor Gal80p to form a complex of 1:1 stoichiometry," *Biochemical Journal*, vol. 363, pp. 515–520, 2002.

[76] N. E. Buchler and M. Louis, "Molecular titration and ultrasensitivity in regulatory networks," *Journal of Molecular Biology*, vol. 384–385, p. 1106, 2008.

[77] L. Bardwell *et al.*, "Repression of yeast Ste12 transcription factor by direct binding of unphosphorylated Kss1 MAPK and its regulation by the Ste7 MEK," *Genes & Development*, vol. 12, pp. 2887–2898, 1998.

[78] Y. Liu and J. M. Belote, "Protein-protein interactions among components of the Drosophila primary sex determination signal," *Molecular and General Genetics*, vol. 248, pp. 182–189, 1995.

[79] R. Benezra *et al.*, "The protein Id: a negative regulator of helix-loop-helix DNA binding proteins," *Cell*, vol. 61, pp. 49–59, 1990.