



UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II



UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

PH.D. THESIS IN

INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING

**MODELING AND CONTROL OF GENE EXPRESSION
DYNAMICS IN YEAST**

GIANSIMONE PERRINO

TUTOR: PROF. DIEGO DI BERNARDO

CO-TUTOR: PROF. MARIO DI BERNARDO

XXIX CICLO

**SCUOLA POLITECNICA E DELLE SCIENZE DI BASE
DIPARTIMENTO DI INGEGNERIA ELETTRICA E TECNOLOGIE DELL'INFORMAZIONE**

Contents

1	Introduction	1
1.1	Background and motivation	1
1.2	Thesis outline	3
1.3	List of main contributions	4
2	Control theory meets synthetic biology	7
2.1	Control theory	7
2.2	The role of control theory in synthetic biology	8
2.2.1	External feedback control	8
3	An experimental platform for controlling an ensemble of yeast cells	19
3.1	Experimental platform characteristics	19
3.1.1	Microfluidics device for the observation and manipulation of yeast cells	21
3.1.2	Hydrostatic pressure actuation	21
3.1.3	Microscopy and image analysis	24
3.1.4	External feedback controller	24
4	A comparative analysis of external feedback control strategies in yeast	27
4.1	An experimental testbed for the assessment of control strategies: the galactose-inducible promoter in <i>S. cerevisiae</i>	28
4.2	A mathematical representation for the transcription driven by the galactose-inducible promoter	29

4.3	Control strategies	30
4.3.1	Proportional–Integral control	31
4.3.2	Model Predictive Control	33
4.3.3	Zero Average Dynamics control	36
4.4	Controlling gene expression from the galactose–inducible promoter: setpoint and tracking control tasks	38
4.4.1	Performance measures	40
4.5	Numerical simulations and <i>in vivo</i> experiments	41
4.5.1	Setpoint control experiments	41
4.5.2	Signal tracking control experiments	41
4.6	Discussion	48
5	Analysis, modeling and control of single–cell gene expression in yeast	49
5.1	Single cell quantification of protein expression	50
5.1.1	Image segmentation and tracking algorithm for yeast cells	51
5.1.2	Single–cell fluorescence time–series data	51
5.2	Analysis of single cell fluorescence data	53
5.2.1	Quantification and analysis of cellular noise at steady–state	53
5.2.2	Quantification and analysis of cellular noise dynamics	54
5.3	Mathematical representation of single–cell gene expression from the galactose–inducible promoter	59
5.3.1	Mixed–effects model as an ideal framework for describing cell to cell variability	59
5.3.2	Mixed–effects model inference	60
5.3.3	Numerical simulations of the mixed–effects model of protein expression from the galactose–inducible promoter in yeast cells	62
5.4	A novel model–predictive–control strategy for controlling an ensemble of yeast cells	65

5.4.1	Numerical simulations of the MEMPC strategy to steer protein expression from the galactose-inducible promoter	65
5.5	Discussion	67
6	Feedback control of human α-synuclein protein expression in yeast	71
6.1	A quantitative study for protein aggregation in yeast: the human α -synuclein case	73
6.2	A pilot study on the feasibility of controlling α -synuclein expression in yeast	74
6.2.1	External feedback controller of α -synuclein expression	77
6.2.2	Identification and modeling of α -synuclein protein expression	77
6.2.3	Tracking control experiments	79
6.3	Experimental investigation of human α -synuclein protein aggregation in yeast	81
6.3.1	Open-loop dynamics of α -synuclein protein expression in the multiple copy SNCA _{A53T} -GFP yeast strain	83
6.3.2	Offset-free Model Predictive Control of α -synuclein protein expression in the multiple copy SNCA _{A53T} -GFP yeast strain	83
6.4	Discussion	92
7	Conclusions	95
A	Materials and methods	97
A.1	Yeast strains	97
A.1.1	Single copy SNCA _{A53T} -GFP strain construction	98
A.1.2	Multiple copy SNCA _{A53T} -GFP strain construction	99
A.2	Microfluidics	100
A.2.1	Master-mold	100
A.2.2	Device fabrication	101

A.3	Microscopy	101
A.4	Image analysis	102
A.4.1	Custom offline tracking algorithm	103
A.5	Computational analysis	104
A.5.1	Model quality metrics	104
A.6	Experimental Protocol	105
A.7	Cell culture	105
A.7.1	Microfluidic device set up	106
A.7.2	Feedback control phase	108
B		109
Bibliography		113

Chapter 1

Introduction

1.1 Background and motivation

Control Engineering aims at driving a physical system in order to attain a specific value of a quantity of interest (such as a boiler that needs to warm water to a desired temperature, or a car cruise-control maintaining a constant speed) despite the presence of disturbances. This is achieved by appropriately varying its inputs (switch on or off a heater in the case of the boiler, or accelerating or braking in the cruise-control) as a function of the difference between the measured value of the output and its desired target value (control error).

At the core of most control schemes lies a negative feedback loop [1], as depicted in Fig. 1.1. The variable to be controlled (system output y) is measured through a sensor and its value is subtracted from the desired value (control reference r). The quantity that is obtained is minimised by the controller, a set of logical and mathematical rules through which an appropriate value of the input u is chosen in order to guarantee that the output y matches the desired reference r . The input u is thus applied to the system by an actuator.

Synthetic biology is a novel research field, in which biomolecular circuits are assembled in living cells with the final goal of controlling cellular behavior for a number of uses, from energy, to environment, to medicine [2].

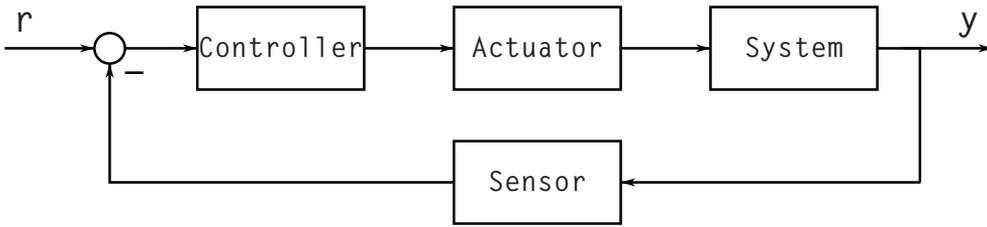


Figure 1.1: **Negative feedback loop.**

However, realisation of synthetic biomolecular circuits is often a lengthy and ad hoc process. Mainly, this is due to nonlinearity, stochasticity, variability and lack of modularity in biomolecular processes [3]. In recent years, control theory has been applied to synthetic biology for tackling several of these problems, leading to promising results [4].

Implementations of negative feedback in synthetic biology fall into two different categories [4]:

- *embedded* feedback control, which has both the controller and the process realised inside the cell by means of biomolecular processes;
- *external* feedback control, which has the whole cell as the process to be controlled, while the controller is implemented in a computer workstation.

In this study, I deal exclusively with external feedback control, since the controller is realised outside the cell.

So far, external feedback control has been extensively applied to control growing conditions of cells in chemostats in terms of temperature and/or CO_2 and it is a current feature of bench-top and industrial chemostats [5, 6]. Only recently, however, the application of control theory principles has been exploited to regulate molecular events in living cells, thanks to innovative microfluidics and optogenetics platforms [7–13].

A multitude of successful attempts to control gene expression, or even signaling pathways, have been described in the literature. They mainly differ in the control input (osmotic pressure, light, small-molecules) and the control strategy adopted. Optogenetics-based light inducible systems have been

exploited to control gene expression in yeast [7, 11, 13], to regulate intracellular signalling dynamics in mammalian cells [8], and to drive protein levels by using light-switchable two-component systems in bacteria [14]. Microfluidic-based devices, allowing a tight control of cellular growing medium and the administration of inducer small-molecules, have been successfully employed to investigate synchronisation properties of synthetic biological clocks in bacterial cells [15], to control the transcription from the *STL1* promoter in yeast *S. Cerevisiae* by varying the osmotic pressure [9], to control the transcription from the endogenous *GAL1* promoter in yeast *S. Cerevisiae* using galactose and glucose as input [10], and to regulate the transcription from a tetracycline-inducible (*CMV-TET*) promoter in mammalian cells by varying the concentration of an antibiotic [12].

The main aim of this Thesis is to explore and extend the methodology to model and control gene expression in population of living cells by applying concepts borrowed from control theory.

So far, the different control strategies proposed in the literature have never been compared in the same experimental model, thus making a direct comparison of their performance impossible. Here, I performed a rigorous comparative analysis of different control algorithms for regulating gene expression from the endogenous *GAL1* promoter in yeast cells.

Furthermore, I present an application of gene expression control to study aggregation of α -synuclein, a pathological hallmark of Parkinson's disease.

1.2 Thesis outline

In Chapter 2, I present control theory in the framework of synthetic biology. I provide an overview of both disciplines, focusing the discussion on the role of control engineering to develop and improve synthetic biology. Moreover, I illustrate the-state-of-the-art of gene expression control in cell populations.

In Chapter 3, I introduce the experimental platform which I used to perform time-lapse experiments on populations of living yeast cells and to implement external feedback control of gene expression.

In Chapter 4, I present a comparative analysis of different control algorithms for regulating gene expression from the endogenous galactose-inducible promoter in yeast cells. To this aim, I implemented and compared two feedback control strategies reported in literature, and conceived an additional strategy employed so far only in power electronic systems. The comparative analysis was carried out both *in silico* (i.e. by numerical simulations) and *in vivo*.

In Chapter 5, I analysed single cell gene expression data from the time-lapse experiments reported in Chapter 4, and characterised intrinsic and extrinsic gene expression noise. I derived a mixed-effects dynamical model to correctly describe the variability in fluorescence level both in individual cells and at the population level. I finally propose a novel model-predictive-control approach, based on mixed-effects models.

In Chapter 6, I describe medically relevant application of feedback control to model and study aggregation dynamics of α -synuclein protein, which is a pathological hallmark of Parkinson's disease. The aim is to attain a quantitative understanding of the aggregation dynamics of α -synuclein by carefully regulating its expression and following its dynamics in living cells.

Finally, in Chapter 7, final considerations are drawn.

1.3 List of main contributions

The main contributions of this study are fourfold.

- Firstly, I contributed to improve the efficiency of feedback control of gene expression in cell populations (Chapter 4).
- Secondly, I contributed to characterise gene expression from endogenous galactose-inducible promoter both at single cell and at population level (Chapter 5).
- Thirdly, I proposed a novel model-predictive-control strategy to predict the overall behavior of the cell population. The novelty lies on employing single cell models to carry out the prediction (Chapter 5).

- Lastly, I studied aggregation dynamics of the α -synuclein protein, that is a pathological hallmark of Parkinson's disease (Chapter 6).

The main results of this study are presented in a number of past [16–18] and ongoing publications.

Chapter 2

Control theory meets synthetic biology

In this Chapter, I introduce control theory in the framework of synthetic biology. A brief introduction of control theory is provided, together with an overview of the novel field of synthetic biology. I then analyse the role of control theory in developing and improving synthetic biology. Lastly, the state-of-the-art of external feedback control in synthetic biology is illustrated.

2.1 Control theory

Control theory aims at driving a physical system in order to attain a specific value of a quantity of interest despite the presence of disturbances. This task is achieved by appropriately changing its inputs as a function of the difference between the measured value of the output and its desired target value (control error).

Control theory has been used in a number of applications for improving the stability, robustness, and performance of physical systems [1]. Indeed, it has been successfully applied in mechanical devices, electrical/power networks, space and air systems, and chemical processes [1].

Nature discovered feedback long ago. It created feedback mechanisms

and exploited them at all levels that are central to homeostasis and life [1]. The human being started to use feedback as a technology almost two millennia ago [1]. Nevertheless, only in recent years the technology has been exploited in a biological context [4]. In this context, one of the most promising field of application is synthetic biology.

2.2 The role of control theory in synthetic biology

Synthetic biology is a novel research field (cf. Fig. 2.1), which aims to engineer new functions in living cells by creating, characterizing and assembling biological parts, devices and systems [19].

The practical realisation of synthetic circuits is often a lengthy and ad hoc process. Nonlinearity, stochasticity, variability and lack of modularity in cell processes arise as problems to be solved when dealing with synthetic biology [3]. In recent years, control theory has been applied to synthetic biology for tackling several of these problems, leading to initial but promising results [4].

2.2.1 External feedback control

External feedback control is an application of control theory to synthetic biology. It has been conceived to overcome the limitations arising from implementing an entire feedback control loop inside a cell (i.e. *embedded* feedback control) [4]. Indeed, the feedback control loop is realised outside the cell.

An external feedback control system can be decomposed into four modules: measurement, control, actuation, and biological process to be controlled. The measurement module measures the reporter fluorescence level of either a single cell or a cell population by employing flow cytometry or fluorescence microscopy. The measured fluorescence level is sent to the control module, normally a computer workstation, that elaborates the measured

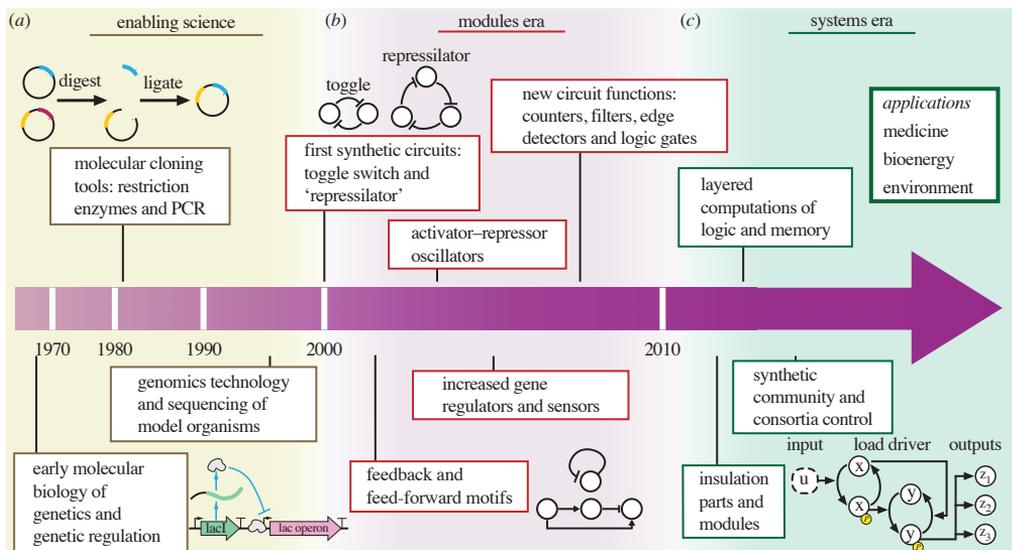


Figure 2.1: **Condensed timeline of synthetic biology.** (a) The development of synthetic biology is grounded on molecular biology, genetic engineering and genomics. (b) The early phases of synthetic biology were focusing mostly on forward engineering simple modules, such as switches and oscillators. (c) After the ‘era’ of modules, synthetic biology is heading towards the era of systems, in which modules will serve as functional units to create more complex and sophisticated systems with potential applications to energy, environment and medicine. From [4].

data to obtain the state of the biological process. The state is then employed by a control algorithm to compute a control input to be applied to the cell in order to steer the biological process dynamics. The control input is actuated by applying external stimuli such as light (optogenetics method) [7, 8, 11, 13] or small-molecules [9, 10, 12]. Finally, the biological process responds to the external stimuli through biomolecular reactions producing a change in a fluorescence reporter, which is measured thus closing the feedback loop.

Applications of control to biological system in literature differ substantially in the control strategy and in the biological process to be regulated [4].

Toettcher *et al.* [8] were the first to present in literature an operating external feedback control. They employed a light-based feedback to control protein-protein interaction dynamics in mammalian cells (cf. Fig. 2.2). A PI controller was chosen to carry out the control objectives (setpoint, linear ramp up, exponential ramp up, and linear ramp down).

Almost simultaneously, Miliás-Argeitis *et al.* [7] applied a model predictive controller to regulate gene expression from a light-switchable system based on PhyB-PIF3 interaction in yeast cells (cf. Fig. 2.3).

Subsequently, Uhlendorf *et al.* [9] implemented an external feedback loop to control gene expression from an endogenous osmopress responsive promoter in yeast cells (cf. Fig. 2.4). It is noteworthy that for the first time in literature, a microfluidics approach was employed to carry out the study. In this case, the cell process is responsive to an osmotic shock, which is mediated by the high osmolarity glycerol (HOG) signaling cascade. Since the presence of short-term (non transcriptional) and long-term (transcriptional) negative feedback loops ensures perfect adaptation of yeast cells to osmotic stress, a more sophisticated feedback control algorithm was needed. Indeed, a model predictive control strategy was chosen to compute the control input.

Melendez *et al.* [11] employed a bang-bang control strategy to steer gene expression dynamics from a light-switchable two-component system in yeast cells (cf. Fig. 2.5).

Menolascina *et al.* conceived and implemented a microfluidics-based

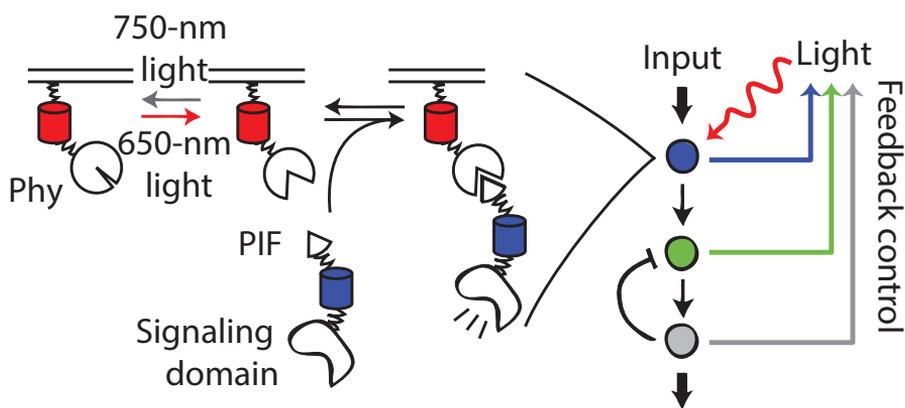


Figure 2.2: **External feedback control of intracellular signaling dynamics via optogenetics.** Schematic of feedback control of the Phy–PIF optogenetic module. Upon ligation to the small-molecule chromophore phycocyanobilin (PCB), membrane-fused, fluorescent Phy fusion proteins can be used to drive fluorescent PIF-tagged proteins to the plasma membrane by exposure to 650 – nm light, and this interaction can be reversed by exposure to 750 – nm light. By automatically tuning input light levels, feedback control sets the activity state at downstream nodes for which live-cell readouts are available. Adapted from [8].

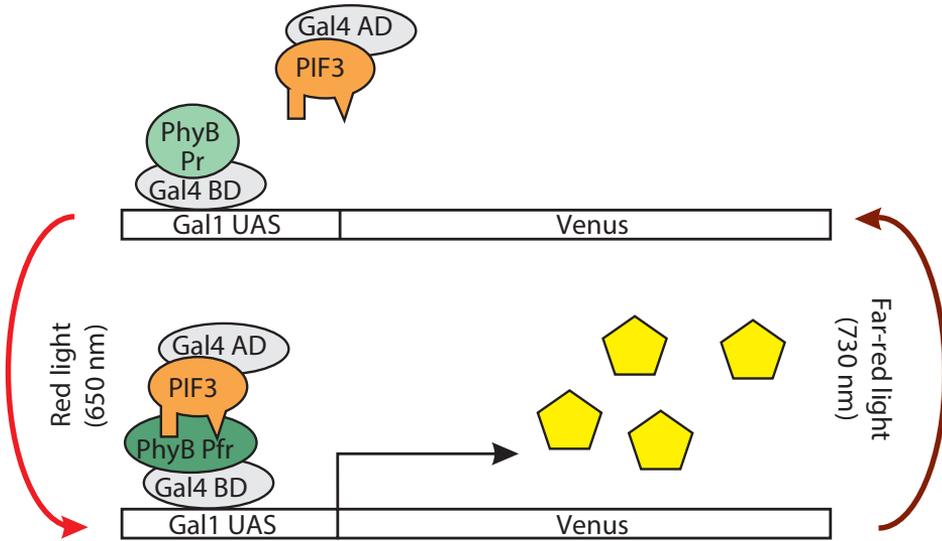


Figure 2.3: **External feedback control of gene expression from *GAL1* inducible–promoter via optogenetics.** Light–switchable gene system based on PhyB–PIF3 interaction. Transformed cells grown in darkness and incubated with the chromophore phycocyanobilin (PCB) synthesize both PhyB(Pr)–GBD and PIF3–GAD fusion proteins. Because PIF3 interacts only with the activated form of PhyB (Pfr), the Gal1 target gene is initially off. Upon exposure to red light, PhyB is rapidly converted into its active Pfr form and binds the PIF3 moiety of PIF3–GAD. The transcription activation domain of Gal4 is therefore recruited to the promoter and induces transcription of the target gene. Exposure to far–red light switches off gene expression by rapidly converting PhyB into its inactive Pr form, causing its dissociation from PIF3–GAD. Adapted from [7].

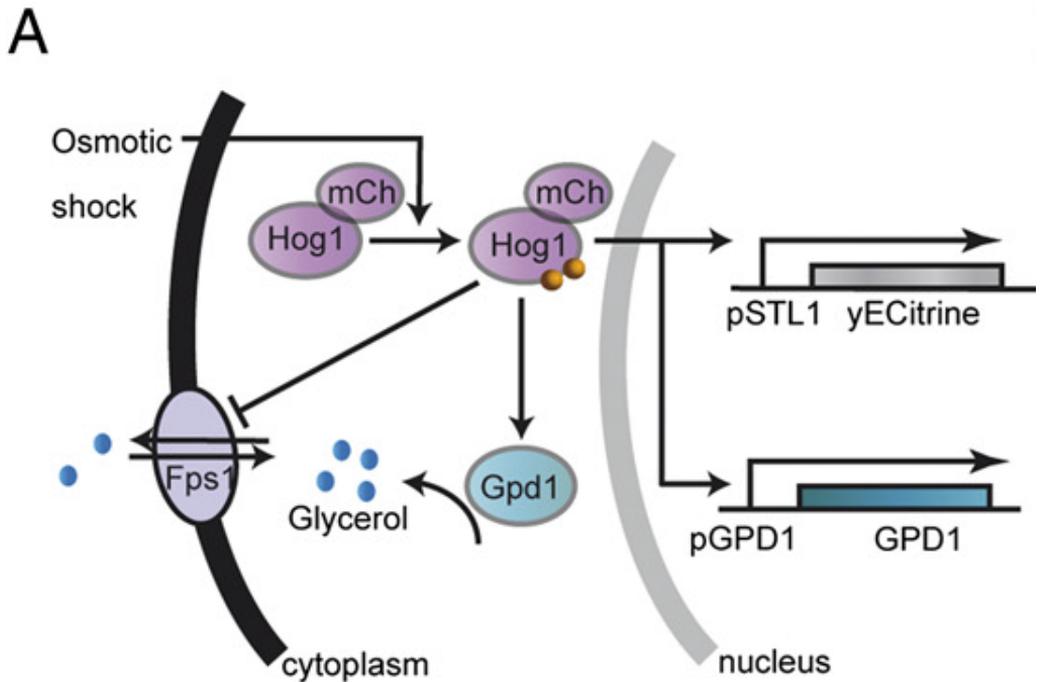


Figure 2.4: **External feedback control of gene expression from endogenous *STL1* promoter via microfluidics.** Schematic of high osmolarity glycerol (HOG) signaling cascade. A hyperosmotic stress triggers the activation and nuclear translocation of Hog1. Short-term adaptation is mainly implemented by cytoplasmic activation of the glycerol-producing enzyme Gpd1 and closure of the aqua-glyceroporin channel Fps1. Long-term adaptation occurs primarily through the production of Gpd1. From [9].

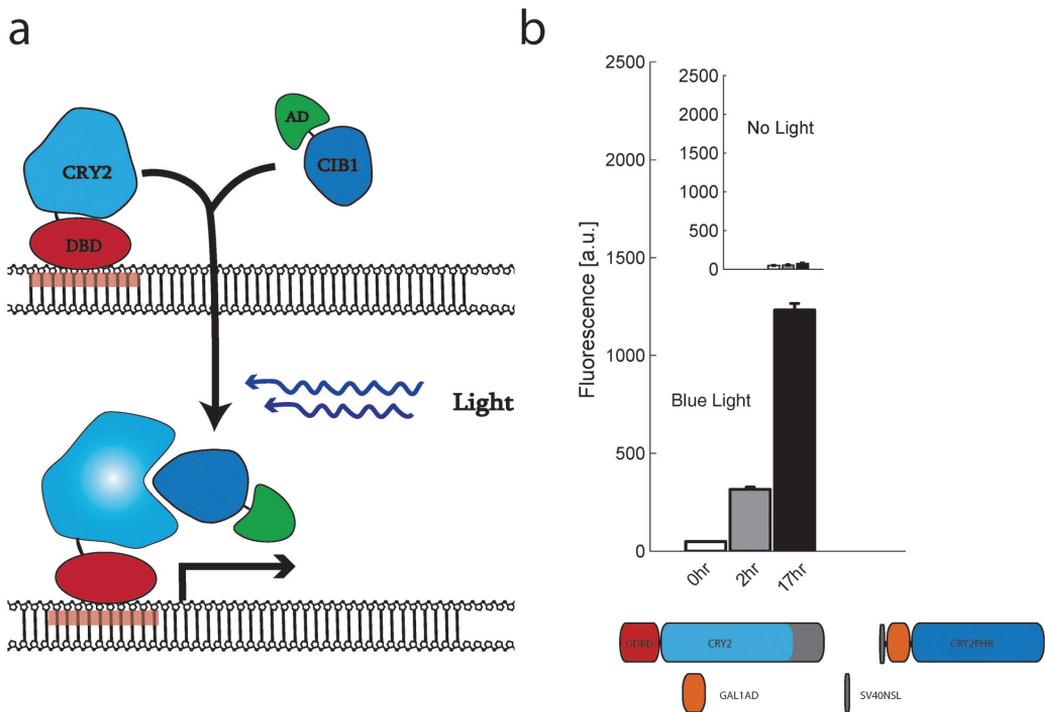


Figure 2.5: **External feedback control of gene expression from light-switchable two-component system via optogenetics.** *Arabidopsis* proteins CRY2 and CIB1 are used to make transcription blue-light responsive. (a) In response to blue light, CRY2 and CIB1 bind. When fused to appropriate DNA-binding and activation domains, transcription becomes blue-light inducible. (b) CRY2 and CIB1 constructs fused to the GAL4 DNA-binding and GAL4 activation domains respectively. When plasmids containing these constructs are co-transformed with a *pGAL1-yEVENUS* reporter plasmid, *yEVENUS* expression becomes blue-light inducible. From [11].

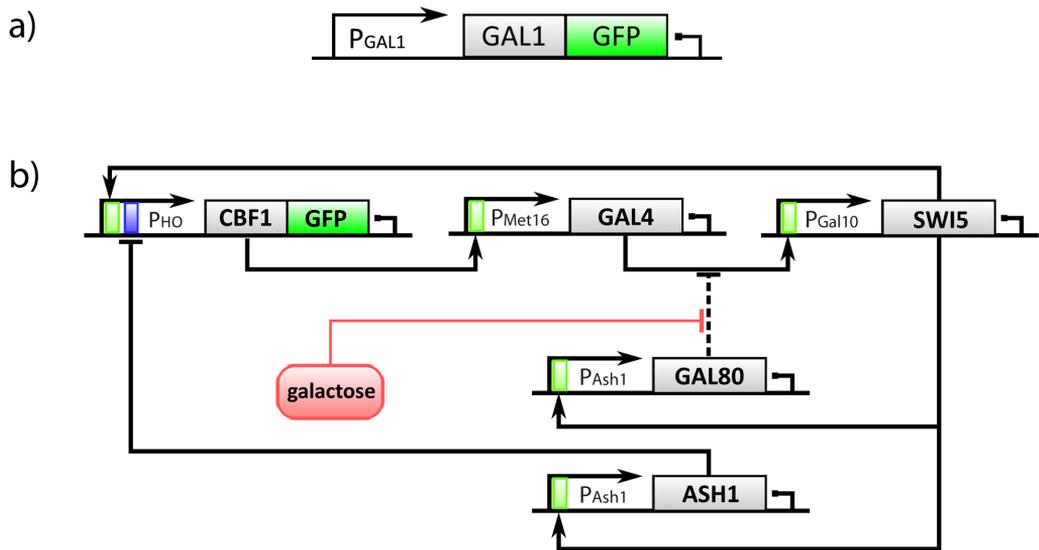


Figure 2.6: **External feedback control of gene expression from endogenous and synthetic gene networks via microfluidics.** (a) Schematic of endogenous *GAL1* promoter gene network. The Gfp protein was integrated downstream of the endogenous *GAL1* promoter [20]. (b) Schematic of synthetic gene network IRMA. IRMA is composed of 5 genes encoding for transcription factors modulating the expression of each other [21]. From [10].

feedback control for regulating gene expression both from endogenous and synthetic gene networks in yeast (cf. Fig. 2.6). A PI controller was used to achieve the control tasks for both the cell processes. Despite the simple control strategy, the control objectives were satisfied also in the case of the complex synthetic gene network (cf. Fig. 2.6b), that was composed by five interacting genes.

Fracassi *et al.* [12] proposed an external feedback control loop for regulating gene expression from the tetracycline-inducible promoter in mammalian cells (cf. Fig. 2.7). Despite the simple control task (setpoint) and the simple control strategies adopted to carry out the study (relay control and PI control), this work is remarkable since it presented for the first time in literature a control of gene expression in mammalian cells.

Recently, Miliás–Argeitis *et al.* [13] significantly improved the results

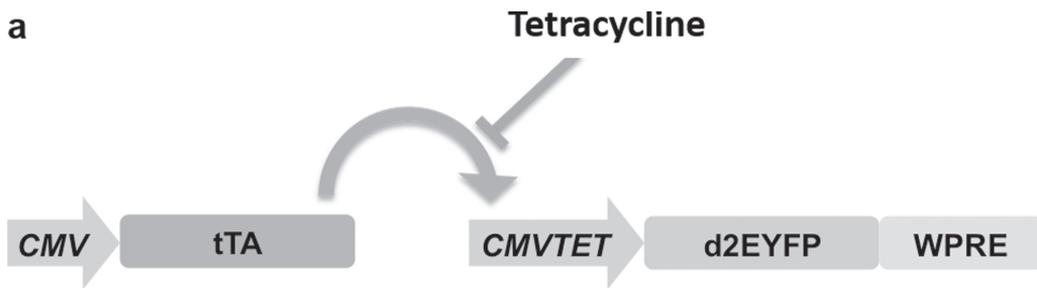


Figure 2.7: **External feedback control of gene expression from the *CMV-TET* inducible-promoter via microfluidics.** Schematic of the *CMV-TET* inducible-promoter. The tetracycline transactivator protein (tTA) is constitutively expressed by a *CMV* promoter. tTA binds the *CMV-TET* promoter, which harbors seven tet-responsive elements (*tetO7*) upstream of a minimal *CMV* promoter, thus driving the transcription of a destabilized enhanced yellow fluorescent reporter protein (d2EYFP). Doxycycline or tetracycline binds tTA and prevents it from binding to the *CMV-TET* promoter. Adapted from [12].

in [7] by presenting a completely automatic system capable of long-term optical feedback control of gene expression in continuous liquid cultures. The light-switchable cyanobacterial two-component system in *Escherichia coli* was used as cell process (cf. Fig. 2.8). Instead, the control tasks were achieved by means of two control strategies: PI control and MPC.

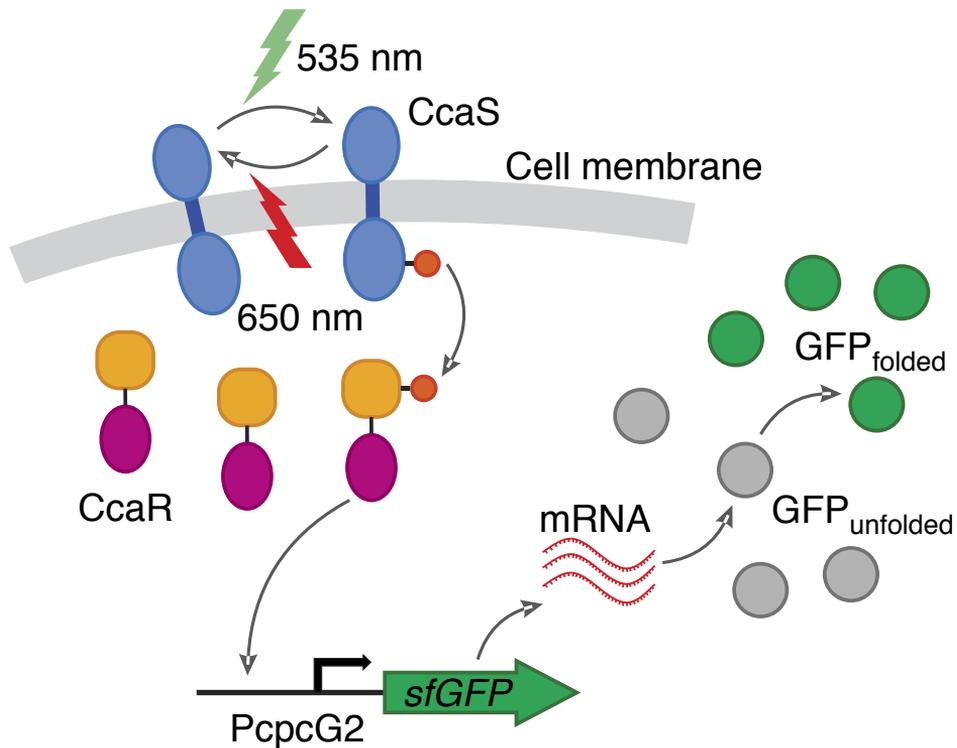


Figure 2.8: **External feedback control of gene expression from the light-switchable cyanobacterial two-component system via optogenetics.** Schematic of the light-switchable two-component system. On absorption of green light, the sensor histidine kinase CcaS is quickly autophosphorylated and transfers its phosphate group to the cognate response regulator CcaR. Phosphorylated (active) CcaR in turn binds to the *cpcG2* promoter to activate transcription of sfGFP. Absorption of red light inactivates CcaS, and transcription is eventually switched off. Adapted from [13].

Chapter 3

An experimental platform for controlling an ensemble of yeast cells

In this Chapter, I introduce the experimental platform for external feedback control of gene expression in living cells. I exploited this integrated experimental set-up to perform time-lapse experiments on populations of living yeast cells, as detailed in Chapters 5 and 6. Further details about its design and implementation can be found in [10, 22, 23].

3.1 Experimental platform characteristics

The experimental platform consists of an integrated set-up comprising a microfluidic device, a time-lapse fluorescence microscope and a set of automated syringes, all controlled by a computer workstation. An overview of the platform is given in Fig. 3.1.

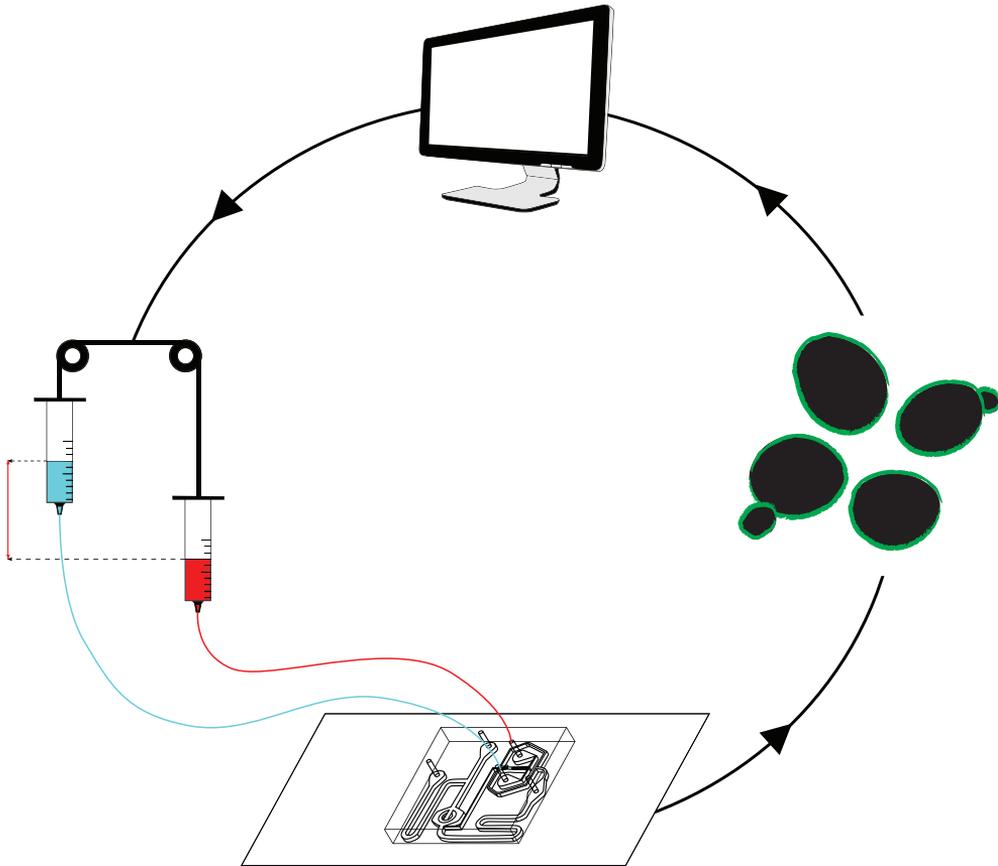


Figure 3.1: **Overview of the experimental platform.** Cells are loaded in a microfluidic device in order to guarantee their survival during time-lapse experiments lasting thousands of minutes. Two different growing media can be administrated to cells by means of an automated system of syringes by adjusting their relative heights and hence differences in hydrostatic pressure to dynamically steer the flow of medium from one of the two syringes to the cells. The syringes are controlled by a computer workstation running the control algorithm.

3.1.1 Microfluidics device for the observation and manipulation of yeast cells

Microfluidics is the science and technology of systems that process or manipulate small amounts of fluids (10^{-9} to 10^{-18} L), employing channels with dimensions of tens to hundreds of micrometres [24].

The application of microfluidics to biology offers a number of advantages [24]. In fact, a microfluidic device allows to grow cells ensuring their vitality along the entire duration of experiment and to dynamically change the cell microenvironment (e.g. providing different drugs or inducing molecules). Moreover, it employs small quantities of samples and reagents thus to reducing experimental costs.

The MFD0005a device (Fig. 3.2) has been used for all the microfluidics experiments reported in this thesis. It was designed by the *Biodynamics Laboratory* of Prof. Jeff Hasty (UCSD) [25].

The geometry of the MFD0005a device is represented in Fig. 3.2a. Yeast cells are trapped into a micro-chamber (Fig. 3.2a – grey area), whose height is $3.5 \mu\text{m}$ allowing them to grow only as a monolayer thus enabling accurate microscopic image analysis by preventing cells to move out of focus. Two different media can be used as inputs to steer the cell behaviour: the flow from ports 1 and 2 reaches the trap through the staggered herringbone mixer (Fig. 3.2c) in the Dial-A-Wave junction (Fig. 3.2b). This mixer is needed to mix the fluids since flows in microfluidics are laminar (no turbulence). A complete descriptions of the MFD0005a device, comprising of all its characteristics, has been published in [25].

I produced replicas of the device MFD0005a thanks to the master-mold that Prof. Jeff Hasty kindly provided us as a blueprint, according to the protocol procedures published in [25].

3.1.2 Hydrostatic pressure actuation

The fluid that reaches the yeast trap is a mixture of two fluids coming from the inlet ports 1 and 2. The Dial-A-Wave junction (Fig. 3.2b) is used as

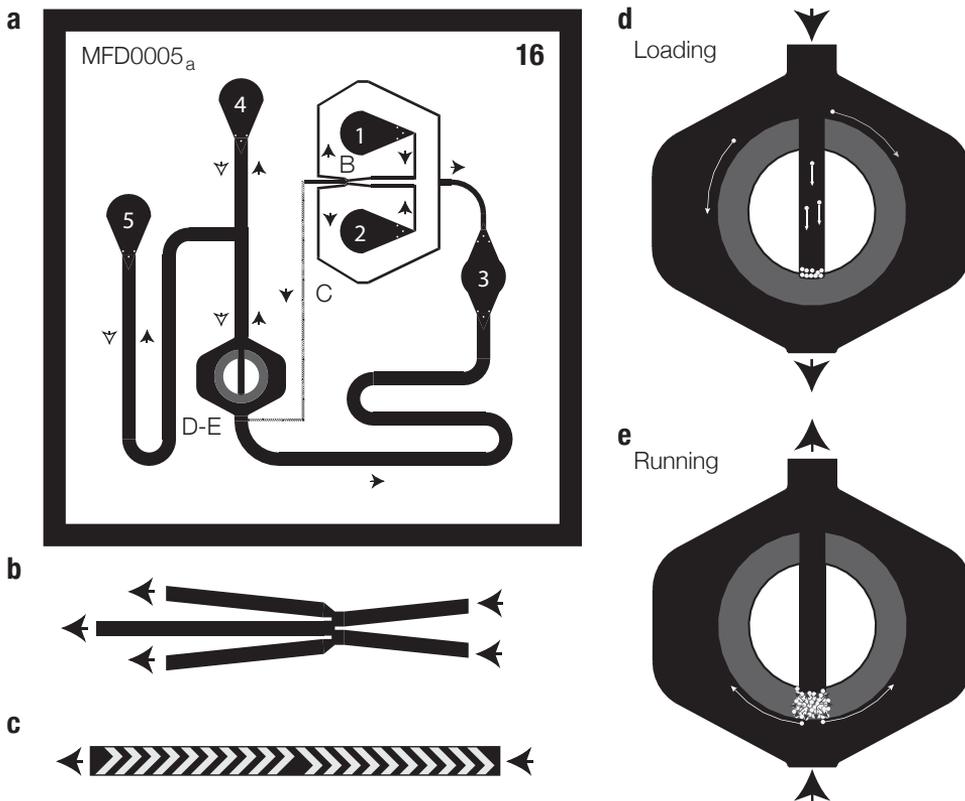


Figure 3.2: **Microfluidic device MFD0005a.** (a) Overview of the MFD0005a chip's architecture. (b) Dial-A-Wave junction. (c) Staggered herringbone mixer (SHM). (d) Trap region under loading conditions. (e) Cell trap upon running of an experiment. Picture adapted from [25].

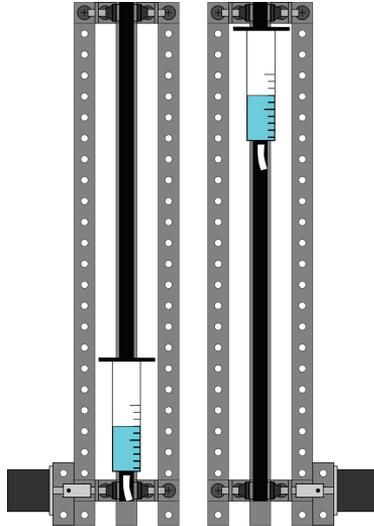


Figure 3.3: **Motorised linear rails.**

mixer, and the blending of the fluids depends from the relative pressure between the two fluids at the inlet ports. Thus, in order to change the fluid that arrives at the trap, it is necessary to modify its relative pressure. Physically, the aim can be achieved by changing the hydrostatic pressure of the syringes linked to the two inlet ports.

The hydrostatic pressure actuation relies on a system of two vertically mounted linear actuators (Fig. 3.3). In this way, it is possible to change the height of the liquid-filled syringes, and hence the relative pressure of the fluids at the inlet ports.

In detail, the actuation system comprises two linear guides, that are conceived to move independently; and two stepper motors, that realise the motion of the syringes through two timing belt and four pulleys (a timing belt and two pulleys for guide). The entire actuation system is represented in Fig. 3.3.

Complete details regarding the sizing and the specifications of the actuation system has been reported in [23].

3.1.3 Microscopy and image analysis

Fluorescence is acquired by means of an inverted fluorescence Eclipse Ti-E microscope (Nikon Instruments) equipped with an automated and programmable stage, an incubator to guarantee fixed temperature (30°C) and gasses (O₂) to cell environment, and a high sensitivity Electron Multiplying CCD (EM-CCD) Camera (iXON Ultra897; Andor Technology Ltd); as depicted in Fig. 3.4.

The microscope and the camera were programmed to acquire, at 5 min intervals, two types of images: (*a*) a phase contrast image, and (*b*) fluorescence images (with the appropriate filters) to monitor fluorescence due to fluorescent proteins and to coloured dyes, as e.g. Sulforhodamine B (Sigma–Aldrich Co.), that can be added to one of the enriched media in order to evaluate the inputs administered to the cells.

A custom image processing algorithm was developed to quantify the fluorescence signals being expressed by the entire population of yeast cells. The measurement units for the fluorescence are considered arbitrary and, thus, a calibration phase at the beginning of each control experiment is needed to calculate a reference value for the fluorescence steady states.

Further details about the image processing algorithm are reported in [22, 23, 26]

3.1.4 External feedback controller

The computer workstation runs the control algorithm, which at each sampling interval:

1. processes images acquired by the microscope to estimate the fluorescence;
2. executes the control algorithm to derive the input u for the next sampling period;
3. controls the set of automated syringes to administrate the calculated input (i.e. induction medium or no induction medium) to the cells.

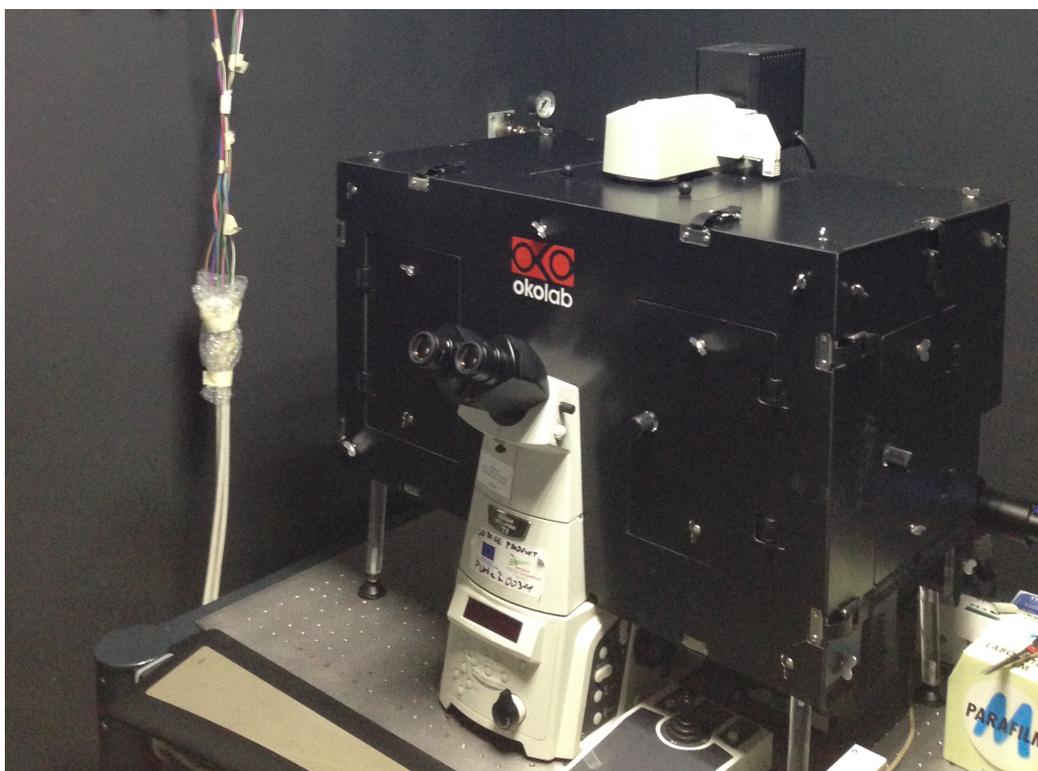


Figure 3.4: Nikon Eclipse Ti-E microscope.

In a previous work, this experimental platform was successfully used to demonstrate that the average fluorescence level of a yeast population can be effectively regulated by means of a simple Proportional–Integral control strategy [10]. However, the possibility of exploiting new control strategies to improve the control performance motivated the study proposed in the following chapters.

Chapter 4

A comparative analysis of external feedback control strategies in yeast

Many successful attempts to control gene expression, or even signaling pathways, have been described in literature (refer to Chapter 2 for details). They mainly differ in the control input (osmotic pressure, light, small-molecules) and the control strategy adopted to steer the amount of protein expression. So far, the different control strategies proposed in the literature have never been compared in the same experimental model, thus making a direct comparison of their performance impossible. This is extremely important for practical applications where knowing advantages and limitations of each strategy can be useful, if not necessary, to select the most appropriate and effective one.

In this Chapter, I compared *in silico*, that is by numerical simulations, and *in vivo*, by experiments with yeast cells, the performance of different control algorithms when applied to the problem of controlling gene expression from the galactose-inducible promoter. In addition to the control strategies that have already been described in the literature and reported in Chapter 2, namely the Proportional-Integral (PI) control and the Model Predictive Control (MPC), I also tested a different control strategy named Zero

Average Dynamics (ZAD) control. ZAD is a control strategy inspired by sliding control techniques [27] in power electronic systems, but that has never been applied to biological systems. Practical considerations of the pros and cons of each control strategy are provided in details.

The content of this chapter has been developed in collaboration with Dr. Gianfranco Fiore and partially published in [16].

4.1 An experimental testbed for the assessment of control strategies: the galactose–inducible promoter in *S. cerevisiae*

The endogenous galactose–inducible promoter is the most widely used inducible promoter in yeast genetics. Thousands of strains, each expressing a different yeast gene, are available to the research community, making this an attractive choice for practical applications of control engineering. The activity of the galactose–inducible promoter is governed by the presence of galactose in the cells' growing medium. This sugar is interpreted as a "switch on" signal for the expression of the *GAL1* gene; conversely when yeasts are fed with glucose, the expression of the *GAL1* gene is repressed [28]. Yeast cells will first consume all the available glucose in the medium before starting metabolising galactose. Hence, the control input can either be glucose (switch off signal) or galactose (switch on signal), but not an intermediate concentration of the two sugars, because cells will not respond to galactose when glucose is present.

I decided to use the galactose–inducible promoter upstream of a reporter gene (GFP fused with the Gal1 protein – cf. Fig. 4.1) in *Saccharomyces cerevisiae* (commonly known as baker's yeast) as a testbed for comparing and assessing the performance of the different control strategies. When dealing with living cells, one of the major issues is represented by the uncertainty affecting transcriptional and translational processes, introducing a remarkable cell–to–cell variability in mRNA and protein production [29]. Rather than

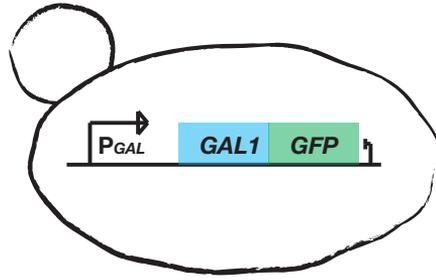


Figure 4.1: Galactose-inducible promoter in *Saccharomyces cerevisiae*.

trying to control stochastic behaviour of cells, here I addressed the simpler problem of regulating the average fluorescence intensity expressed by all cells as the quantity to be controlled (y), thus averaging out the effects due to intrinsic and extrinsic sources of noise [30].

In vivo control experiments were carried out using the integrated experimental set-up described in detail in Chapter 3. As reported in Chapter 2, Dr. Filippo Menolascina and Dr. Gianfranco Fiore demonstrated in a previous work that the average fluorescence level of a yeast population can be effectively regulated with this real-time platform using a simple Proportional-Integral controller [10].

4.2 A mathematical representation for the transcription driven by the galactose-inducible promoter

The control strategies I compared in this Chapter need a mathematical model of the process being controlled to compute the control input (u). Thus, in order to implement these regulators, I used the mathematical representation for the transcription driven by galactose-inducible promoter derived by Dr. Gianfranco Fiore and reported in [23, 31].

The mathematical representation that I chose is a state-space linear

model defined as:

$$\dot{x}(t) = A x(t) + B u(t) \tag{4.1}$$

$$y(t) = C x(t)$$

where the matrices A , B and C are defined as:

$$A = \begin{bmatrix} -d_1 & 0 \\ v_2 & -d_2 \end{bmatrix} \tag{4.2}$$

$$B = \begin{bmatrix} b_1 \\ 0 \end{bmatrix} \tag{4.3}$$

$$C = \begin{bmatrix} 0 & c_2 \end{bmatrix} \tag{4.4}$$

In Eq. (4.1), u is the only external stimulus to the model and it is assumed to be equal to 1 when cells are fed with galactose, whereas, when glucose is provided to yeasts, it is assumed to be equal to 0 (note that these values are related to the concentration of galactose in the growing medium).

The numerical values for the model parameters are taken from [16, 23], and are reported in the following:

$$A = \begin{bmatrix} -0.0063 & 0 \\ 0.0274 & -0.0166 \end{bmatrix} \tag{4.5}$$

$$B = \begin{bmatrix} 0.0037 \\ 0 \end{bmatrix} \tag{4.6}$$

$$C = \begin{bmatrix} 0 & 1.0343 \end{bmatrix} \tag{4.7}$$

4.3 Control strategies

In designing the control strategies, two major constraints were identified:

- the sampling time;
- the admissible values of the control input.

I set the sampling time (T) equals to 5 min as a compromise to minimise phototoxicity to the cell but maintain a reasonable temporal resolution. The control input u can instead assume only two values (galactose–ON, glucose–OFF). Thus, at each sampling time kT , the control algorithms can only choose the duration of galactose pulse (ON), which can vary from 0 min to 5 min, and it is defined as the duty–cycle:

$$d_k = \frac{t_{ON}}{T}, \quad (4.8)$$

i.e. the percentage of the time interval during which the cells are fed with galactose. The control input u is mathematical described as follows, where ON means that cells are fed with galactose enriched medium, and conversely OFF with glucose enriched medium:

$$u(t) = \begin{cases} u_{MAX} = ON & kT \leq t < (k + d_k)T \\ u_{MIN} = OFF & (k + d_k)T \leq t < (k + 1)T \end{cases} \quad (4.9)$$

4.3.1 Proportional–Integral control

The Proportional–Integral (PI) control algorithm uses the control error

$$e(t) = r(t) - y(t) \quad (4.10)$$

to compute, at each sampling time (kT), the duty–cycle value (d_k). Specifically, d_k has a value proportional to the weighted sum of two contributions, one proportional to the actual error $e(t)$ and the other proportional to the sum of the past values of the error (the integral term). The proportionality constants K_p and K_i are called respectively proportional and integral gains, and their values were calculated with the Ziegler–Nichols’ open–loop tuning method [1] applied to the mathematical model of the galactose–inducible promoter. Thus, the gains were set to $K_p = 13.49$ and $K_i = 0.17$.

Given the constraints on the control input as well as on the sampling time described above, a modulation on the PI output was implemented to calculate the duty cycle (d_k) as:

$$d_k = \frac{\hat{u} - u_{MIN}}{u_{MAX} - u_{MIN}}. \quad (4.11)$$

where \hat{u} is the output of the PI regulator saturated between $u_{MIN} = 0$ and $u_{MAX} = 1$. To avoid delays and overshoots introduced by the saturation of the regulator output [1], an anti-windup block was added to the feedback loop.

PI control implementation

In practical control implementations, the control signal $\hat{u}(t)$ is fed to the process being regulated by means of actuators (i.e. motors, valves, pumps). These components have physical limitations: motors have limited speed and acceleration, valves cannot be more than fully opened or fully closed and pumps cannot go slower than stopped. Thus the control signal acting on the system is saturated between the minimum (u_{MIN}) and the maximum (u_{MAX}) values achievable with the actuator being used.

If the control variable \hat{u} exceeds the saturation limits, the actuator will constantly run at its saturation limits, thus the feedback loop becomes unresponsive to the error being measured and the error remains nonzero. In the case of PI feedback control strategy, the error is integrated and the integral term may become very large, hence the control signal remains saturated even if the error changes and, it may take a long time before the integrator and the controller output return inside the saturation range. This situation is called *integrator windup* and leads to large transient in system response.

Anti-windup compensation. To fulfil the constraints on the control input (u), a modulation was implemented on the PI output. Specifically, the modulation introduces a saturation on the regulator output; thus leading to a possible integrator windup. To overcome this issue, I introduced in the

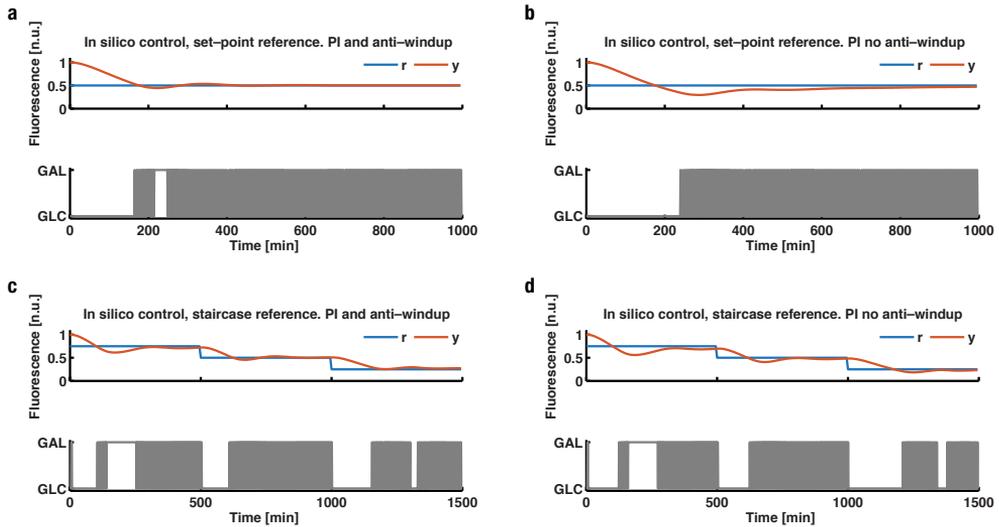


Figure 4.2: *In silico* control with and without windup compensation. The blue line is the reference signal (r). The orange line is the simulated fluorescence level (y). The grey line is the control input (u). (a–d) Four *in silico* control experiments performed on the mathematical model of the galactose-inducible promoter by the means of the PI control algorithm in the case of set-point control with (a) and without anti-windup compensation (b) and, to accomplish the tracking of the staircase reference signal with (c) and without anti-windup compensation (d).

feedback control loop an anti-windup scheme proposed in [1].

I compared *in silico* the performance of the PI regulator with and without the anti-windup compensation scheme (cf. Fig. 4.2). The system output (y) reaches the control objective faster when being controlled with the anti-windup compensation (setpoint control task – cf. Fig. 4.2a–b) and exhibits less oscillations around the reference (staircase control reference – cf. Fig. 4.2c–d).

4.3.2 Model Predictive Control

Model Predictive Control (MPC) is an iterative optimisation-based technique which uses a mathematical model of the process being controlled to predict the future values of a cost function of the control error within a finite

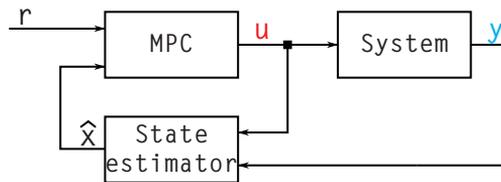


Figure 4.3: **Model Predictive Control scheme.**

prediction horizon and to find the best value of the control input (in this case the duty-cycle value (d_k)) that minimises the cost function over the prediction horizon [32].

Specifically, in my implementation, at each sampling time (kT), the MPC algorithm computes the optimal control input that minimises the sum of the squared control error (SSE):

$$SSE \triangleq \sum_{i=k+1}^{k+N} (N+1+k-i) \epsilon_i^2 = \sum_{i=k+1}^{k+N} (N+1+k-i) (\hat{y}_i - r_i)^2 \quad (4.12)$$

where \hat{y} is the output provided by the mathematical model of the galactose-inducible promoter (cf. Section 4.2), which is computed by a Kalman state estimator, able to reconstruct system states from the measured output y , as depicted in Fig. 4.3. The integer $N = 12$ (corresponding to 60 min) defines the length of the prediction horizon in terms of sampling intervals. The forgetting factor ($N + 1 + k - i$) weights the error samples more at the beginning of the prediction horizon than at the end; this guarantees faster corrections of output deviations from the reference. The optimisation was carried out by adopting the MATLAB (The MathWorks, Inc.) implementation of the Genetic Algorithm described in [33].

The result of the optimisation is an array of N optimal duty-cycles d_{k+i} , $i \in [1, N]$. In the absence of external disturbances and other sources of uncertainty, the optimal input computed by the MPC could, in principle, be applied to the yeast cells over the entire prediction horizon. However, in order to make the control action robust to any source of uncertainty and variability, the feedback loop is closed by applying only the first element of

the calculated control input and at the next sampling time $(k + 1)T$ when the entire procedure is repeated.

MPC implementation

To reduce the computational effort to solve the optimal control problem, I decided to use a discretised version of the mathematical model introduced in Section 4.2, defined as

$$x_{k+1} = A x_k + B u_k \tag{4.13}$$

$$y_k = C x_k$$

and assuming that the control input (u) is piecewise constant during the sampling period (Zero-order hold method described in [34]). $x_k = [x_{k,1} \ x_{k,2}]^T$ is the vector of system state, $u_k = u(kT)$ is the control input, $y_k = y(kT)$ is the measured output, with k being a natural number $\{k \in \mathbb{N}\}$; and $x_0 = [x_{0,1} \ x_{0,2}]^T$ is the vector of the initial condition. In this case, the matrices A , B and C are:

$$A = \begin{bmatrix} a_{1,1} & 0 \\ a_{2,1} & a_{2,2} \end{bmatrix} \tag{4.14}$$

$$B = \begin{bmatrix} b_1 \\ 0 \end{bmatrix} \tag{4.15}$$

$$C = \begin{bmatrix} 0 & c_2 \end{bmatrix} \tag{4.16}$$

The control input, as in the continuous time case, affects directly only the first system equation via the coefficient b_1 , and it can take values $u_{MAX} = 1$ and $u_{MIN} = 0$.

4.3.3 Zero Average Dynamics control

The Zero Average Dynamics (ZAD) algorithm relies on a feedback strategy devised for the regulation of power converters [35, 36], and allows to directly calculate the duty-cycle (d_k) of a switching control input [35, 37].

ZAD control is a practical implementation of Sliding Mode Control [27], where the control objective consists in attracting and then maintaining onto a fixed surface $s(x) = 0$ (denoted as the sliding surface) the states of the system by appropriately switching the available inputs.

In the ZAD control approach, the sliding condition has to be fulfilled only on average over each sampling period (kT), thus allowing to directly calculate the duty-cycle (d_k) via the solution of the following integral equation:

$$\mathbb{E}_T[s(x(t))] = \frac{1}{T} \int_{kT}^{(k+1)T} s(x(t)) dt = 0 \quad (4.17)$$

where \mathbb{E}_T indicates the operator taking the average over the time interval T .

For controlling galactose-inducible promoter dynamics onto the desired reference signal, I considered the following sliding surface, which was derived by the mathematical model introduced in Section 4.2:

$$s(x) = (x_2 - x_{2_{ref}}) + (\dot{x}_2 - \dot{x}_{2_{ref}}) \quad (4.18)$$

where x_2 is the state variable describing the dynamics of the fluorescent reporter (GFP). Note that $\dot{x}_{2_{ref}}(t) = 0$ in the case of setpoint regulation.

ZAD implementation

The solution of Eq. 4.17 can be computationally expensive. Thus, to overcome this issue, I considered the piecewise-linear approximation of the sliding surface $s(x)$ proposed in [37]:

$$s(x(t)) = \begin{cases} s_k + (t - kT) \dot{s}_k^{\text{on}} & kT \leq t < (k + d_k)T \\ s_k + d_k T \dot{s}_k^{\text{on}} + (t - (k + d_k)T) \dot{s}_k^{\text{off}} & (k + d_k)T \leq t < (k + 1)T \end{cases} \quad (4.19)$$

where s_k , \dot{s}_k^{on} , and \dot{s}_k^{off} are defined as:

$$\begin{aligned} s_k &= s(x_k) \\ \dot{s}_k^{\text{on}} &= \dot{s}(x_k) \Big|_{u=1} \\ \dot{s}_k^{\text{off}} &= \dot{s}(x_k) \Big|_{u=0} \end{aligned} \quad (4.20)$$

Considering the mathematical model of gene expression from the galactose-inducible promoter introduced in Chapter 4, the sliding surface is defined as:

$$s(x) = (x_2 - x_{2_{ref}}) + (\dot{x}_2 - \dot{x}_{2_{ref}}) \quad (4.21)$$

By substituting the piecewise-linear approximation (4.19) into Eq. 4.17, I obtained:

$$\begin{aligned} \mathbb{E}_T[s(x(t))] &= \frac{1}{T} \int_{kT}^{(k+d_k)T} [s_k + (t - kT) \dot{s}_k^{\text{on}}] dt \\ &\quad + \frac{1}{T} \int_{(k+d_k)T}^{(k+1)T} [s_k + d_k T \dot{s}_k^{\text{on}} + (t - (k + d_k) T) \dot{s}_k^{\text{off}}] dt \end{aligned} \quad (4.22)$$

and by computing the integral in (4.22):

$$\mathbb{E}_T[s(x(t))] = 0 \implies \frac{1}{2} d_k^2 T (\dot{s}_k^{\text{off}} - \dot{s}_k^{\text{on}}) - d_k T (\dot{s}_k^{\text{off}} - \dot{s}_k^{\text{on}}) + s_k + \frac{1}{2} T \dot{s}_k^{\text{off}} = 0 \quad (4.23)$$

The duty cycle d_k can then be calculated by solving the quadratic equation (4.23), thus obtaining:

$$d_k = \frac{-T (\dot{s}_k^{\text{on}} - \dot{s}_k^{\text{off}}) \pm \sqrt{T (\dot{s}_k^{\text{on}} - \dot{s}_k^{\text{off}}) (2 s_k + T \dot{s}_k^{\text{on}})}}{-T (\dot{s}_k^{\text{on}} - \dot{s}_k^{\text{off}})} \quad (4.24)$$

Moreover, considering that

$$\dot{s}_k^{\text{off}} - \dot{s}_k^{\text{on}} = -b_1 a_{2,1} < 0 \implies \dot{s}_k^{\text{on}} - \dot{s}_k^{\text{off}} > 0 \quad (4.25)$$

then, the solutions of (4.23) are

$$d_k = 1 \mp \sqrt{\frac{2 s_k + T \dot{s}_k^{\text{on}}}{T (\dot{s}_k^{\text{on}} - \dot{s}_k^{\text{off}})}} \quad (4.26)$$

As the duty cycle assumes values only in $[0, 1]$, the only admissible solution is

$$d_k = 1 - \sqrt{\frac{2 s_k + T \dot{s}_k^{\text{on}}}{T (\dot{s}_k^{\text{on}} - \dot{s}_k^{\text{off}})}} \quad (4.27)$$

Furthermore, to avoid saturation, I imposed this constraint

$$0 \leq \frac{2 s_k + T \dot{s}_k^{\text{on}}}{T (\dot{s}_k^{\text{on}} - \dot{s}_k^{\text{off}})} \leq 1 \quad (4.28)$$

4.4 Controlling gene expression from the galactose-inducible promoter: setpoint and tracking control tasks

I compared the performance of the three control strategies (PI, MPC and ZAD) when performing two different control tasks, as shown in Fig. 4.4:

- **setpoint control**, where the average GFP fluorescence expressed from the galactose-inducible promoter must reach and maintain a desired reference level;
- **signal tracking control** where the average GFP fluorescence must follow (or track) a desired time-varying signal.

Specifically, in the setpoint control (Fig. 4.4a), the desired fluorescence r was set equal to 50% of the initial average fluorescence expressed by the cell population during the calibration phase of 180 min. During the calibration phase, cells are kept in galactose, in order to set the unit of measure of fluorescence, which may vary due to technical and biological variability in each experiment. In the signal tracking control, three different references r are used:

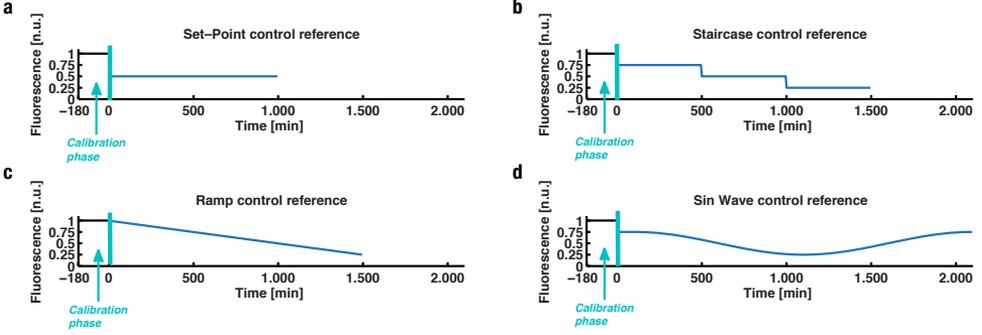


Figure 4.4: **Reference signals for setpoint and tracking control tasks.** (a) The desired setpoint r (blue line) is equal to 50% of the average fluorescence measured during the initial calibration phase of 180 min (black line). (b) The desired level of fluorescence (r) is a three-step descending staircase signal, each step is set at a given percentage (75%, 50% and 25%) of the average fluorescence measured during the initial calibration phase of 180 min. (c) The desired level of fluorescence (r) is a linear descending ramp starting at 100% of the average fluorescence measured during the initial calibration phase of 180 min and going down to 25%. (d) The desired level of fluorescence (r) is a steady state signal equal to 75% of the average fluorescence measured during the calibration phase, with a duration of 100 min; followed by a sinusoidal wave of period $T = 2000$ min.

- a descending staircase function (Fig. 4.4b) where each step lasts 500 min, beginning at 75% of the fluorescence measured during the calibration phase, then stepping down to 50% and then 25%;
- a linear descending ramp of 1500 min (Fig. 4.4c) starting at the 100% of average fluorescence measured during the calibration phase, and decreasing down to 25%;
- a sinusoidal wave $s(t)$ of period $T = 2000$ min (Fig. 4.4d) defined as

$$s(t) = 0.5 + 0.25 \sin \left(\frac{2\pi}{T} (t - 100) + \frac{\pi}{2} \right). \quad (4.29)$$

4.4.1 Performance measures

For assessing and comparing control performance obtained from the different control algorithms, I used metrics based on the analysis of the control error e . These measures, in general, are adopted for optimising the tuning of PI and PID controller gains on the basis of the control outcome [1].

In details, the performance measures used in this work are: ISE , IAE and $ITAE$.

ISE The *Integral Square Error* is defined as:

$$ISE = \int_0^t e(\tau)^2 d\tau \quad (4.30)$$

and integrates the square of the error over the time. It penalises large errors more than smaller ones.

IAE The *Integral Absolute Error* is defined as:

$$IAE = \int_0^t |e(\tau)| d\tau \quad (4.31)$$

and integrates the absolute error of the control over time.

ITAE The *Integral Time Absolute Error* is a weighted version of the IAE , and is calculated as:

$$ITAE = \int_0^t \tau |e(\tau)| d\tau. \quad (4.32)$$

Integrating the absolute error multiplied by the time, $ITAE$ penalises more persisting errors than those at the start of the response.

4.5 Numerical simulations and *in vivo* experiments

4.5.1 Setpoint control experiments

I first tested *in silico*, by numerical simulations, the three control strategies described above, by simulating the behavior of yeast cells by means of identified dynamical model. Numerical simulations' results are shown in Fig. 4.5. All the control strategies are able to reach and maintain the reference fluorescence value without exhibiting oscillations at steady-state. Performance measures (ISE, IAE, ITAE in Fig. 4.5d) are of the same order of magnitude for all the control strategies; interestingly the ZAD controller is able to achieve satisfying results with a reduced number of input switches (five and six fold less than respectively MPC and PI). This is advantageous in the experimental setting because it reduces unnecessary stress to cells.

In vivo control experiments confirm results from numerical simulations, demonstrating that the three strategies are indeed all able to reach and maintain the desired fluorescence level (cf. Fig. 4.6). As predicted by the *in silico* simulations, the ZAD controller employs fewer galactose pulses (cf. Fig. 4.6c) and displays smaller oscillations around the setpoint than the MPC strategy (cf. Fig. 4.6b).

4.5.2 Signal tracking control experiments

Numerical simulation of the descending staircase tracking control task shows that the three control strategies have very different performances. The PI is not able to properly follow the reference signal (Fig. 4.7a). This is to be expected, since the PI controller was designed specifically to solve setpoint control tasks [1]. The MPC algorithm, with its intrinsic predictive ability, achieves a good performance, specifically noticeable in the proximity of the steps' edges (Fig. 4.7b). Indeed, the MPC is able to foresee changes in the reference signal and to adjust the control input accordingly, by starting to "switch off" the system in advance. The ZAD control algorithm (Fig. 4.7c)

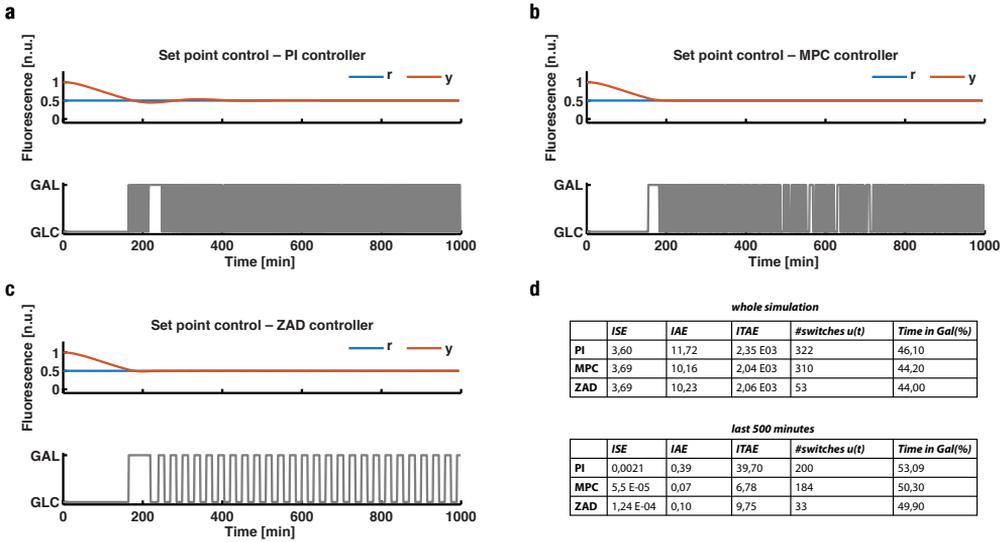


Figure 4.5: *In silico* setpoint control task. The blue line is the reference signal (r). The orange line is the simulated fluorescence level (y). The grey line is the control input (u). (a–c) Three *in silico* setpoint control experiments performed on the mathematical model of the galactose-inducible promoter by the means of the PI (a), MPC (b) and ZAD (c) controllers. The initial level of fluorescence is assumed to be equal to 1 (n.u.). The control action starts at time $t = 0$ min and ends at $t = 1000$ min. d) Performance measures: Integral Square Error (ISE), Integral Absolute Error (IAE), Integral Time Absolute Error (ITAE), number of switches of the control input, and the percentage of time during which the model is provided with the 'ON' input.

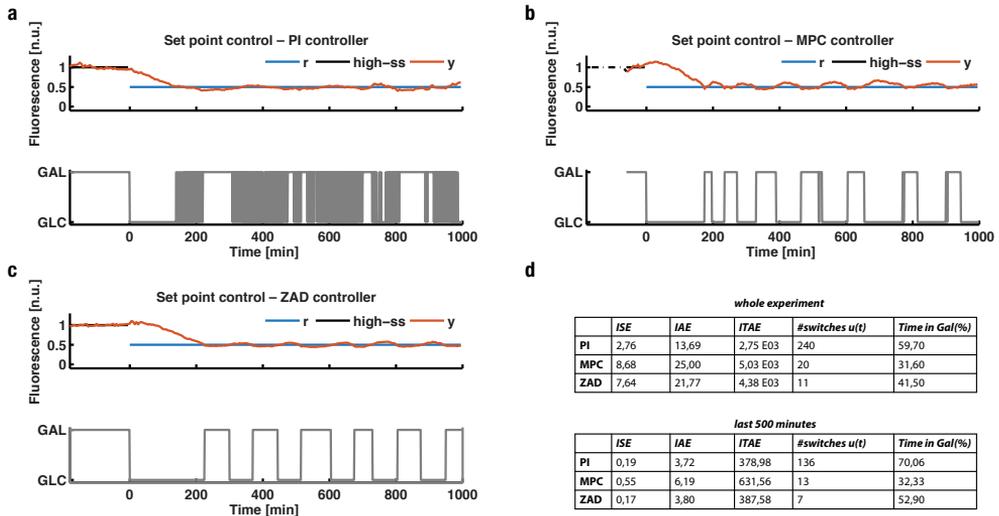


Figure 4.6: *In vivo* setpoint control task. The black line is the average fluorescence intensity during the calibration phase of 180 min. The blue line is the reference signal (r). The orange line is the measured fluorescence level (y) across the yeast population. The grey line is the control input (u). (a–c) Three *in vivo* setpoint control experiments by the means of the PI (a), MPC (b) and ZAD (c) controllers. The control action starts at time $t = 0$ min and ends at $t = 1000$ min. d) Performance measures: Integral Square Error (ISE), Integral Absolute Error (IAE), Integral Time Absolute Error (ITAE), number of switches of the control input, and the percentage of time during which the model is provided with the 'ON' input.

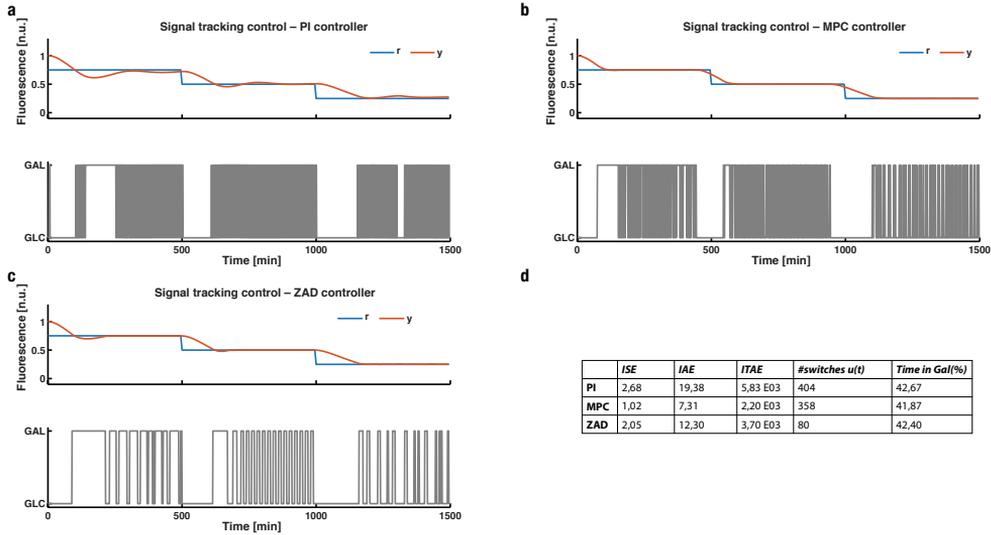


Figure 4.7: **Numerical simulation of the staircase tracking control task.** The blue line is the reference signal (r). The orange line is the simulated fluorescence level (y). The grey line is the control input (u). (a–c) Three *in silico* staircase tracking control experiments performed on the mathematical model of the galactose–inducible promoter by the means of the PI (a), MPC (b) and ZAD (c) controllers. The initial level of fluorescence is assumed to be equal to 1 (n.u.). The control action starts at time $t = 0$ min and ends at $t = 1000$ min. d) Performance measures: Integral Square Error (ISE), Integral Absolute Error (IAE), Integral Time Absolute Error (ITAE), number of switches of the control input, and the percentage of time during which the model is provided with the 'ON' input.

achieves satisfying results, comparable to the MP Controller (except in the proximity of the steps' edges), but with a smaller number of galactose pulses.

In vivo tracking control task for the descending staircase tracking control task (Fig. 4.8) confirm *in silico* results. The PI controller (Fig. 4.8a) poorly tracks the reference r , despite the high number of control input switches. The MPC, as already demonstrated by numerical simulations, has a much better performance, quantitatively confirmed by the performance indices (cf. Fig. 4.8b and Fig. 4.8d). As in the case of the *in silico* simulations, the ZAD controller (cf. Fig. 4.8c and Fig. 4.8d) achieves a performance comparable to that of the MPC (even if not as good in the proximity of the steps' edges)

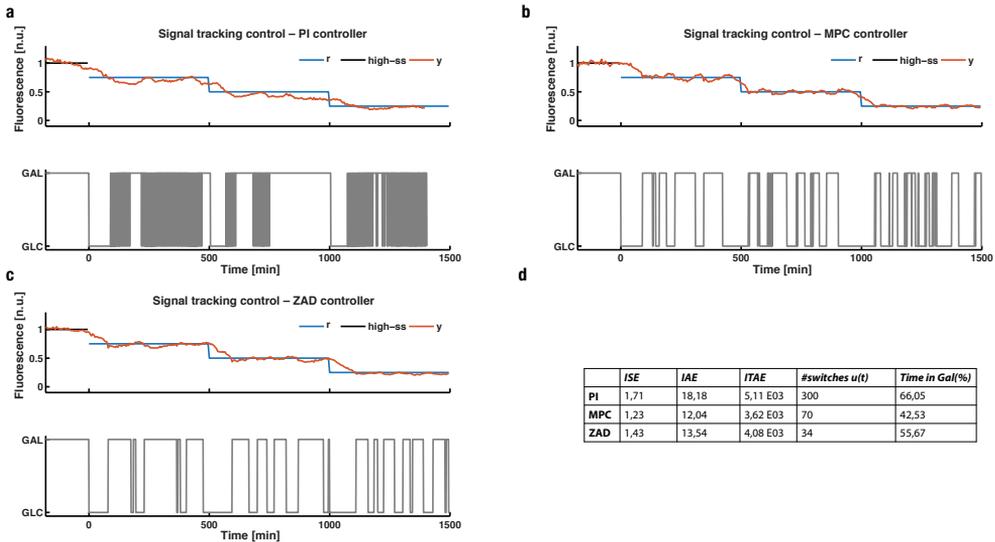


Figure 4.8: ***In vivo* tracking control task for the descending staircase tracking control task.** The black line is the average fluorescence intensity during the calibration phase of 180 min. The blue line is the reference signal (r). The orange line is the measured fluorescence level (y) across the yeast population. The grey line is the control input (u). (a–c) Three *in vivo* staircase tracking control experiments by the means of the PI (a), MPC (b) and ZAD (c) controllers. The control action starts at time $t = 0$ min and ends at $t = 1000$ min. d) Performance measures: Integral Square Error (ISE), Integral Absolute Error (IAE), Integral Time Absolute Error (ITAE), number of switches of the control input, and the percentage of time during which the model is provided with the 'ON' input.

by employing fewer control input switches than the MPC.

Because of the poor tracking results achieved by the PI controller, I decided to compare only the MPC and ZAD strategies when tracking the ramp and the sinusoidal signals. Both *in silico* numerical simulations (Fig. 4.9) and *in vivo* (Fig. 4.10) experiments confirmed that the ZAD controller is able to guarantee a performance (Fig. 4.9e and Fig. 4.10e) similar to that of the MPC strategy, but again with a reduced number of control input switches.

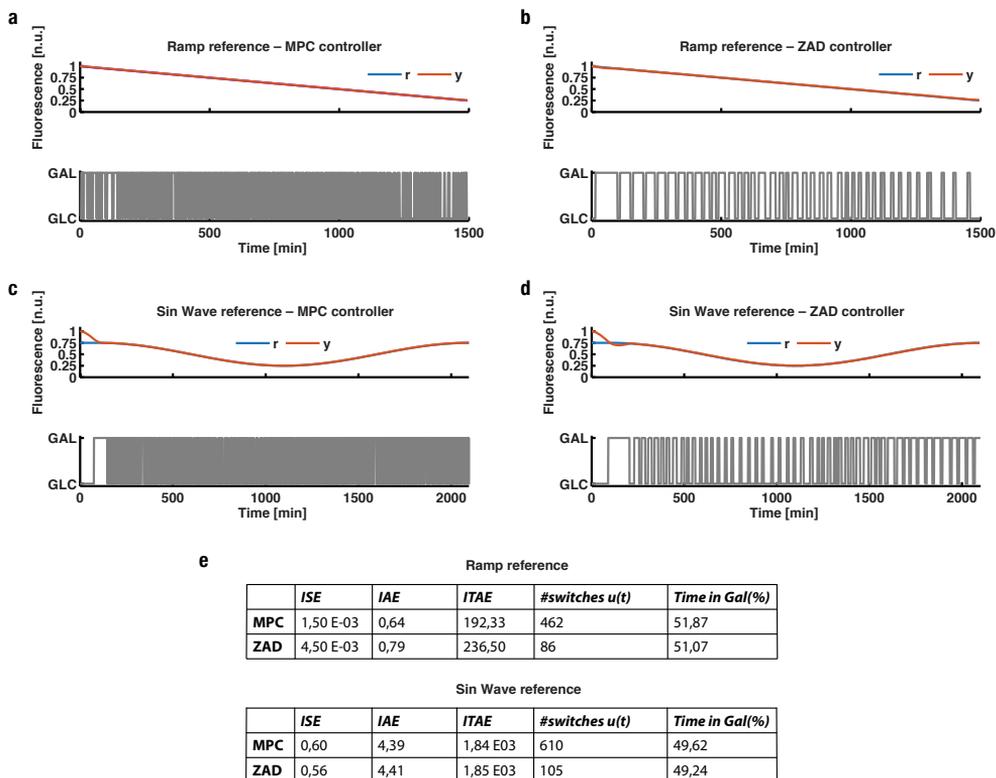


Figure 4.9: **Numerical simulations of the ramp and sine wave tracking control tasks.** The blue line is the reference signal (r). The orange line is the simulated fluorescence level (y). The grey line is the control input (u). (a–b) Two *in silico* ramp tracking control experiments performed on the mathematical model of the galactose-inducible promoter by the means of the MPC (a) and ZAD (b) controllers. The initial level of fluorescence is assumed to be equal to 1 (n.u.). The control action starts at time $t = 0$ min and ends at $t = 1500$ min. (c–d) Two *in silico* sin wave tracking control experiments performed on the on the mathematical model of the galactose-inducible promoter by the means of the MPC (c) and ZAD (d) controllers. The initial level of fluorescence is assumed to be equal to 1 (n.u.). The control action starts at time $t = 0$ min and ends at $t = 2100$ min. e) Performance measures: Integral Square Error (ISE), Integral Absolute Error (IAE), Integral Time Absolute Error (ITAE), number of switches of the control input, and the percentage of time during which the model is provided with the 'ON' input.

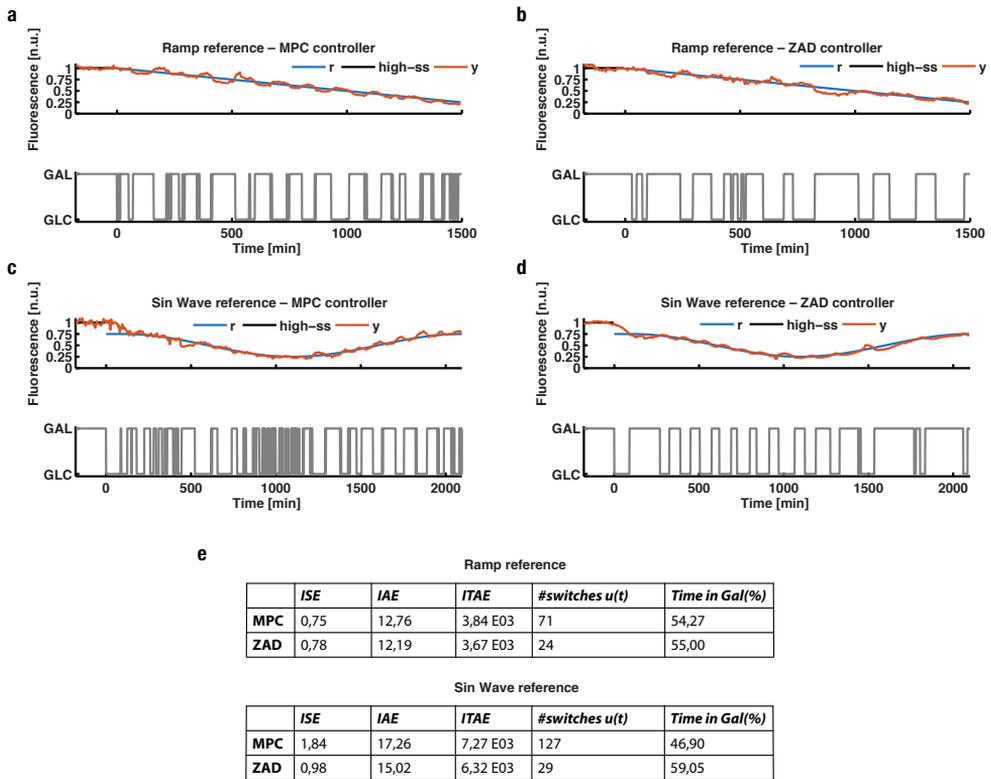


Figure 4.10: *In vivo* experiments of ramp and sine wave tracking control tasks. The black line is the average fluorescence intensity during the calibration phase of 180 min. The blue line is the reference signal (r). The orange line is the measured fluorescence level (y) across the yeast population. The grey line is the control input (u). (a–b) Two *in vivo* ramp tracking control experiments by the means of the MPC (a) and ZAD (b) controllers. The control action starts at time $t = 0$ min and ends at $t = 1500$ min. (c–d) Two *in vivo* sine wave tracking control experiments by the means of the MPC (c) and ZAD (d) controllers. The control action starts at time $t = 0$ min and ends at $t = 2100$ min. e) Performance measures: Integral Square Error (ISE), Integral Absolute Error (IAE), Integral Time Absolute Error (ITAE), number of switches of the control input, and the percentage of time during which the model is provided with the 'ON' input.

4.6 Discussion

In this Chapter, I provided a comparative analysis, *in silico* and *in vivo*, of three different control strategies for regulating gene expression from the galactose-inducible promoter. To this end, I implemented and compared PI and MPC controllers, which have been previously reported in the literature [7–10] and proposed an additional strategy, the ZAD controller [35].

I demonstrated that both MPC and ZAD control strategies can be successfully employed to control gene expression from the galactose-inducible promoter to generate any desired time-varying concentration of the reporter protein (GFP). These controllers require a quantitative model of the system to be controlled. This is not a strong limitation, since it is possible to identify a dynamical input-output model of the biological system under investigation using standard system identification techniques, which work very well at least for simple inducible promoters [16, 23, 31].

The PI controller, as expected from control theory [1] and from numerical simulations as reported in this Chapter, performs similarly to the MPC and ZAD strategies only in the setpoint control task, whereas it is the worst performer in the case of signal-tracking experiments.

The MPC and ZAD controllers perform similarly well in all the control tasks. The main differences are that the MPC performs slightly better than ZAD for fast switching reference signals (such as the staircase signal in Fig. 4.8), however it requires a higher number of input switches when compared to the ZAD controller. The ZAD technique may be advantageous in those applications in which a high cost is associated to the actuation such as when the input administration can cause stress to the cells (e.g light stimuli, antibiotic, osmotic shocks etc.).

In conclusion, the choice of the control strategy to employ will depend on which kind of control task needs to be achieved (setpoint or tracking), the complexity of the synthetic circuit to be controlled, the availability of a descriptive mathematical model of the circuit to be controlled, the cost associated to the actuation effort and, whether a minimal stress to the cells is required (i.e. a small number of input switches).

Chapter 5

Analysis, modeling and control of single-cell gene expression in yeast

Gene expression is an intrinsically noisy process, since the amount of protein produced by a particular gene is affected by stochastic fluctuations due to low copy numbers of regulatory molecules contained in living cells [29]. The stochasticity, or noise, of gene expression has been largely investigated in the literature [38–42]. In summary, gene expression noise can be classified as intrinsic, when it is caused by stochastic effects inherent the biomolecular process of gene expression, and extrinsic, when it is caused by variations in the level or activity of other cellular components involved in the process [29]. In eukaryotes, like yeast and mammalian cells, extrinsic effects predominate [43], conversely in prokaryotes, like bacterial cells, intrinsic effects predominate [44].

Recently, a mixed-effects dynamical modelling framework has been shown to correctly capture the dynamical behavior of gene expression from the STL1 promoter at both single cell and population levels in yeast cells [45]. In this framework, expression variability across cells is assumed to be caused by stable differences among cells, which are modeled with a deterministic linear time-invariant model with cell-specific parameters but with the same

structure for all the cells. This class of models makes it possible to implement more realistic simulators to describe the dynamics of gene expression in individual cells, and hence simulate not only the average fluorescence level but also the variance of the fluorescence across the cell population [45, 46].

In this Chapter, I developed and used an offline cell segmentation and tracking algorithm to collect single cell gene expression data from the time-lapse experiments presented in Chapter 4. Thus, I performed an analysis of the single cell data, characterising the intrinsic and extrinsic noise of gene expression from the galactose-inducible promoter. I then verified that a mixed-effect dynamical model could correctly describe the observed variability in protein expression both in individual cells and at the population level. Moreover, I proposed a novel model-predictive-control approach able to steer the average fluorescence level across a population of yeast cells, by using single cell models to describe and predict the overall dynamics of the cell population. This control strategy is different from the one that I and others have applied in the past, since it is not based on an average model describing the dynamics of the average fluorescence level of the entire cell population, but on a set of models describing gene expression at single cell level. I proposed to name this control strategy, maybe improperly, Mixed-effects Model Predictive Control (MEMPC).

The contents of this Chapter are part of an ongoing work and has been partially published in [17].

5.1 Single cell quantification of protein expression

Single cell time-series of protein expression were obtained by means of a custom offline segmentation and tracking algorithm, able to calculate the fluorescence level of each yeast cell in a time-lapse experiment. For this aim, I adapted and improved the algorithm first proposed in [26] and extended in [23].

5.1.1 Image segmentation and tracking algorithm for yeast cells

The custom image processing algorithm was developed and implemented as a series of *scripts* and *functions* in the MATLAB (The MathWorks, Inc.) environment. The algorithm receives in input the sequence of images in a time-lapse experiment, and returns in output a set of time-series related to each one of the tracked cells, consisting of the single-cell fluorescence levels and mother-daughter relationships among the cells in order to track also the phylogeny of each yeast cell.

Basically, in order to measure the fluorescence level of a single cell over time, a three-step image analysis approach was devised:

1. detection of the cells' position within each image (segmentation);
2. tracking of the movements of each cell in the entire image sequence (tracking);
3. generation of a lineage tree describing the mother-daughter relationships among cells (phylogenetic analysis).

Further details on the implementation of the segmentation and tracking algorithm are provided in Appendix A.

5.1.2 Single-cell fluorescence time-series data

I ran offline the image processing algorithm on the time-lapse control experiments described in Chapter 4, thus obtaining a total of ten different data sets. I decided to consider only the fluorescence level expressed by single cells, and to discard mother-daughter relationships among cells. Data sets were enumerated with progressive numbers, and the correspondence between data sets and time-lapse experiments are reported in Table 5.1.

An example to elucidate of the single cell fluorescence data collected in each experiment is shown in Fig. 5.1.

Table 5.1: **Correspondence between data sets and time-lapse experiments.**

	Time-lapse experiment	Reference figure
Data set 1	Setpoint – PI	Fig. 4.6a
Data set 2	Setpoint – MPC	Fig. 4.6b
Data set 3	Setpoint – ZAD	Fig. 4.6c
Data set 4	Signal tracking – PI	Fig. 4.8a
Data set 5	Signal tracking – MPC	Fig. 4.8b
Data set 6	Signal tracking – ZAD	Fig. 4.8c
Data set 7	Ramp – MPC	Fig. 4.10a
Data set 8	Ramp – ZAD	Fig. 4.10b
Data set 9	Sine wave – MPC	Fig. 4.10c
Data set 10	Sine wave – ZAD	Fig. 4.10d

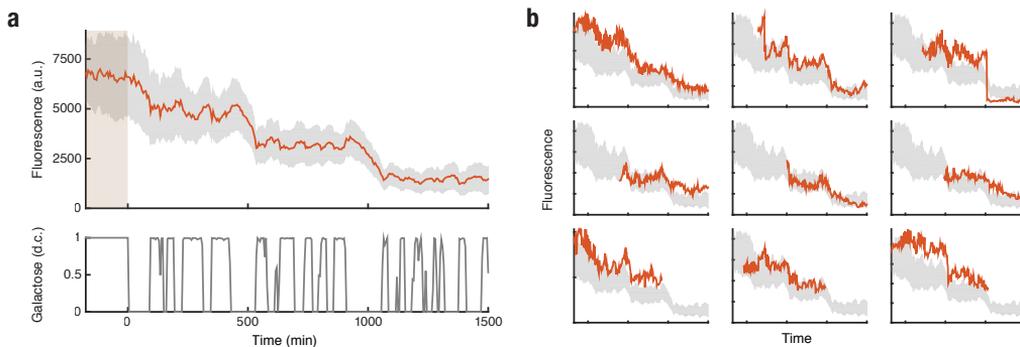


Figure 5.1: **Single cell fluorescence data extrapolated from data set 5.** Fluorescence measurements were obtained by means of an offline tracking algorithm that computes the single cell fluorescence for each cell in each image. **(a)** Averaged fluorescence level across the cell population. The orange line is the mean of single cell fluorescence data. The shaded grey area is the standard deviation of the single cell fluorescence data. The shaded box is the initial calibration phase of 180 min at the beginning of each time-lapse experiment, as described in Chapter 4. The grey line in bottom panel is the duty cycle (d.c.) of the galactose pulse administered to the cells during the time-lapse experiment. **(b)** Each panel reports fluorescence level for a single cell (orange line). Single cells were chosen randomly as examples among cell ensemble.

5.2 Analysis of single cell fluorescence data

In order to derive a mathematical representation describing the behavior of protein expression both at single cell and cell population levels, I had to understand and characterise gene expression noise from the data. Thus, I performed an analysis to quantify the variability of gene expression from the galactose-inducible promoter at single cell and population levels. The single cell fluorescence data were extrapolated from the data sets presented in the previous Section 5.1.

5.2.1 Quantification and analysis of cellular noise at steady-state

Variability in protein expression was quantified both at single cell and population levels by considering only fluorescence data measured during the initial calibration phase of each time-lapse experiment (e.g., shaded box in Fig. 5.1a). Before and during this phase, cells were kept in galactose enriched medium, in order to fully activate the expression of the reporter protein from the galactose-inducible promoter. Hence, during the calibration phase, protein expression can be considered to be at steady state (*galactose steady state*).

The distribution of single-cell fluorescence level of cells growing in galactose were analysed for a total of ten replicates, and the results are depicted in Fig. 5.2. Each panel of Fig. 5.2 (a–j) corresponds to an experimental replicate and it consists of three insets. The central inset reports the fluorescence distribution at the single cell level, i.e. the variability in protein expression in the cell due to the inherent stochasticity of the biomolecular reactions (intrinsic noise); the left inset reports the fluorescence distribution across the population, i.e. the cell to cell variability in protein expression, which represents the total noise afflicting the biomolecular process of gene expression, and it is composed by intrinsic and extrinsic noise [29]. Since in eukaryotes extrinsic effects predominate over intrinsic effects [43], I considered the cell to cell variability as a source of only extrinsic noise. Cell to

cell variability was obtained by analysing fluorescence of all the cells in the population at each time point. Instead, single cell variability was obtained by analysing fluorescence of a single cell for all the time points. Lastly, the right inset compares the coefficient of variation ($CV = \frac{\sigma}{\mu}$) of the intrinsic and extrinsic noise [29].

At the population level, cells exhibited very similar distributions of fluorescence across time, as can be observed in the left insets of Fig. 5.2a–j reporting the fluorescence distribution across the cell population as a box plot. It can be observed that mean and standard deviation of the fluorescence remain similar as time progresses (Fig. 5.2a–j – Left inset). I also estimated the fluorescence distribution within each cell, which can be considered to be caused by intrinsic noise (Fig. 5.2a–j – Central inset). It can be appreciated that the mean fluorescence, and to a lesser extent the standard deviation, change for each cell (Fig. 5.2a–j – Central inset). Interestingly, standard deviations for population data (Fig. 5.2a–j – Left inset) are much greater compared to the standard deviations of single cell data (Fig. 5.2a–j – Central inset). I also compared the level of intrinsic and extrinsic noise by estimating the coefficient of variation, demonstrating the predominance of the extrinsic noise over intrinsic noise (Fig. 5.2a–j – Right inset). Taken together, these results confirm that yeast cells exhibit a large cell to cell variability (i.e. extrinsic noise) and a low variability at the single cell level (i.e. intrinsic noise), in line with what is expected from eukaryotic cells [43].

5.2.2 Quantification and analysis of cellular noise dynamics

Variability in protein expression was quantified both at single cell and population levels by considering fluorescence data measured during the entire time-lapse experiments, during which cells undergo a series of growth medium switches between galactose-rich and glucose-rich growth media. Since in the previous section I demonstrated that the cells exhibit much higher extrinsic noise than intrinsic noise, in what follows I decided to ne-

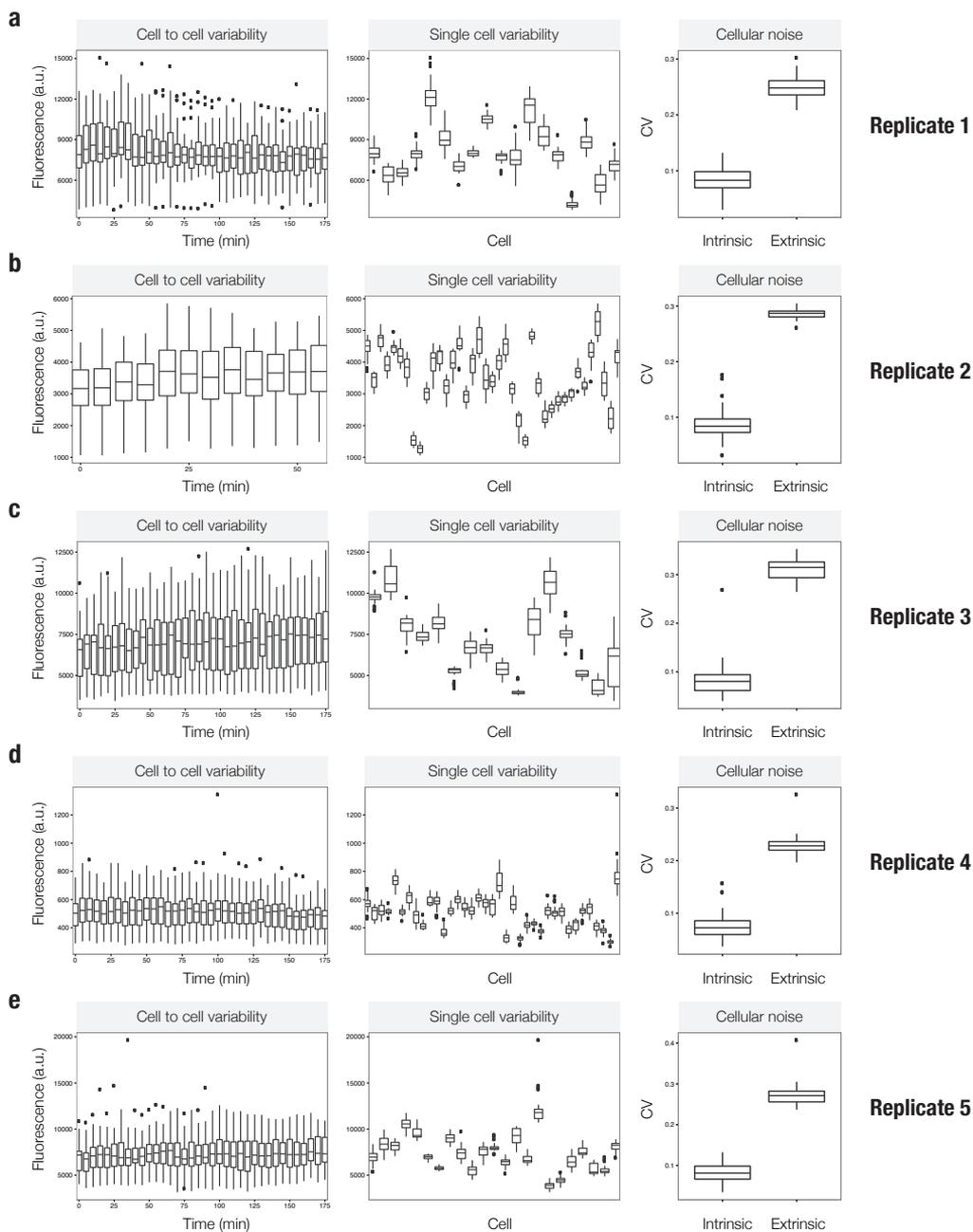


Figure 5.2: **Cell to cell variability, single cell variability and cellular noise when cells are grown in galactose.** (a–j) Each panel represents an experimental replicate, for a total of ten replicates. (**Left inset**) Fluorescence distribution across the population, i.e. the cell to cell variability in protein expression. (**Central inset**) Fluorescence distribution at the single cell level, i.e. the variability in protein expression in the single-cell. (**Right inset**) Coefficient of variation computed for the intrinsic and extrinsic noise.

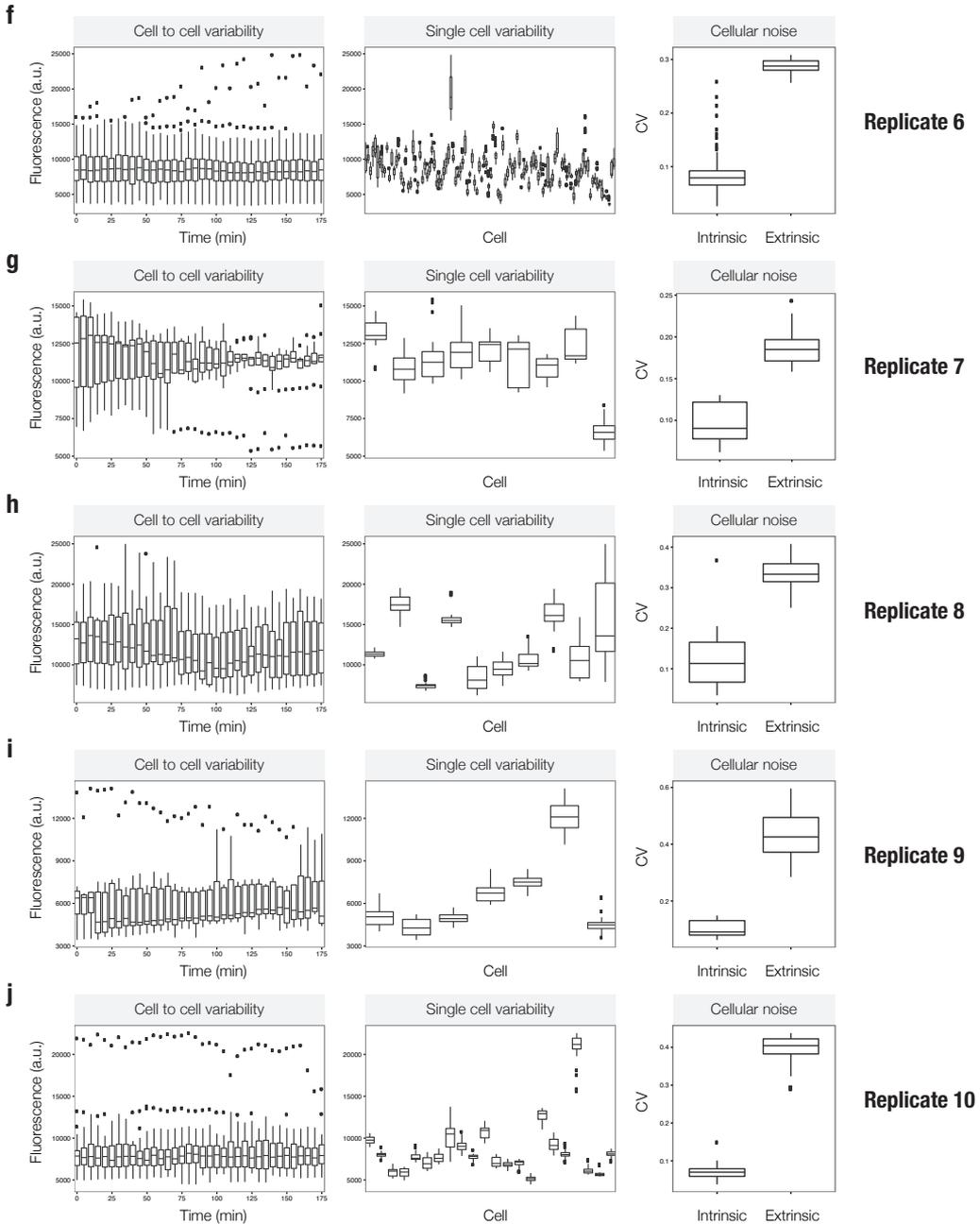


Figure 5.2: Cell to cell variability, single cell variability and cellular noise during the galactose steady state. (...)

Table 5.2: **Duration of each phase for the several time-lapse experiments.**

Time-lapse experiment	Initial calibration	T1	T2	T3
Setpoint – PI	180 min	200 min	620 min	180 min
Setpoint – MPC	60 min	200 min	620 min	180 min
Setpoint – ZAD	180 min	200 min	620 min	180 min
Signal tracking – PI	180 min	500 min	500 min	500 min
Signal tracking – MPC	180 min	500 min	500 min	500 min
Signal tracking – ZAD	180 min	500 min	500 min	500 min
Ramp – MPC	180 min	500 min	500 min	500 min
Ramp – ZAD	180 min	500 min	500 min	500 min
Sine wave – MPC	180 min	100 min	1075 min	1075 min
Sine wave – ZAD	180 min	100 min	1075 min	1075 min

glect intrinsic noise. Hence, cellular noise (η_t) consists only of the extrinsic noise (η_e) with a coefficient of variation (CV) estimated from single cell fluorescence data grouped for each time point. I divided each time-lapse experiment into four consecutive temporal phases, in order to better understand and characterise noise dynamics: *galactose steady state*, T1, T2, and T3. Table 5.2 details the duration of each phase for each of the time-lapse experiments used to in this analysis.

The noise distributions during the entire time-lapse experiments are depicted in Fig. 5.3. The noise distributions are grouped according the different set-point and signal tracking control tasks, in order to understand if control strategies could affect the cell noise at population level (i.e. extrinsic noise). Interestingly, control strategies did not affect cellular noise neither in the set-point (Fig. 5.3a) nor in the signal tracking control tasks (Fig. 5.3b–d). Moreover, as expected from the literature [47], the coefficient of variation is a function of the absolute fluorescence levels, decreasing exponentially with the increase of fluorescence levels (e.g. Fig. 5.3b).

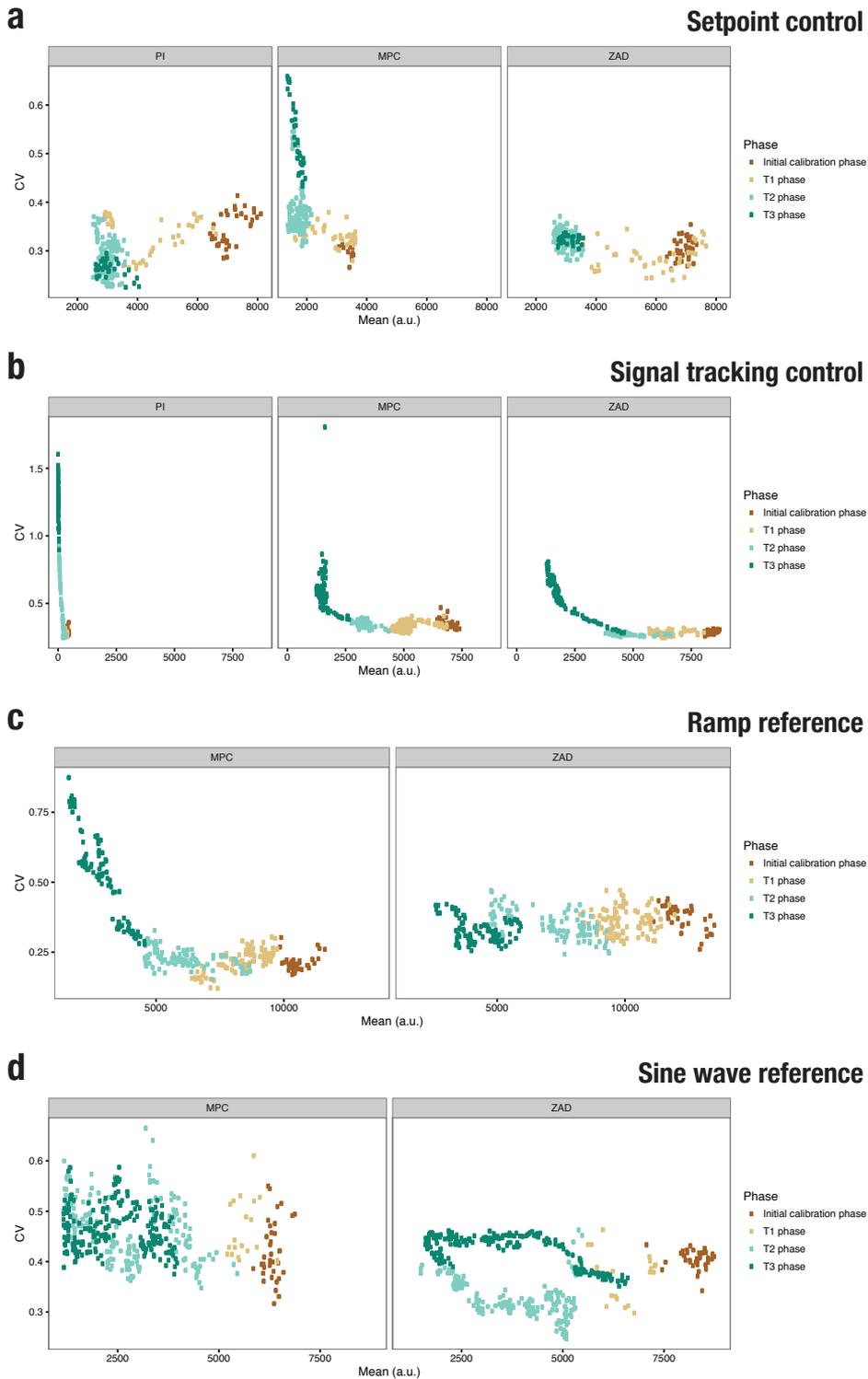


Figure 5.3: **Cellular noise during time-lapse control experiments.** Coefficients of variation (CV) computed on single cell gene expression data grouped for time points.

5.3 Mathematical representation of single-cell gene expression from the galactose-inducible promoter

Starting from the findings of Section 5.2, I decided to improve the modeling of fluorescence dynamics caused by the expression of the reporter gene from the galactose-inducible promoter by adopting a mixed-effects modelling approach, which has been recently applied in yeast cells [45]. In this way, cell-to-cell variability can be modeled by mixed-effects linear models, thus capturing protein expression dynamics both in single cell and at population level.

5.3.1 Mixed-effects model as an ideal framework for describing cell to cell variability

Mixed-effects models are suitable to describe mathematically ensemble of similar entities, such as a cells within a population. Each entity of the ensemble is described by a mathematical model with a fixed structure, equal for all the entities in the ensemble, and a set of parameters different for each entity [45]. By adopting a mixed-effects approach for modeling gene expression from the galactose-inducible promoter, the dynamics of the fluorescence level in a single cell can be described as a single-input single-output (SISO) system of two coupled linear ordinary differential equations defined as:

$$\frac{d}{dt} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -\alpha & 0 \\ \gamma & -\delta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \beta \\ 0 \end{bmatrix} u, \quad x(0) = x_0 \quad (5.1)$$

$$y = x_2$$

where x is the vector state, x_0 is the vector of initial conditions, u is the input (glucose or galacote), and y (the measured fluorescence in a single cell) is the output. Note that in Eq. (5.1) u is the only external stimulus

to the model and it is assumed to be equal to 1 when single cell is fed with galactose, whereas, when glucose is provided to cell, it is assumed to be equal to 0 (note that these values are related to the concentration of galactose in the growing medium).

Intrinsic noise could be modeled by transforming the set of ordinary differential equations in a set of stochastic differential equations, considering the presence of multiplicative and additive white Gaussian noise on the state equations. However, I decided to neglect these effects to simplify both numerical simulations and analytical approaches.

As stated before, the mixed-effects model is defined by a mathematical structure fixed for all the entities in the ensemble, in this case the structure of the state-space linear model; with model parameters $\psi = \{\alpha, \beta, \gamma, \delta\}$ changing from one entity to another entity. I considered the mathematical structure of the mixed-effects model, comprising the vector state x and the set of parameters ψ , without physical meanings. I also assumed that the set of parameters ψ varies within the population.

5.3.2 Mixed-effects model inference

As in [45], I assumed that model parameters $\psi = \{\alpha, \beta, \gamma, \delta\}$ were distributed across the cell population with a multivariate log-normal distribution, i.e. $\ln(\psi) \sim \mathcal{N}(\mu, \Sigma)$ is a multivariate normal distribution defined by the means vector μ and the covariance matrix Σ .

Thus, the mixed-effects model is defined mathematically by its structure, represented in Eq. (5.1); and the multivariate log-normal distribution, which is defined by its center of mass, i.e. a vector of means; and its spread, i.e. a covariance matrix, across cell population. I inferred the distribution's parameters with the simplest approach presented in [45], that the authors named *naive approach*. In the *naive approach*, the distribution's parameters are inferred in two steps. In the first step, the task is to estimate numerically the best parameters that describe the expression dynamics of each single cell; and then, in the second step, to compute the statistics of the underlying log-normal distribution from the collected sets of parameters.

I extrapolated the single cell fluorescence data from the data sets presented in Section 5.1, considering only the cells present since the initial calibration phase of each time-lapse control experiment. I then applied a grey-box state-space identification technique to fit the model parameters $\psi = \{\alpha, \beta, \gamma, \delta\}$ to fluorescence data available for each cell [48].

I discarded the single cell model whose fitting percentage (FIT – see Appendix A for further details about the formula) was smaller than the 50%, thus obtaining a total of $N = 112$ sets (one set for each cell) of model parameters $\{\alpha_n, \beta_n, \gamma_n, \delta_n\}$. Thus, I collected each set of model parameters in the matrix Ψ defined as

$$\Psi = \begin{bmatrix} \alpha_1 & \beta_1 & \gamma_1 & \delta_1 \\ \alpha_2 & \beta_2 & \gamma_2 & \delta_2 \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_N & \beta_N & \gamma_N & \delta_N \end{bmatrix} = \begin{bmatrix} \psi_1^T \\ \psi_2^T \\ \vdots \\ \psi_N^T \end{bmatrix}. \quad (5.2)$$

The matrix Ψ describes the 112 observations of the four model parameters, and each row vector ψ_n^T denotes an observation of the four parameters.

Starting from the estimated model parameters collected in Ψ , I obtained the vector of means (μ) and the covariance matrix (Σ) of the underlying multivariate log-normal distribution by means of the method of moments, according to the procedure reported in [45]:

$$\begin{aligned} \mu &= \frac{1}{N} \sum_{n=1}^N \phi_n \\ \Sigma &= \frac{1}{N-1} \sum_{n=1}^N (\phi_n - \mu)(\phi_n - \mu)^T \end{aligned} \quad (5.3)$$

where $\phi_n = \log(\psi_n)$.

The estimated statistics for the multivariate log-normal distribution by

means of the method of moments were:

$$\mu = \begin{bmatrix} -4.2260 \\ -3.8739 \\ -5.4030 \\ -5.0355 \end{bmatrix} \quad (5.4)$$

and

$$\Sigma = \begin{bmatrix} 1.7494 & 1.8165 & -0.2996 & -0.2654 \\ 1.8165 & 3.5977 & -1.7806 & -0.2930 \\ -0.2996 & -1.7806 & 2.3108 & 0.8920 \\ -0.2654 & -0.2930 & 0.8920 & 0.8560 \end{bmatrix} \quad (5.5)$$

5.3.3 Numerical simulations of the mixed-effects model of protein expression from the galactose-inducible promoter in yeast cells

By numerical simulations, I tested the performance of the identified mixed-effects model by considering an ensemble of 100 cells, whose model parameters were sampled randomly from the multivariate log-normal distribution. I reported the model parameters for the cell ensemble in Appendix B.

Numerical simulations were carried out by simulating the response of the cell ensemble to the same input as the one applied during the time-lapse control experiments and detailed in Chapter 4. By way of example, I show in Fig. 5.4 the results of the numerical simulation for the input depicted in the bottom panel of Fig. 5.1a. The mixed-effects model was able to capture dynamics of protein expression both for the entire population (Fig. 5.4a) and for single cells (e.g., Fig. 5.4b). The overall behaviour of the cell population was captured through mean (Fig. 5.4a; orange line) and standard deviation (Fig. 5.4a; gray area) of fluorescence level across the cell population. Fig. 5.4b shows the simulated fluorescence levels for nine cells chosen randomly in the population. By comparing experimental (Fig. 5.1) and simulated (Fig. 5.4) data, I concluded that a mixed-effects model can successfully describe the dynamics of gene expression from the galactose-inducible pro-

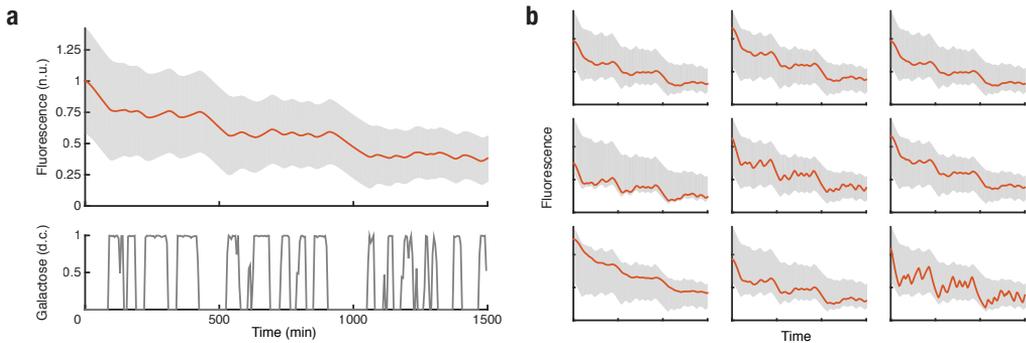


Figure 5.4: **Numerical simulation of the mixed-effects dynamical model.** An open loop experiment was simulated by applying the same input as the one used in the time-lapse experiment depicted in Fig. 5.1a. **(a)** Averaged behavior of simulated fluorescence dynamics. The orange line is the mean of simulated single cell fluorescence data. The grey area is the standard deviation of simulated single cell fluorescence data. The grey line in the bottom panel is the input given to the mixed-effects model and to the average model during simulation. **(b)** Each panel reports simulated fluorescence levels for one single cell (orange line). Cells were chosen randomly from cell ensemble.

motor both at the population and single cell levels, as previously reported for the STL1 promoter [45].

I then verified whether the mixed-effects model was able to describe the experimentally observed cell to cell variability (extrinsic noise) by quantifying the coefficient of variation (CV) on the simulated single cell expression levels at each time point.

The numerical simulations confirmed that a mixed-effects model is able to replicate the experimentally observed extrinsic noise (Fig.5.5). As in the experimental data, the specific control strategy did not affect the variability across the cell population. Interestingly, the CV tends to decrease when the average fluorescence level reaches a setpoint value, as can be appreciated in Fig. 5.5a–b; however the CV computed on experimental data did not show this behavior (Fig. 5.3a–b). The reason for this discrepancy maybe due to the exclusion of the intrinsic noise component that was not modeled in the mixed-effects model.

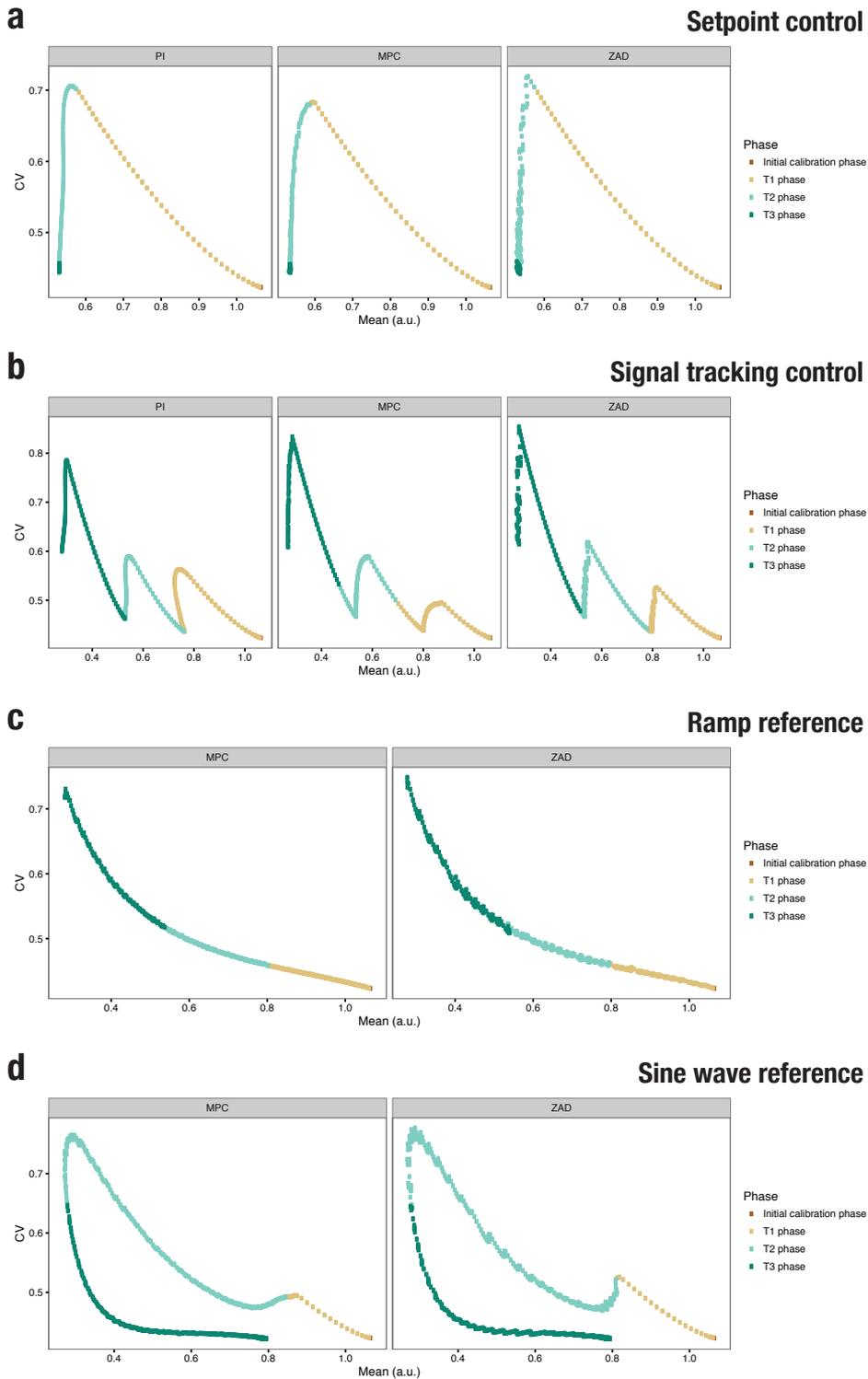


Figure 5.5: **Cell noise during numerical control simulations.** Coefficients of variation (CV) computed on single cell gene expression data grouped for time points.

5.4 A novel model–predictive–control strategy for controlling an ensemble of yeast cells

In order to improve the control performance, I proposed a novel MPC strategy, that employs single cell models to predict the overall effect of each cell on the average behavior of the cell population. In agreement with the mixed–effects modeling approach, I assumed a fixed dynamical structure for the single cell models, and a set of parameters different for each cell, as previously described in Section 5.3. Each single cell model was then used to predict the average behavior of the cell population in time, exploiting the characteristic of mixed–effects model to capture the cell to cell variability. I called this approach Mixed–effects Model Predictive Control (MEMPC).

The proposed MPC scheme with single cell models is depicted in Fig. 5.6b. In order to understand the differences between the proposed controller and the one presented in Chapter 4, I also report in Fig. 5.6a the MPC scheme with population–averaged model. The main difference between the control schemes lies in the *mean operator*, that in the case of Fig. 5.6a is upstream of the population–averaged model; conversely in the case of Fig. 5.6b it is found downstream of the single cell models.

5.4.1 Numerical simulations of the MEMPC strategy to steer protein expression from the galactose–inducible promoter

In order to assess performance of the proposed MEMPC strategy, I performed numerical simulations on the same ensemble of 100 cell sampled in Section 5.3.3 (see Appendix B for single cell model parameters). I chose as control tasks the reference signals presented in Fig. 4.4a–b. In order to add an uncertainty to the model parameters of the single cells used by the MEMPC, I identified again (with a grey–box state–space identification) the model parameters for each cell in the population using the simulated input–output data depicted in Fig. 5.4.

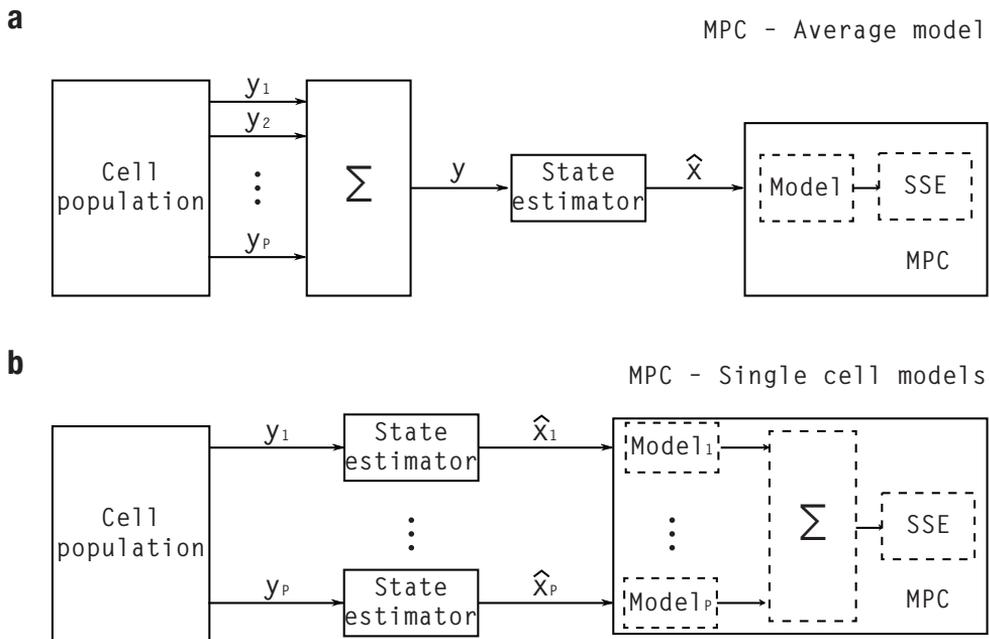


Figure 5.6: **Comparison of Model Predictive Control schemes.** (a) MPC scheme with population-averaged model. (b) MPC scheme with single cell models.

The proposed MEMPC strategy was compared to the MPC strategy presented in Chapter 4, which adopts a population-averaged model to predict the average fluorescence level of the cell population. The average model was identified on the same simulated input-output data (Fig. 5.4a) used to identify the parameters of each single cell model.

Simulated control experiments were performed by means of the MPC strategy using either the average model describing the population-averaged dynamics (Fig. 5.7a-b), or the single cell models (Fig. 5.7c-d). Both implementations achieved the control tasks. Numerical simulations with MEMPC however showed a much smoother control input, hinting to a better performance in predicting the future behaviour of the system being controlled.

5.5 Discussion

In this Chapter, I verified that a mixed-effects dynamical model can correctly describe the variability in fluorescence level both in individual cells and at the population level. In order to achieve this task, I developed a custom offline image segmentation and cell tracking algorithm, and used it to collect single cell gene expression data from the time-lapse experiments presented in Chapter 4. I then performed an analysis on these single cell data, characterising the intrinsic and extrinsic noise of gene expression. I found that extrinsic noise is predominating over intrinsic noise in yeast cells. I then derived a mathematical representation of protein expression neglecting intrinsic noise using the mixed-effects modeling framework.

Finally, I proposed a novel model-predictive-control approach, based on single cell models, for predicting the overall behaviour of the cell population. I decided to name it, maybe improperly, Mixed-effects Model Predictive Control (MEMPC). Numerical simulations confirmed that the proposed strategy could successfully achieve the regulation of gene expression from the galactose-inducible promoter in yeast cells. Interestingly, the proposed approach generates a smoother control input compared to classical implementation of MPC based on a population-averaged model.

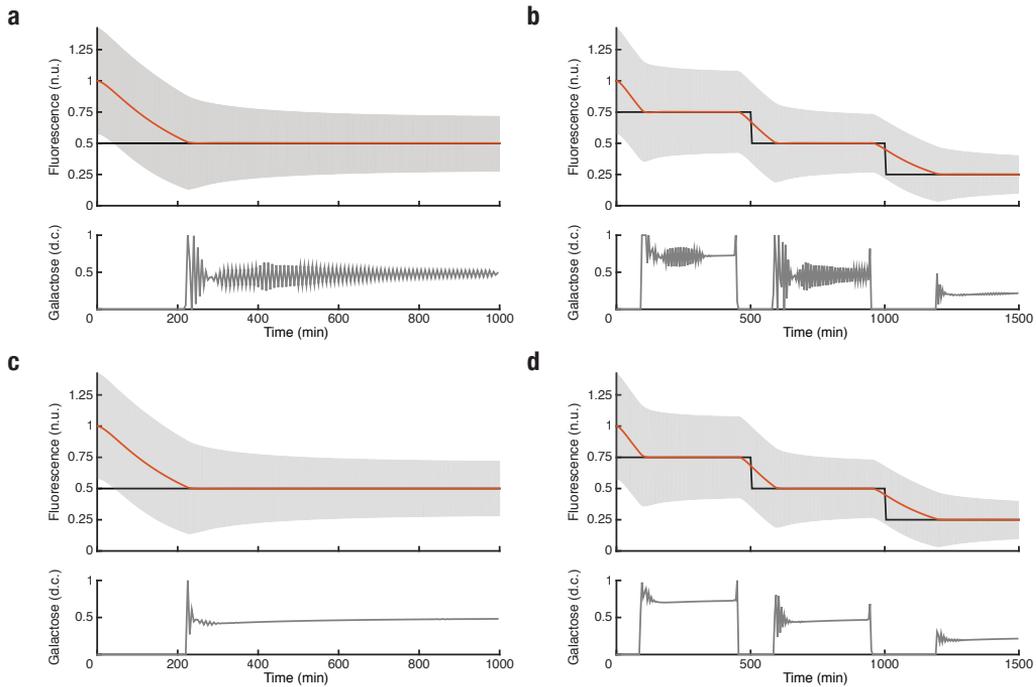


Figure 5.7: **Numerical simulations of the time-lapse control experiments.** The black line is the reference signal. The orange line is the mean of simulated single cell fluorescence data. The shaded grey area is the standard deviation of the simulated single cell fluorescence data. The grey line is the duty cycle (d.c.) of the galactose pulse administered to the cells during simulated time-lapse experiment. Control tasks were achieved by means of a Model Predictive Control strategy using either an average model (**a–b**) or single cell models (**c–d**).

Despite the control performances of the MEMPC and MPC controllers being similar, at least in terms of reference tracking, MEMPC is much more flexible and could have a great impact in practical applications. For example, the proposed MEMPC strategy can be used to control only a subpopulation of cells, by excluding in real-time unhealthy cells that have extreme behaviours (no expression or full expression of the reporter protein independently of the induction medium concentration). MEMPC could also be useful when additional *safety* constraints must be satisfied, such as preventing any cell from expressing the protein being controlled above a certain *toxic* threshold. In this case, the MEMPC could be used to steer average population dynamics, but deviations from the control objective would be allowed if one cell (or a given percentage) is predicted to exceed the *safety threshold*. In addition to the above applications, the MEMPC strategy enables the prediction of the variance of the population over time, and if the experimental system were to allow more than one control input, it could be used to control the variance of the population as well as its mean. In summary, further experimental work is needed to assess the performance of the MEMPC strategy in a real-case scenario, however this control strategy offers a new tool for control applications of living cells.

Chapter 6

Feedback control of human α -synuclein protein expression in yeast

In this Chapter, a practical application of the external feedback control of gene expression is presented. In detail, I explored the feasibility of using the experimental platform presented in Chapter 3 to model and study a pathological hallmark of Parkinson's disease (PD), the most common among neurodegenerative disorders, affecting about 1% of people aged 65 or older worldwide [49].

PD, as other neurodegenerative disorders, is characterised by the progressive disruption of specific neuronal population partly due to the formation of abnormal protein aggregates that interfere with normal cell functions [50]. A neuropathological hallmark of PD is the aggregation of the α -synuclein protein in intraneuronal proteinaceous inclusions, termed Lewy bodies (LBs) or Lewy neurites, that are toxic for neurons [51]. Since the α -synuclein accumulation contributes to PD pathogenesis (further details in Fig. 6.1), it is important to investigate and understand the dynamics of α -synuclein aggregation.

So far, a quantitative understanding of the dynamics of α -synuclein protein aggregation in living cells is lacking. Thus, I investigated the aggre-

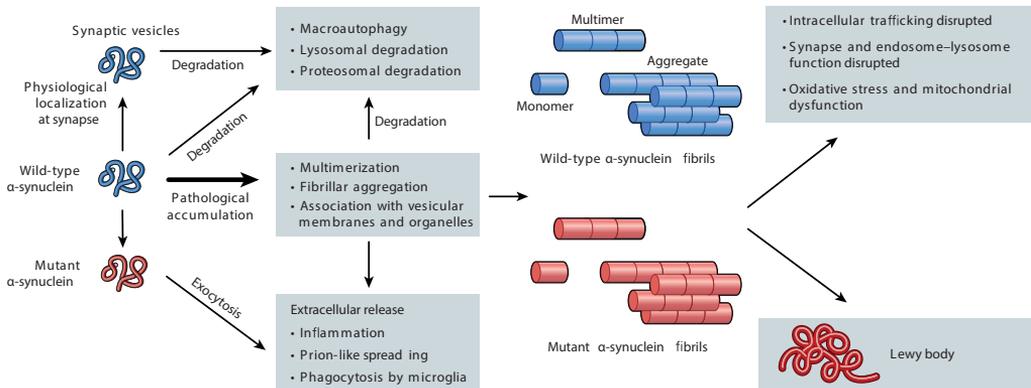


Figure 6.1: The proposed pathological functions of α -synuclein in neurons. When expressed at a moderate level, wild-type α -synuclein (*blue*) associates with synaptic vesicles at axon terminals. Typically, α -synuclein undergoes degradation either through lysosomal or proteasomal pathways. Thus, the pathological accumulation of wild-type α -synuclein can result from an increase in its production or from ineffective degradation or mis-trafficling. These defective mechanisms have also been associated with rare, familial PD-associated mutations in α -synuclein (*red*) as well as mutations in trafficking-associated genes linked to PD and environmental factors such as ageing [52]. An excess of α -synuclein in the form of monomers, multimers or aggregates can disrupt intracellular trafficking and synaptic function and contributes to the formation of LBs. Adapted from [51]

gation dynamics of the human α -synuclein protein in yeast by exploiting the potentiality of the external feedback control and of the experimental platform introduced in Chapter 3. The aim was to attain a quantitative understanding of the dynamics of α -synuclein protein's aggregation by carefully regulating its expression and following its dynamics in real-time in living cells. By regulating the expression level of the α -synuclein protein at different set-points, I can assess quantitatively the threshold protein expression level and the dynamics that lead to the formation of α -synuclein aggregates.

The following work has been conducted in collaboration with Dr. Cathal Wilson of the Telethon Institute of Genetics and Medicine as regards the manipulation of yeast strains, and Dr. Marco Santorelli as regards the molecular cloning of the constructs that were inserted into the yeast strains. Moreover, the contents of this Chapter are part of an ongoing work which has been partially published in [18].

6.1 A quantitative study for protein aggregation in yeast: the human α -synuclein case

Neurodegenerative disorders are associated to the formation of abnormal protein aggregates that interfere with the normal functions of neurons, causing the progressive disruption of the neuronal population. The dysfunction of α -synuclein protein, which is encoded in *SNCA* gene [53], is involved in PD and related neurodegenerative disorders [51]. Mutations in α -synuclein protein are also associated with rare forms of early-onset familial PD [53–59].

The toxicity of the α -synuclein protein has been characterised in several cell-based and organism-based models [60]. For example, yeast strains overexpressing normal and mutant human α -synuclein protein fused to a green fluorescent reporter protein under the galactose-inducible promoter successfully recapitulate protein aggregation observed in PD patients [61]. The overexpression of the human α -synuclein protein in yeast cells mimics

the situation of the aging neurons when the capacity of the quality-control (QC) system to cope with accumulating misfolded proteins is exceeded [61]. One copy of the α -synuclein construct is not able to saturate the QC system of the yeast (Fig. 6.2a). However, when inserting multiple copies of the human α -synuclein gene in yeast, formation of protein aggregates is observed, followed by reduced growth rate and death of yeast cells (Fig. 6.2b). Thus, the toxicity of the α -synuclein protein can be studied using yeast strains carrying multiple copies of the galactose-inducible α -synuclein construct as depicted in Fig. 6.2b. However, as soon as the promoter is activated by growing these cells in galactose enriched medium, α -synuclein toxicity causes cell death thus preventing a thorough investigation and quantification of aggregation dynamics.

External feedback control of the galactose-inducible promoter can overcome these limitations and enable quantitative analysis of α -synuclein dynamics in yeast strains carrying multiple copies of the *SNCA-GFP* construct. Specifically, as depicted in Fig. 6.3, external feedback control of gene expression from the galactose-inducible promoter can be used to increase α -synuclein expression at discrete steps, starting from a fully repressed promoter (glucose steady state), thus enabling precise quantification of the expression level needed for the protein to aggregate as well as analysis of the aggregation dynamics of α -synuclein, both in wild type and mutant forms.

6.2 A pilot study on the feasibility of controlling α -synuclein expression in yeast

I started the investigation on the feasibility of controlling human α -synuclein expression by performing a pilot study on yeast cells carrying only a single copy of *SNCA_{A53T}-GFP* construct driven by the galactose-inducible *GAL1* promoter. This yeast strain expresses human mutant A53T α -synuclein at a moderate level which prevents protein aggregation (Fig. 6.2a). Further details about the yeast strain and how it was generated are provided in Appendix A.

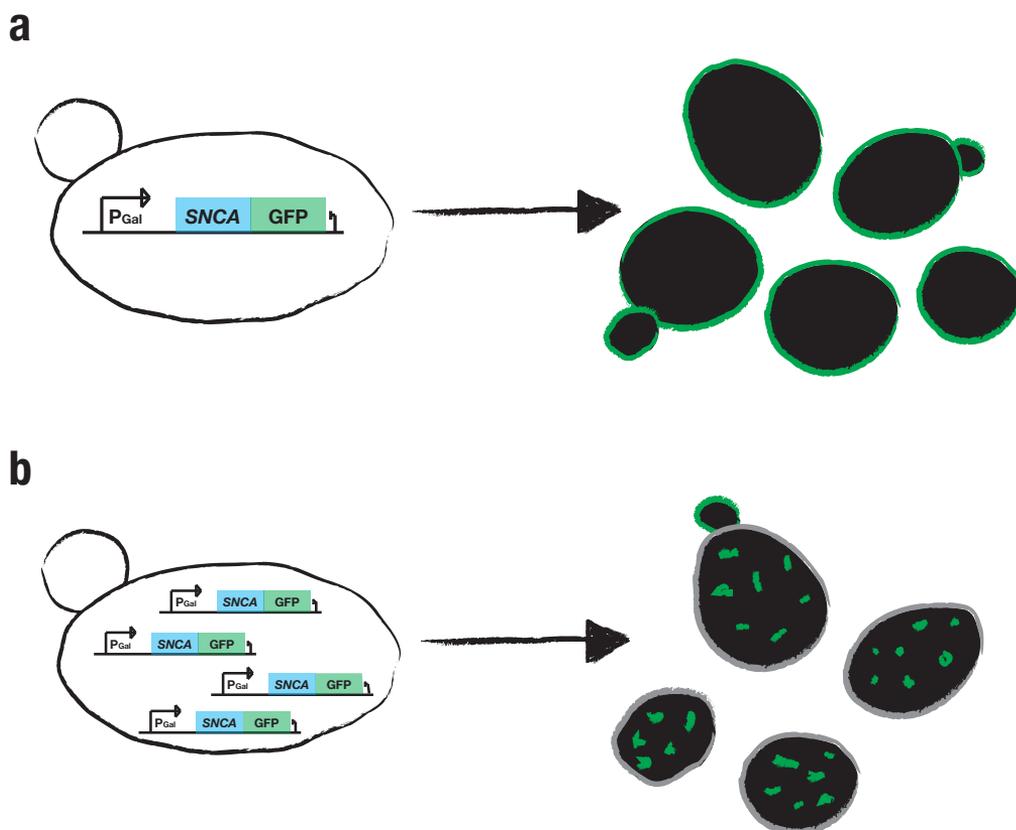


Figure 6.2: **Expression of α -synuclein protein in yeast.** (a) One copy of the α -synuclein construct is not able to saturate the QC system of the yeast. In this case, α -synuclein protein concentrates at the plasma membrane, and small amounts concentrate in the cytoplasm. (b) Multiple copies of the α -synuclein construct are able to saturate the QC system of the yeast, exhibiting cytoplasmic inclusions and reduced membrane localization.

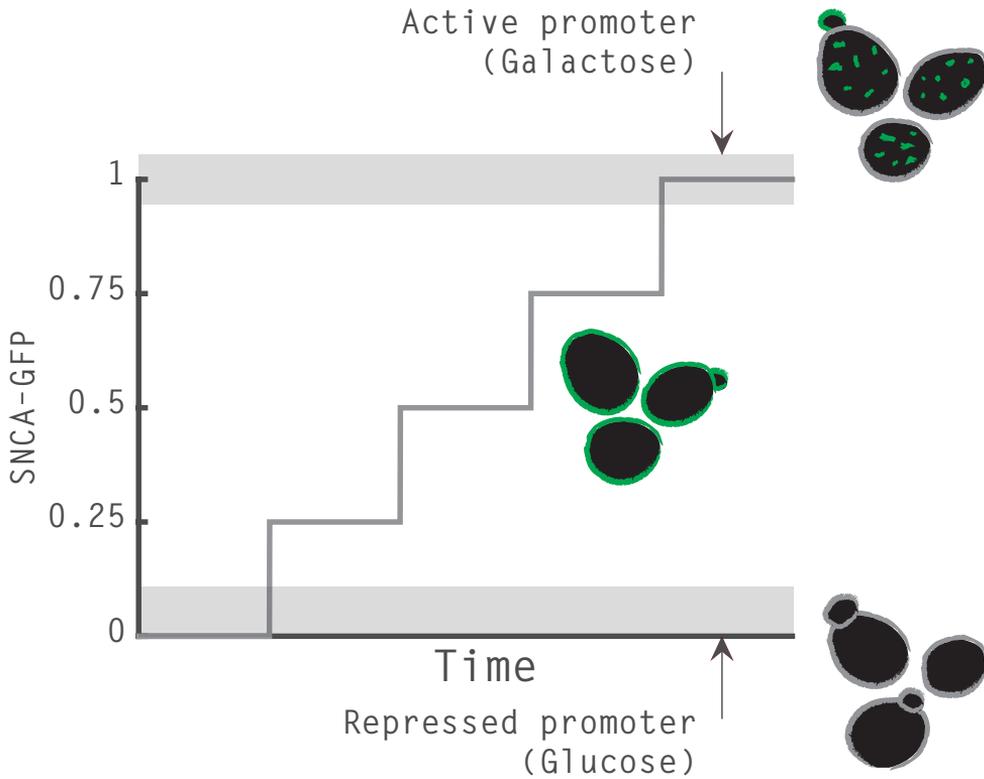


Figure 6.3: **Proof-of-concept to investigate α -synuclein aggregation dynamics.** So far, the α -synuclein accumulation was studied only in a qualitative manner, since the expression could be activated or repressed by means of an inducible promoter without the possibility to maintain intermediate concentration levels. External feedback control of gene expression from the galactose-inducible promoter makes it possible to quantitatively study the dynamics of α -synuclein expression and aggregation.

6.2.1 External feedback controller of α -synuclein expression

The controller is based on the model predictive control strategy presented in Section 4.3 . The implementation details of the control algorithm are described in Section 4.3.2, except for the differences presented in the following paragraphs.

Since the system dynamics of the α -synuclein protein expression is slower compared to that of the experimental testbed used in Chapter 4, the length of the prediction horizon (T_p) was doubled, and set to 120 min. In this way, the number of sampling intervals to be considered in Eq. 4.12 is equals to the integer number $N = 24$.

Moreover, the optimisation was performed by adopting the Sequential Quadratic Programming (SQP) method described in Chapter 14 of Nocedal and Wright [62]. The SQP method is a constrained nonlinear optimization algorithm and it is implemented as a solver method in the *fmincon* function of the *Optimization Toolbox* (The MathWorks, Inc.).

6.2.2 Identification and modeling of α -synuclein protein expression

The model predictive controller needs a dynamical model of the system being controlled to compute the control input (u). Thus, I derived a state-space linear model of the single copy SNCA_{A53T}-GFP yeast strain. To this end, a system identification time-lapse experiment was performed, as depicted in Fig. 6.4. Before the experiment, yeast cells were kept in galactose enriched medium in order to fully activate the galactose promoter (galactose steady state). The identification time-lapse experiment was carried out for 40 hrs, feeding yeast cells alternatively with galactose and glucose enriched media for 480 min (Fig. 6.4). The average fluorescence of the cell population was quantified at each sampling time and taken as the system output (Fig. 6.4, orange line).

I decided to describe the dynamics of the α -synuclein expression by

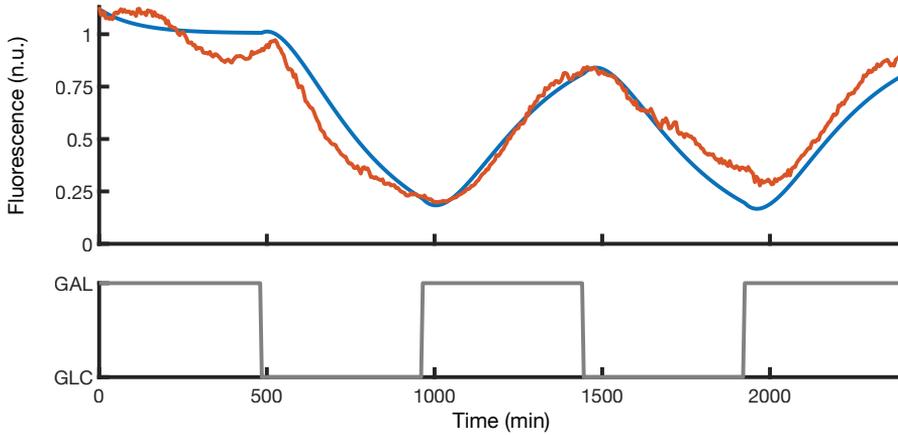


Figure 6.4: **Time-lapse experiment for the identification of α -synuclein protein expression.** The grey line is the input u provided to the yeast cells. The orange line is the output y measured as the average fluorescence (GFP) of the cell population. The blue line is the simulated output to the same stimuli as the one provided for the identification experiment, using the identified discrete model.

means of a single-input single-output (SISO) linear discrete time system consisting of two coupled difference equations:

$$\begin{aligned} x_{k+1} &= A x_k + B u_k \\ y_k &= C x_k \end{aligned} \tag{6.1}$$

where $x_k \in \mathbb{R}^2$ is system state, $u_k \in [0, 1]$ is the input, and $y_k \in \mathbb{R}$ is the measured output. I assumed that the input (u) is piece-wise constant during the sampling period T (zero-order hold method as described in [34]).

The model parameters were estimated from the input-output data shown in Fig. 6.4 (grey and orange lines). Thus, the state-space identification was carried out by employing the *ssest* function of the *System Identification Toolbox* (The MathWorks, Inc.). The algorithm used the prediction error minimization (PEM) approach to estimate the numerical model parameters from the identification input-output data (for further details see Chapter 7

in [48]). The estimated model parameters are:

$$A = \begin{bmatrix} 0.9623 & 0.0095 \\ 0.0067 & 0.9709 \end{bmatrix} \quad (6.2)$$

$$B = \begin{bmatrix} -0.0001461 \\ 0.006252 \end{bmatrix} \quad (6.3)$$

$$C = \begin{bmatrix} 9.569 & -0.0424 \end{bmatrix} \quad (6.4)$$

The identified model was able to recapitulate the experimental data across the identification scenario, as depicted in Fig. 6.4 (blue line); obtaining a fitting percentage (FIT% – see Appendix A for further details about the formula) equal to 95.81%.

6.2.3 Tracking control experiments

As I would like to increase the protein expression at discrete steps, I chose two reference signals:

1. a descending staircase function where each step lasts 750 min, beginning at 75% of the maximum fluorescence when cells are grown in galactose, then stepping down to 50% and then 25%;
2. an ascending staircase function where the first and the second step last 750 min, and the third 500 min, beginning at 25% of the maximum fluorescence value, then stepping up to 50% and then 75%.

Numerical simulations.

I tested the feasibility of achieving the control tasks by simulating the behavior of cell population by means of the identified population-averaged model illustrated in Section 6.2.2. Numerical simulations of the control experiments confirmed the ability of the MPC controller to follow the desired time-varying reference signals (Fig. 6.5).

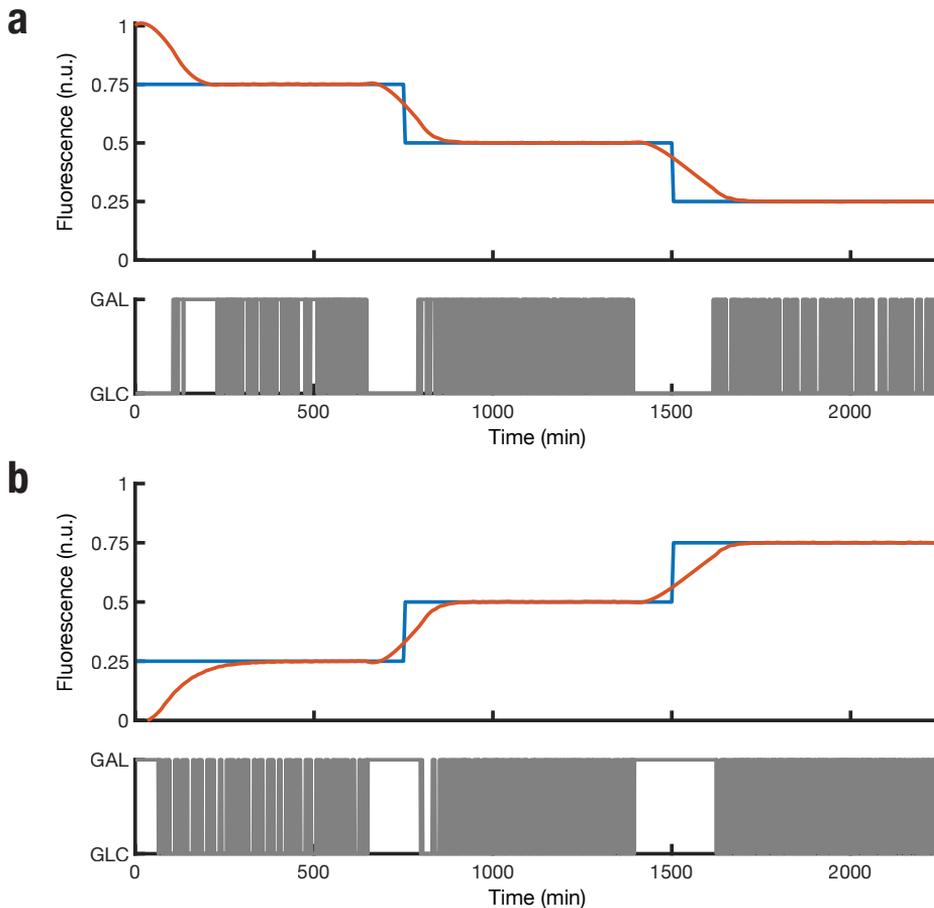


Figure 6.5: **Numerical simulation of tracking control tasks using the MPC strategy.** The blue line is the reference signal. The orange line is the simulated value of the green fluorescent reporter. The grey line is the administered control input. **(a)** Control experiment simulated on the mathematical model of the strain carrying human mutant A53T α -synuclein construct in single copy with a descending staircase reference signal. The initial level of fluorescence is assumed to be equal to 1. The control action starts at $t = 0$ min and ends at $t = 2250$ min. **(b)** Control experiment simulated on the mathematical model of the strain carrying human mutant A53T α -synuclein construct in single copy with an ascending staircase reference signal. The initial level of fluorescence is assumed to be equal to 0. The control action starts at $t = 0$ min and ends at $t = 2250$ min.

In vivo time-lapse control experiments.

Numerical simulations presented in the previous section confirmed that the MPC strategy could achieve a satisfactory control performance. Therefore, I performed the control experiments *in vivo*, whose results are shown in Fig. 6.6. Prior to each control experiment, cells were inoculated in proper enriched medium to induce, or repress, the expression of the α -synuclein construct, depending on the reference signal (descending or ascending staircase). In the case of the descending staircase, cells were inoculated in galactose/raffinose enriched medium, and then the culture was repeatedly diluted to achieve a desired concentration on the day the cells were injected into the microfluidics device. A calibration phase lasting 180 min was performed at the beginning of each experiment to normalize the measured fluorescence value during the control experiment (Fig. 6.6a). Conversely, in the case of ascending reference signal, cells were inoculated in raffinose enriched medium, and then repeatedly diluted to achieve the desired concentration. The raffinose enriched medium allows a faster induction of the galactose-inducible promoter when the cells are fed with galactose. The calibration phase in this case lasts 15 min and the mean of the fluorescence emitted by a constitutively expressed mCherry fluorescent reporter was used to estimate the maximal fluorescence value associated to the full induction of the galactose promoter (Fig. 6.6b).

The experimental results confirmed the numerical simulations, demonstrating the ability of the methodology to study quantitatively the dynamics of α -synuclein protein aggregation.

6.3 Experimental investigation of human α -synuclein protein aggregation in yeast

The experimental results of the pilot study presented up to now firmly demonstrate that the α -synuclein expression can be controlled at intermediate concentration levels by applying principles drawn from control theory. I

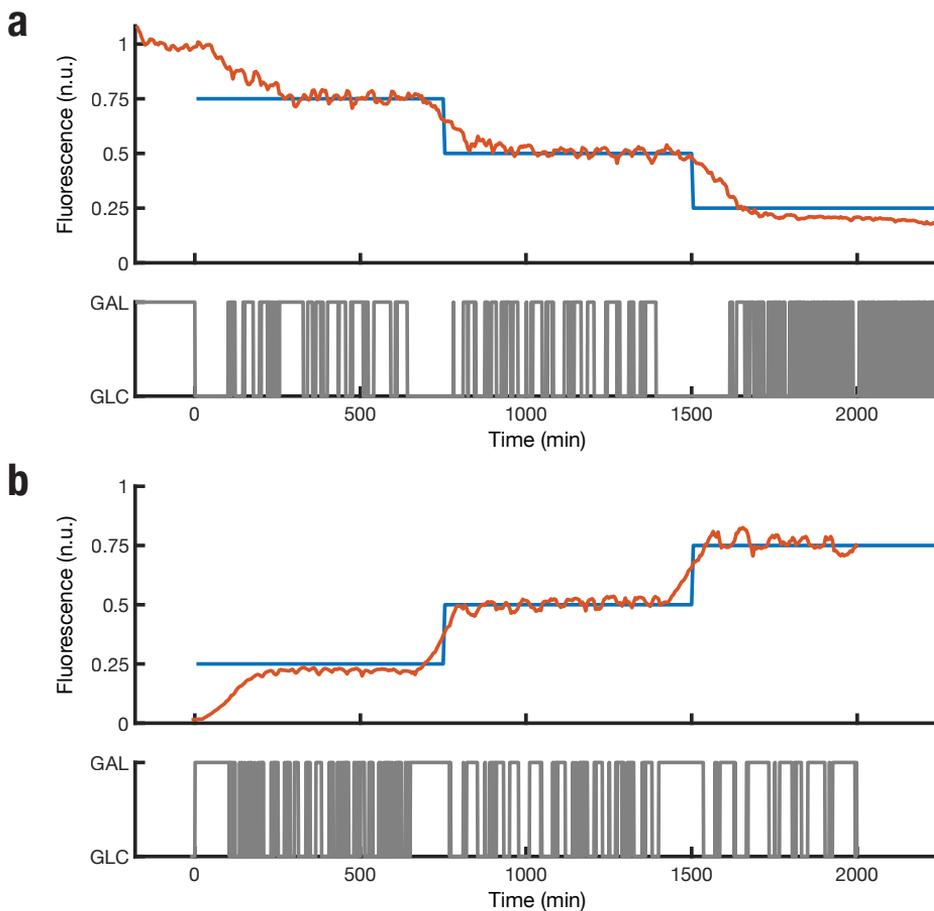


Figure 6.6: Experiment of tracking control tasks using the MPC strategy in yeast cells. The blue line is the reference signal. The orange line is the measured value of the green fluorescent reporter. The grey line is the administered control input. **(a)** Control experiment performed on the strain carrying human mutant A53T α -synuclein construct in single copy with a descending staircase reference signal. The calibration phase lasts 180 min, and the mean value of the measured green fluorescence across the cell population is set equal to 1. The control action starts at $t = 0$ min and ends at $t = 2250$ min. **(b)** Control experiment performed on the strain carrying human mutant A53T α -synuclein construct in single copy with an ascending staircase reference signal. The calibration phase lasts 15 min, and the mean value of the measured green fluorescence across the cell population is set equal to 0, whereas the high steady state is set to a fluorescence value proportional to the mean of measured mCherry fluorescent protein across the cell population. The control action starts at $t = 0$ min and ends at $t = 2000$ min.

then investigated the dynamics of α -synuclein aggregation by applying the same methodology described above.

As stated in Section 6.1, α -synuclein aggregation can be induced in yeast strains only by inserting multiple copies of the SNCA_{A53T}-GFP construct under the GAL1 galactose-inducible promoter. Hence, we generated such a yeast strain as described in Appendix A.

6.3.1 Open-loop dynamics of α -synuclein protein expression in the multiple copy SNCA_{A53T}-GFP yeast strain

In order to confirm α -synuclein protein aggregation in the multiple-copies yeast strain, I first performed experiments in open-loop. Specifically, I performed an *in vivo* time-lapse experiment by simply switching cells from raffinose enriched medium to galactose enriched medium in order to induce the expression of the α -synuclein protein. The complete experimental procedure is reported in Appendix A.

The experimental results of the *in vivo* open-loop experiment are shown in Fig. 6.7. The mean value of the fluorescence measured across the cell population shows that protein expression is induced as expected following growth medium switch (Fig. 6.7a, orange line). Moreover, the α -synuclein protein is expressed at high levels and thus it aggregates during the course of the experiment, as shown in Fig. 6.7.

6.3.2 Offset-free Model Predictive Control of α -synuclein protein expression in the multiple copy SNCA_{A53T}-GFP yeast strain

In vivo control experiments in the pilot study were carried out by means of a model predictive control strategy. Experimental results demonstrate the feasibility of using the control strategy to steer *in vivo* SNCA gene expression. However, they also highlighted the presence of a steady state error

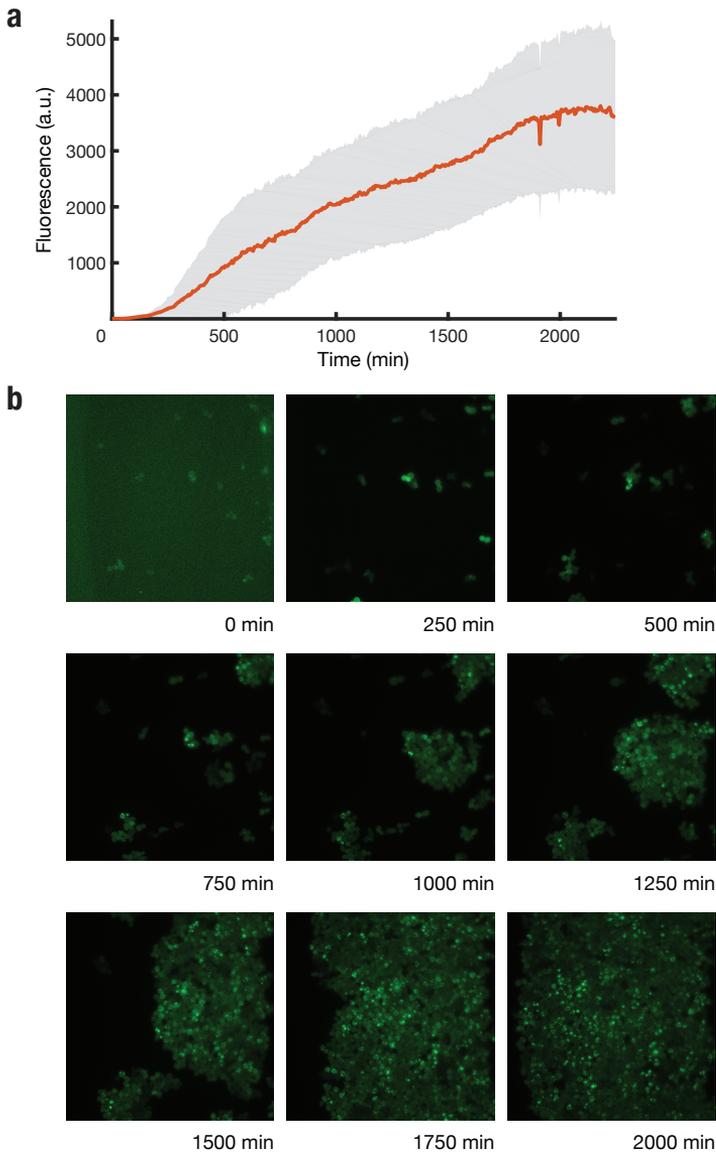


Figure 6.7: **Open-loop dynamics of human mutant A53T α -synuclein protein expression from the galactose-inducible promoter in the multiple copy SNCA_{A53T}-GFP yeast strain.** Yeast cells were grown in raffinose enriched medium, and then switched to galactose enriched medium at 15 min. (a) The orange line is the mean fluorescence level (GFP) measured across the cell population. The grey area is the standard deviation of the expression levels measured across the cell population. (b) Representative microscopy images acquired from the green channel showing the expression of the α -GFP during the time-lapse experiment.

when the fluorescence has to track the lowest value of the reference signal in Fig. 6.6. Since the numerical simulations did not show any steady state error, I deduced that a model mismatch influenced the tracking performance of the MPC.

For this reason, I redesigned the MPC following the approach first proposed by Pannocchia and Rawlings in order to reduce the steady-state error [63]. In this approach, the model predictive control algorithm achieves offset-free control by augmenting the system model with an additional state representing a constant disturbance. Thus, the state of the original system is translated onto the manifold that suppresses the effects of the disturbance on the controlled variables.

The state-space formulation of the augmented model is:

$$\begin{aligned}x_{k+1} &= A x_k + B u_k \\d_{k+1} &= d_k \\y_k &= C x_k + d_k\end{aligned}\tag{6.5}$$

where $x_k \in \mathbb{R}^2$ is system state, $d_k \in \mathbb{R}$ is the integrated disturbance on the controlled variable (y), $u_k \in [0, 1]$ is the input, and $y_k \in \mathbb{R}$ is the measured output. Considering the integrating disturbance as a state variable, I can recast the disturbance model in the following augmented system

$$\begin{aligned}\begin{bmatrix} x_{k+1} \\ d_{k+1} \end{bmatrix} &= \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} x_k \\ d_k \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u_k \\y_k &= \begin{bmatrix} C & I \end{bmatrix} \begin{bmatrix} x_k \\ d_k \end{bmatrix}\end{aligned}\tag{6.6}$$

The state of the system, and thus also the additional disturbance are numerically estimated from the measured fluorescence data by employing a Kalman filter designed for the augmented system as depicted in Eq. 6.6,

according to the procedure described in Pannocchia and Rawlings [63]. The estimation of the disturbance (d) achieves an offset-free steady state in the model predictive control strategy, as described in 6.2.1.

As next step in analysing the aggregation dynamics of α -synuclein protein, I performed a series of time-lapse control experiments to understand the aggregation point of the α -synuclein protein in the cell population. The time-lapse experiments were carried out *in vivo* by means of a offset-free model predictive control strategy, as exposed in Section 6.3.2.

As a control task, I decided to track an ascending staircase signal so that α -synuclein protein expression is steered from no expression to high expression in discrete steps (e.g. Fig. 6.5b and Fig. 6.6b). This choice allows to precisely quantify at which expression level aggregation occurs and also to observe and model the protein aggregation dynamics.

On setting the reference signal for α -synuclein protein expression from the multiple copy SNCA_{A53T}-GFP yeast strain.

Fluorescence intensity is proportional to α -synuclein protein expression, however the scaling constant is not known and prone to vary in each experiment due to biological (gene expression noise) and technical constraints (e.g. wearing of the fluorescence lamp). Moreover, it is not possible to fully induce the expression of α -synuclein in the multiple copy SNCA_{A53T}-GFP yeast strain, because the mutant α -synuclein protein will aggregate, thus distorting fluorescence readings, therefore I cannot directly measure the maximum fluorescence level of α -synuclein-GFP ($GFP_{m.f.l.}$) expressed by the cells. The $GFP_{m.f.l.}$ value is needed to set the unit of measure of the reference signal (expressed as a % of the maximum fluorescence).

To solve this issue, I hypothesised that the α -synuclein's $GFP_{m.f.l.}$ prior to aggregation is proportional to the fluorescence level emitted by the mCherry protein expressed in the yeast strain under a constitutive promoter (as detailed in Appendix A).

In order to estimate the $GFP_{m.f.l.}$ prior to aggregation, I analysed the open-loop experiment depicted in Fig. 6.7. I observed that the protein ag-

gregation starts between 250 min and 500 min. This observation is confirmed by the fluorescence images for the green fluorescence channel (e.g. Fig. 6.7b). I then performed the following steps:

1. I filtered the fluorescence time-series in the time interval between 250 min and 500 min using polynomial interpolation. To this aim, I employed a robust implementation of the *LOESS* method, a non-parametric regression method available in the *smooth* function of the *Curve Fitting Toolbox* (The MathWorks, Inc.) [64, 65]. In detail, the *smooth* function implements local regression using weighted linear least squares and a second degree polynomial model. Moreover, the robust implementation of the *LOESS* method can be exploited to reduce the method sensitivity to outliers. I considered a span of 0.25% for the smoothing of the fluorescence levels.
2. I considered as the time in which aggregation begins, the time point in the interval between 250 min and 500 min at which the fluorescence level changes in slope. Therefore, I computed numerically the second derivative of the smoothed time-course data, and found that its maximum occurs at 270 min and that at this time point the green fluorescence level (i.e. α -synuclein) is approximately 665 (a.u.). This value is set as the $GFP_{m.f.l.}$ prior to aggregation.
3. Since fluorescence values are not comparable across experiments because of biological and technical issues, the $GFP_{m.f.l.}$ I estimated in the previous step needs to be normalised to the red (mCherry) fluorescence signal in the cell population, which can instead be measured in each experiment. To this end, I computed a scaling factor as follows:

$$SF = \frac{GFP_{m.f.l.}}{mCherry_{glucose}} \quad (6.7)$$

where $mCherry_{glucose}$ is the red fluorescence value measured across cell population during the initial calibration phase of the open-loop experiment during which cells were kept in glucose.

The $mCherry_{glucose}$ is computed in each control experiment by measuring it during the calibration phase. Assuming that the value of SF is constant across experiment, the value of $GFP_{m.f.l.}$ can be estimated simply as:

$$GFP_{m.f.l.} = SF \times mCherry_{glucose}. \quad (6.8)$$

Understanding the aggregation dynamics of human α -synuclein protein.

Figure 6.8 shows the results of the reference tracking experiment in the multiple copy SNCA_{A53T}-GFP yeast using an ascending staircase as reference signal. The control objective was achieved (Fig. 6.8a, upper panel), however the protein did not aggregate during the experiment (Fig. 6.8b).

Therefore, I performed an additional control tracking experiment setting the maximum value of the reference signal to 4 (i.e. four times the estimated $GFP_{m.f.l.}$) as shown in Fig. 6.9. Specifically, I decided to divide the second time-lapse control experiment in two parts:

1. a closed-loop control phase lasting 2000 min;
2. an open-loop phase lasting 1000 min in which I just provided galactose enriched medium to the cells in order to maximally overexpress the protein and check if it aggregates.

The experimental results are shown in Fig. 6.9. Also in this case, the protein did not aggregate during the closed-loop control phase (Fig. 6.9b), whereas it aggregated during the open-loop phase.

The third time-lapse experiment in chronological order is equal to the second one in terms of control objectives, although this time I set the reference signal to reach a maximum value of 16 (i.e. sixteen times the estimated $GFP_{m.f.l.}$ in Eq. 6.8). The experimental results of the *in vivo* experiment are shown in Fig. 6.10. This time the protein aggregates during the controlled expression phase of the experiment (e.g. in Fig. 6.10b). I thus decided to stop the experiment, since the aggregation point was reached at the second step of the reference signal.

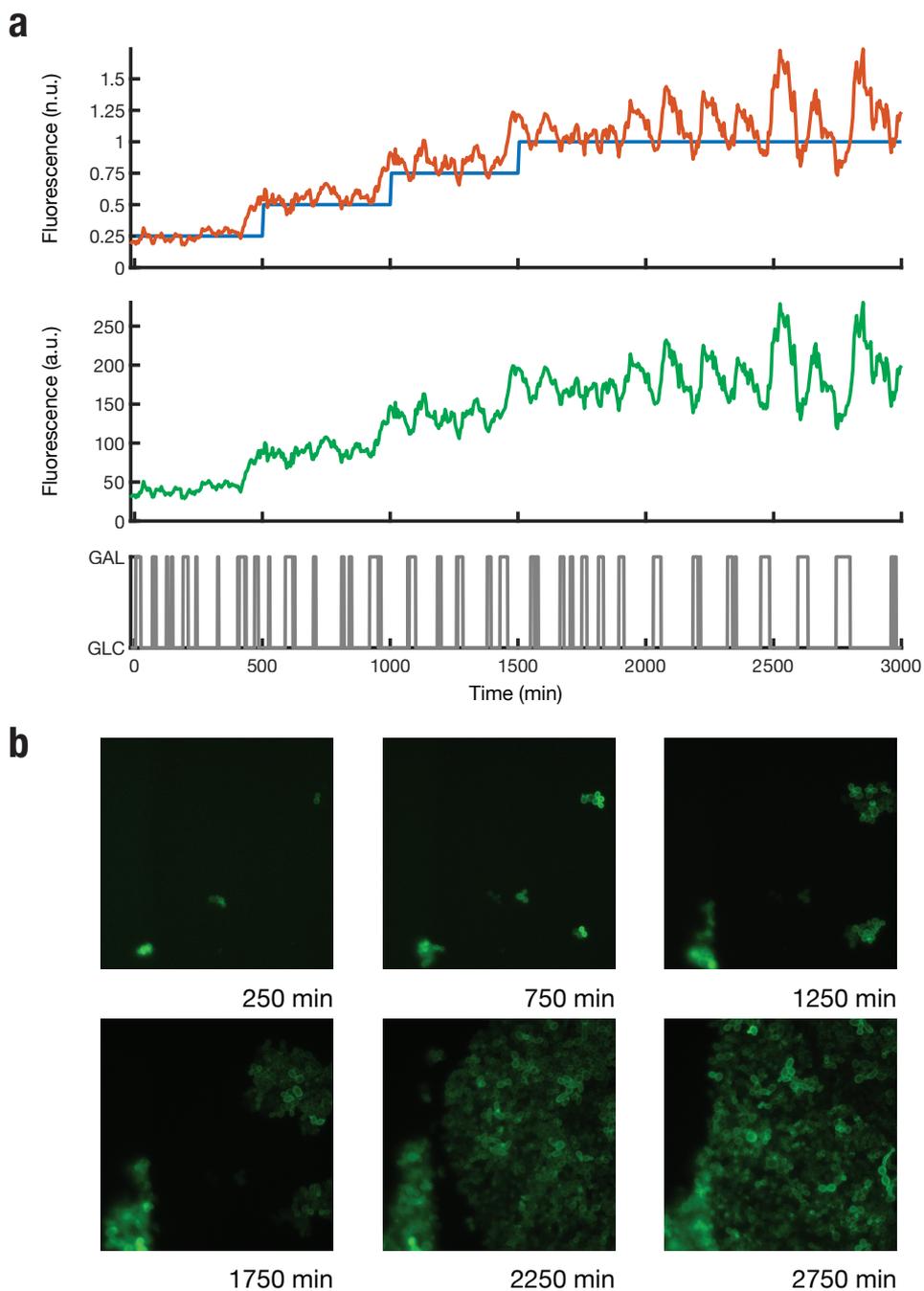


Figure 6.8: **Signal tracking control experiment of human mutant A53T α -synuclein expression.** The value of 1 in the reference signal is the estimated $GFP_{m.f.l.}$ computed as in Eq. 6.8.

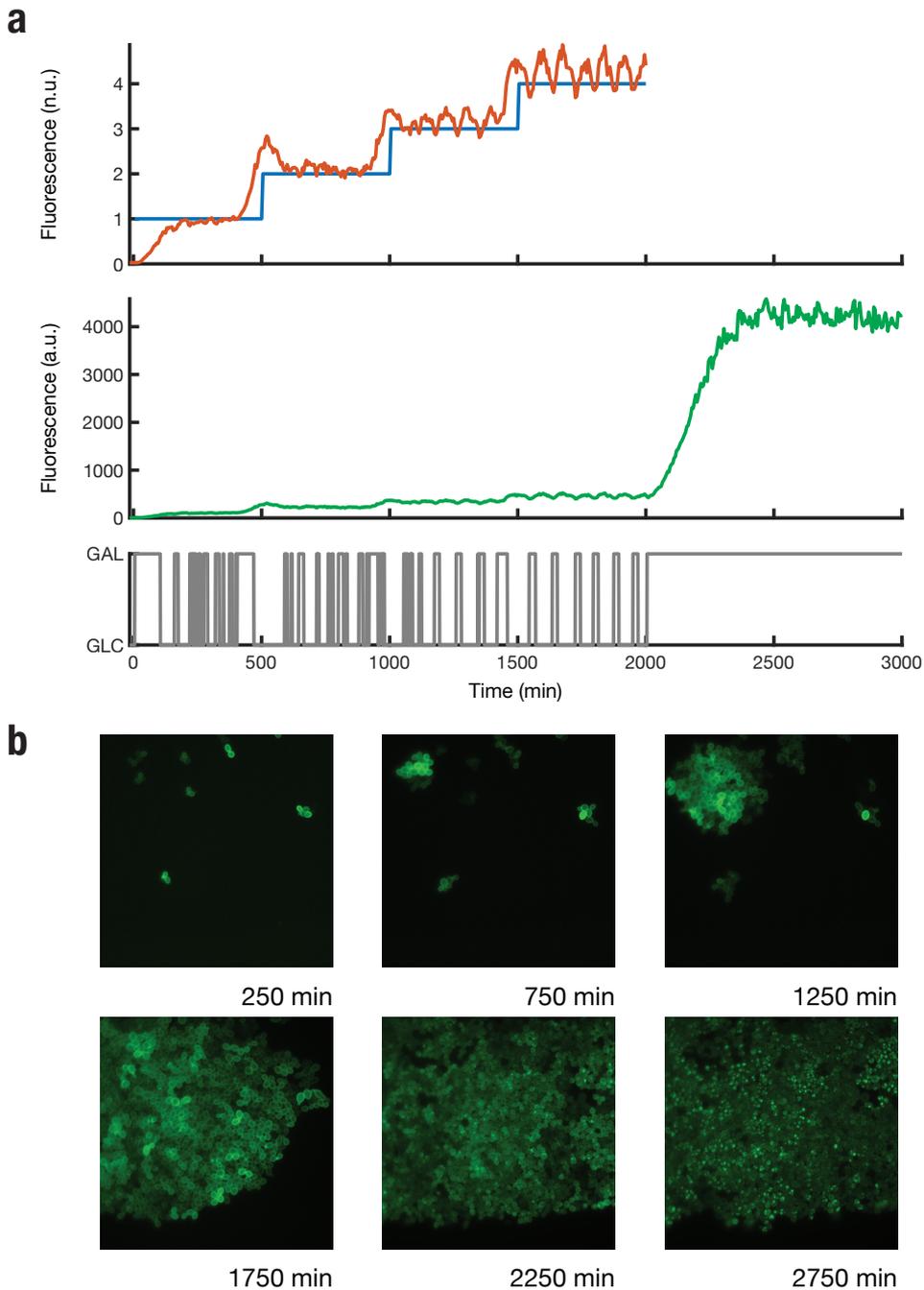


Figure 6.9: **Signal tracking control experiment of the human mutant A53T α -synuclein.** Time-lapse signal tracking control experiment where the maximum level of the reference signal is fourfold the $GFP_{m.f.l.}$ computed in Eq. 6.8.

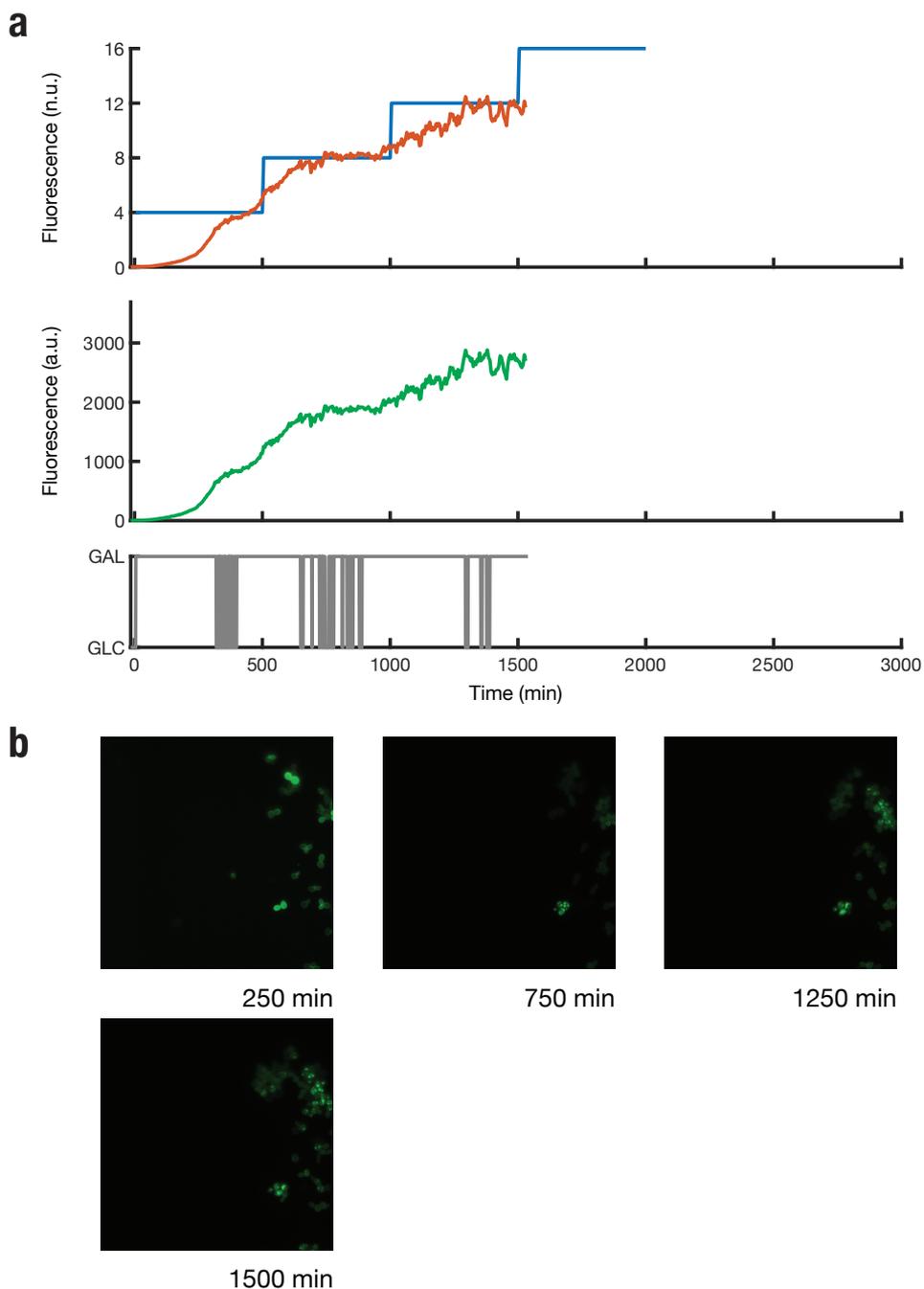


Figure 6.10: **Signal tracking control experiment of the human mutant A53T α -synuclein.** Time-lapse signal tracking control experiment where the maximum level of the reference signal is sixteen-fold the $GFP_{m.f.l.}$ computed in Eq. 6.8.

In the final time-lapse experiment, I controlled the expression of the α -synuclein protein at intermediate concentration level between the first and the second plateaus of Fig. 6.10. Therefore, I set the reference signal to reach a maximum value of 10 (i.e. ten times the estimated $GFP_{m.f.l.}$ in Eq. 6.8). The experimental results are depicted in Fig.6.11. As it can be appreciated from the Fig. 6.11, the α -synuclein protein shows a threshold effect (highlighted in Fig. 6.12 and Fig. 6.13; grey area), since it starts aggregating within a fixed range of fluorescence levels.

Concluding, the entire analysis is recapitulated in Fig. 6.12, which highlights the threshold effect (Fig. 6.12 and Fig. 6.13; grey area).

6.4 Discussion

In this Chapter, I presented a possible application scenario for the experimental platform described in Chapter 3. I exploited the potential of external feedback control of gene expression to study in yeast a pathological hallmark of Parkinson's disease, i.e. the aggregation of the human α -synuclein protein.

So far, a quantitative understanding of the dynamics involving the accumulation of the α -synuclein protein is lacking. Here, I performed a quantitative analysis of the aggregation dynamics of the mutant form A53T of the human α -synuclein protein. I have discovered that the accumulation of the mutant form is characterised by a threshold level. Below this level, the protein does not aggregate, whereas above it, the protein accumulates in cells, and the aggregated corps are visible as cytoplasmic inclusions.

The work presented in this Chapter is part of an ongoing work, further analysis are required to completely dissect the accumulation dynamics of the α -synuclein protein. For example, it will be interesting to study the aggregation properties exhibited at the single cell level. Moreover, a comparison of the aggregation dynamics among yeast strains expressing wild type and mutant forms of human α -synuclein protein can elucidate further details about the pathological hallmark, since the mechanisms underlying the protein aggregation are still unclear.

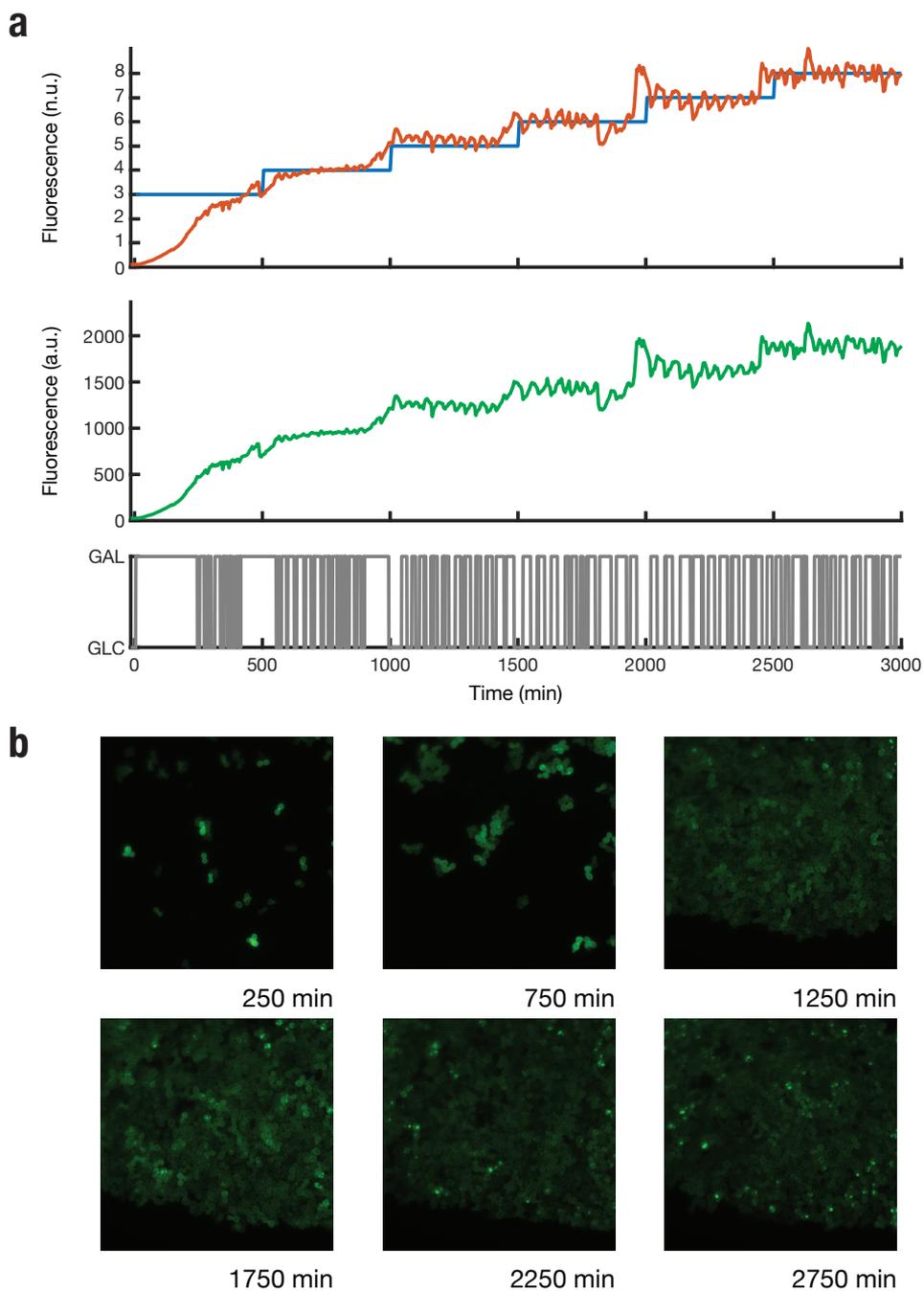


Figure 6.11: **Signal tracking control experiment of the human mutant A53T α -synuclein.** Time-lapse signal tracking control experiment where the maximum level of the reference signal is ten-fold the $GFP_{m.f.l.}$ computed in Eq. 6.8.

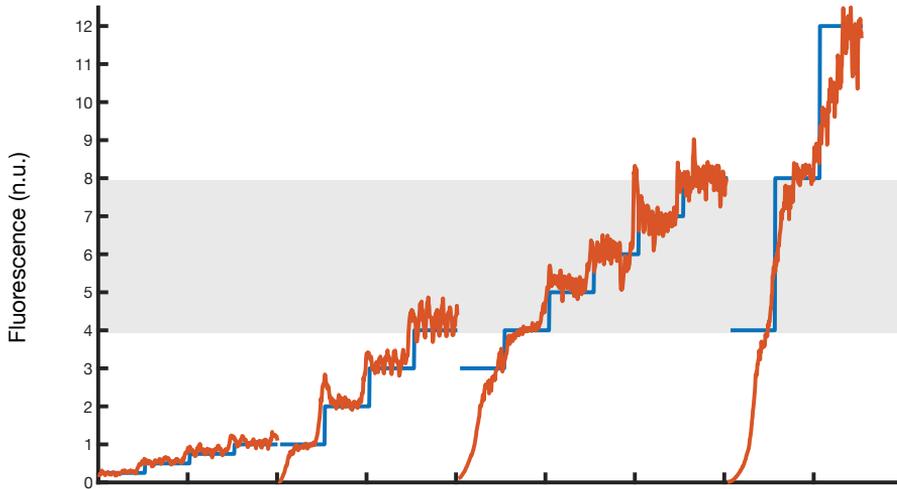


Figure 6.12: **Overview of signal tracking control experiments.** The expression levels (orange lines) of the four time-lapse control experiments, reported in normalized units, are shown together in order to highlight the threshold effect (grey area) exhibited by the protein expression.

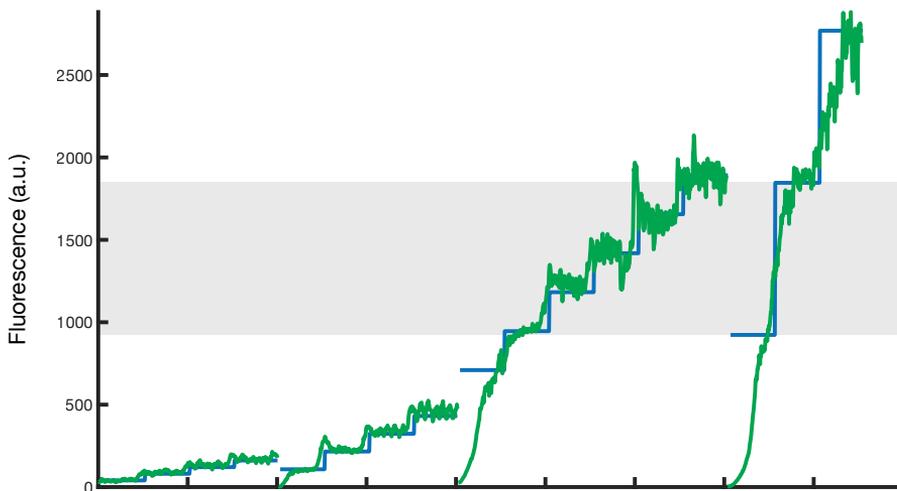


Figure 6.13: **Overview of signal tracking control experiments.** The expression levels (green lines) of the four time-lapse control experiments, reported in arbitrary units, are shown together in order to highlight the threshold effect (grey area) exhibited by the protein expression.

Chapter 7

Conclusions

In this Thesis, I contributed to the emerging role of control theory in synthetic biology. Exploiting the concept of negative feedback loop, I extended the field of controlling cellular processes in real-time to biomedically relevant applications. Indeed, I devised novel approaches to model and control gene expression dynamics in yeast from the endogenous galactose-inducible *GAL1* promoter. I performed most of the experiments here described by means of the experimental platform presented in Chapter 3.

The first contribution of this Thesis is presented in Chapter 4, where I provide a comparative analysis of different control strategies that can be used to control gene expression in living cells. So far, a similar comparison has never performed in literature, thus representing an important contribution to the field of control theory in synthetic biology. I carried out the analysis by comparing three control strategies: proportional-integral control, model-predictive-control, and zero-averaged-dynamics control. I demonstrated that both MPC and ZAD control strategies can be successfully employed to control gene expression from the galactose-inducible promoter to generate any desired time-varying concentration of the reporter protein (GFP). Instead, the PI controller performed similarly to the MPC and ZAD strategies only in the setpoint control task, whereas it was the worst performer in the case of signal-tracking experiments.

The second achievement of this Thesis is proposed in Chapter 5, that

deals with the characterisation of gene expression from endogenous galactose-inducible promoter both at single cell and at population level. I verified that a mixed-effects dynamical model can correctly describe the variability in fluorescence level both in individual cells and at the population level. To this aim, I performed an analysis on single cell data obtained from the time-lapse control experiments presented in Chapter 4. The main aim of the analysis was to characterise both the intrinsic and extrinsic noise arising from the gene expression. I then derived a mathematical representation of protein expression neglecting intrinsic noise using the mixed-effects modeling framework. The third contribution is a novel model-predictive-control approach, based on single cell models in order to predict the overall behaviour of the cell population (Chapter 5). Numerical simulations confirmed that the proposed strategy may successfully achieve the regulation of gene expression from the galactose-inducible promoter in yeast cells.

Finally, the fourth contribution of this Thesis is described in Chapter 6, where I presented a biomedically relevant application of external feedback control of gene expression from the endogenous *GAL1* promoter. In particular, I dissected in yeast a pathological hallmark of Parkinson's disease, i.e. the aggregation of the human α -synuclein protein.

Appendix A

Materials and methods

A.1 Yeast strains

All yeast strains used to carry out this study are here described.

The comparative analysis of the control strategies presented in Chapter 4 has been accomplished through the GAL1–GFP yeast strain (Gal1–GFP–KanMX, Gal10–mCherry–NatMX) kindly provided us by Prof. David Botstein [20]. In this strain the Gal1 protein, expressed by the galactose–inducible promoter, is fused to a green fluorescent protein (GFP).

The aggregation dynamics of the α –synuclein protein has been elucidated in Chapter 6 by means of two different yeast strains:

1. single copy SNCA_{A53T}–GFP yeast strain (Gal1pr–SNCA_{A53T}–GFP–KanMX);
2. multiple copy SNCA_{A53T}–GFP yeast strain (Gal1pr–SNCA_{A53T}–GFP–KanMX, Gal1pr–SNCA_{A53T}–GFP–HphMX, YC_{plac111} Gal1pr–SNCA_{A53T}–GFP–LEU2).

The yeast strains have been constructed in collaboration with Dr. Cathal Wilson and Dr. Marco Santorelli. To this aim, we have inserted into the strains the Gal1pr–SNCA_{A53T}–GFP construct in single and multiple copy. The construct has been taken from the the plasmid (pRS304–Syn–A53T) kindly provided us by Prof. Susan Lindquist [61]. Moreover, both yeast

strains have been manipulated to express the mCherry fluorescent protein from a constitutive promoter.

A.1.1 Single copy SNCA_{A53T}–GFP strain construction

1. α -Synuclein A53T construct fused to a Green Fluorescent Protein (GFP) under control of the galactose-inducible promoter (Gal1pr–SNCA_{A53T}–GFP) was amplified by PCR from the plasmid pRS304–Syn–A53T [61], using the forward primer

O.Gal.Fwd

5' – CAGCTGAAGCTTCGTACGCTGCAGGTCGACAGTACGGA
TTAGAAGCCGCC – 3'

and reverse primer

O.Cyc.Rev

5' – GGCGGGGACGAGGCAAGCTAAACAGATCTCAAATTA
GCCTTCGAGCGTCC – 3'

and cloned *SalI/BglII* into the vector pYM27 [66].

2. After sequencing, a cassette containing α -synuclein sequence together with the Kan resistance gene was amplified from the respective pYM27 construct using primers with a 50 bp overhang corresponding to the 5' and 3' sequences of the yeast dubious ORF YMR082C using the primers

Gal.Syn.Fwd

5' – TGATTATCTAAGCAGCAATCCCCTTGTCTACAAAACA
GAAACTGGAAGAAGTACGGATTAGAAGCCGCGAG – 3'

and

S2.Syn.Rev

5' – ACGCAGACCCATTCGAGGGGCTCATTGGAAACACGTAG
TCGACATTAGTTATCGATGAATTCGAGCTCGTT – 3'.

3. A yeast strain with a constitutively expressed cytosolic marker (TEF2pr–mCherry; [67]) was transformed with the amplicon and transformant was selected on kanamycin-containing plates.

-
4. Insertion of the α -synuclein cassette into the YMR082C locus by homologous recombination was verified by PCR from genomic DNA prepared from the strain.

A.1.2 Multiple copy SNCA_{A53T}-GFP strain construction

1. To insert a second copy of the construct into the genome, the Gal1pr-SNCA_{A53T}-GFP cassette was cloned *SalI*/*BglII* into the vector pYM25, as described above.

2. After sequencing, the cassette containing the α -synuclein sequence together with the Hph resistance gene was amplified from the pYM25 construct using primers with a 40 bp overhang corresponding to the 5' and 3' sequences of the yeast dubious ORF YFR054C using the primers

YFR054C-KO-pYM-F

5' – TAATGCCGAAGTATTACGTACTACGAAAGTTAAGACTA
TGCGTACGCTGCAGGTCGAC – 3'

and

YFR054C-KO-pYM-R

5' – TAGAGTGTAATCTCGCCAAACCCAGTAATAGCATCGTT
TAATCGATGAATTCGAGCTCG – 3'.

3. The yeast strain expressing a single copy of SNCA_{A53T}-GFP-KanMX and TEF2pr-mCherry was transformed with the amplicon, and transformants were selected on kanamycin/hygromycin-containing plates.
4. Insertion of the α -synuclein cassette into the YFR054C locus by homologous recombination was verified by PCR from genomic DNA prepared from the transformants.

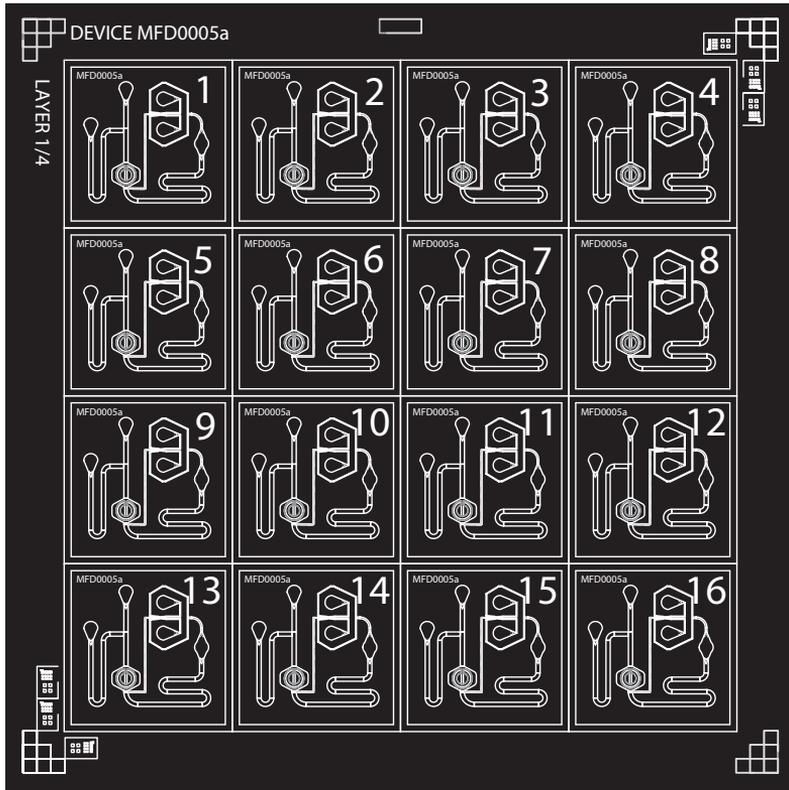


Figure A.1: **Blueprint of device MFD0005a.** Adapted from [25].

A.2 Microfluidics

A.2.1 Master–mold

The microfluidic device MFD0005a has been conceived and designed by the *Biodynamics Laboratory* of Prof. Jeff Hasty (UCSD) [25]. The device fabrication has been carried out with a replica molding technique by means of the master–mold kindly provided us by Prof. Jeff Hasty as a blueprint (Fig.A.1).

A.2.2 Device fabrication

The devices are fabricated with polydimethylsiloxane (PDMS; Sylgard 184, Dow Corning).

1. Before the fabrication, the master-mold is exposed to chlorotrimethylsilane (Sigma-Aldrich Co.) for 10 *min* in order to create an anti-slicking layer for PDMS.
2. PDMS with a 10 : 1 ratio (base to curing agent, w/w) is poured on top of the master-mold and cured at 80°C for 2 hrs.
3. Successively, the PDMS layer is cut and peeled from the master-mold.
4. The inlet ports of the devices are pierced with a micro-puncher (0.5 mm; World Precision Instruments, Inc.).
5. Afterwards, the PDMS devices are washed overnight in isopropyl alcohol (Sigma-Aldrich Co.). At the same time, thin glass slides (Thickness No. 1.5; Paul Marienfeld GmbH & Co. KG) are cleaned in acetone (Sigma-Aldrich Co.) and isopropyl alcohol (Sigma-Aldrich Co.). Both PDMS devices and cleaned glass slides are left overnight to dry under vacuum.
6. A PDMS device and a glass slide are undergone to plasma treatment for 1 min in a plasma cleaner machine (ZEPTO version B; Diener electronic GmbH + Co. KG), and then are bonded together irreversibly.
7. Finally, an inspection is performed on the fabricated devices to check for possible defects.

A.3 Microscopy

The experimental platform presented in Chapter 3 exploits an inverted fluorescence Nikon Eclipse Ti-E microscope (Nikon Instruments) to acquire raw images of cells into microfluidic device. The microscope is equipped with

an iXON Ultra897 EMCCD camera (Andor Technology Ltd) and a Perfect Focus System (PFS; Nikon Instruments). The PFS has been conceived to combat axial focus fluctuations, due to thermal noise, in real-time during long-term imaging acquisitions.

The microscope and the camera were programmed to acquire, at 5 min intervals, three different type of image:

1. a phase contrast image,
2. a fluorescence image in the green spectrum (Piston GFP Bandpass Emission filter; Nikon),
3. a fluorescence image in the red spectrum (TRITC HYQ filter; Nikon).

All the types of image have been acquired with a 40X dry objective (CFI Plan Fluor DLL 40X, Nikon Instruments).

Image acquisition has been performed with the NIS-Elements AR v. 3.22 (Nikon Instruments) control software.

A.4 Image analysis

The image analysis has been carried out in MATLAB R2012b (The MathWorks, Inc.) by employing the custom image processing algorithm developed by Menolascina et al. [10]. In summary, it was conceived and developed to quantify the fluorescence signals being expressed by the entire population of yeast cells.

This aim has been achieved implementing the following pipeline:

1. Raw images (phase contrast and fluorescence images) are acquired by the microscope and fed to the image processing algorithm.
2. The image processing algorithm elaborates the phase contrast image to detect the yeast cell inside the frame.
3. Once yeast cells are detected inside the frame, a binary mask is created to define the regions in which yeast cells have been found.

4. The binary mask is overlapped to raw fluorescence images (logical AND operator) to select only the fluorescence intensity emitted by cells.
5. The average fluorescence intensity is calculated.

Further details about the image processing algorithm can be found in [22, 23, 26].

A.4.1 Custom offline tracking algorithm

The single cell quantification of protein expression presented in Chapter 5 has been achieved by means of a custom offline segmentation and tracking algorithm. To this aim, I adapted and improved the algorithm proposed in [26].

The custom image processing algorithm has been developed and implemented as a function in the MATLAB (The MathWorks, Inc.) environment. The algorithm receives in input the sequence of images in a time-lapse experiment, and returns in output a set of time-series related to each one of the tracked cells, consisting of the single-cell fluorescence levels and mother-daughter relationships among the cells in order to track also the phylogeny of each yeast cell.

Basically, in order to measure the fluorescence level of a single cell over time, a three-step image analysis approach was devised:

1. detection of the cells' position within each image (segmentation);
2. tracking of the movements of each cell in the entire image sequence (tracking);
3. generation of a lineage tree describing the mother-daughter relationships among cells (phylogenetic analysis).

The segmentation step was rearranged from the custom image segmentation algorithm described above. Instead, the tracking step was developed as an improvement of the algorithm presented in [26].

Since yeast cells make small frame-to-frame movements, tracking and cell division detection can be performed by finding the correspondences between the objects detected in two consecutive frames (phase contrast images) spotting a minimum cost configuration. This association cost increases as long as the displacement between the centroids of the corresponding objects. The minimum cost configuration can be determined by setting up and solving a linear programming problem (LPP).

GPU-parallel software implementation in CUDA

The proposed segmentation and tracking algorithm, implemented as a MATLAB function, is not suitable for real-time applications, like the time-lapse control experiments presented in Chapter 4. In order to reduce the computational time requested by the algorithm, an efficient GPU-parallel software implementation was developed in collaboration with Dr. Diego Romano and Dr. Livia Marcellino, exploiting the NVIDIA CUDA (Compute Unified Device Architecture) environment.

The computational time was sped up by transforming the tracking module of the algorithm, that holds the CPU for the 40% of the total time requested by the algorithm to be executed, in a GPU-parallel component according to the procedure described in [68].

A.5 Computational analysis

A.5.1 Model quality metrics

Model quality metrics are computational tools that can be used to assess the quality of identified mathematical models, compare different models, and choose the best one.

The fitting percentage ($\text{FIT}_{\%}$) has been used as index to evaluate the quality of the identified mathematical models. It provides a measure of the percentage of the output variation that is reproduced by the mathematical

model and is given by the following formula:

$$\text{FIT}_{\%} = 100 \left(1 - \frac{\|y_{measured} - y_{model}\|}{\|y_{measured} - \bar{y}_{measured}\|} \right) \quad (\text{A.1})$$

where:

- $y_{measured}$ is the measured output data.
- $\bar{y}_{measured}$ is the mean value of $y_{measured}$.
- y_{model} is the simulated or predicted response.
- $\|\cdot\|$ denotes the 2 – norm operator of a vector.

A.6 Experimental Protocol

A.7 Cell culture

GAL1–GFP strain

1. On day 0, batch culture is inoculated in 10 mL Synthetic Complete medium (SC) + galactose/raffinose (2%).
2. The batch culture is grown overnight at 30°C until it reaches log or mid–log phase.
3. On day 1, the batch culture is then normalized for OD₆₀₀.

Single copy SNCA_{Δ53T}–GFP strain

1. On day 0, batch culture is inoculated in 10 mL Synthetic Complete medium (SC) enriched either with galactose/raffinose (2%) or raffinose (2%), according the control task. The galactose/raffinose enriched medium is used to allow the induction of the galactose–inducible promoter, conversely the raffinose enriched medium is employed to repress the induction.

2. The batch culture is grown overnight at 30°C until it reaches log or mid-log phase.
3. On day 1, the batch culture is then normalized for OD₆₀₀.

Multiple copy SNCA_{A53T}-GFP strain

1. On day 0, batch culture is inoculated in 10 mL Synthetic Complete medium without leucine (SC-L) + raffinose (2%).
2. The batch culture is grown overnight at 30°C until it reaches log or mid-log phase.
3. On day 1, the batch culture is then normalized for OD₆₀₀.

A.7.1 Microfluidic device set up

1. On day 1, 60 mL syringes (Becton, Dickinson and Company, NJ) filled with 10 mL SC + galactose/raffinose (2%) and SC + glucose (2%) media are prepared, as well as sink syringes (filled with 10 mL ddH₂O); capillaries and needles are used to allow connection to the microfluidic device. Sulforhodamine B (Sigma-Aldrich Co.) is added to the syringe filled with galactose-enriched medium.
2. Temperature in the micro-environment surrounding the moving stage of the microscope is allowed to settle at 30°C.
3. Before connecting media and sink syringes, the microfluidic device MFD0005a wetting is carried out as described in [25].
4. After air bubbles are removed, media and water filled 60 mL syringes are attached to the device and correct functioning is checked by inspecting the red-fluorescence emitted by Sulforhodamine B as a result of the automatic height control of syringes. This allowed us to carry out a correct calibration of the actuation strategy before the actual experiment is run.

5. At this point cells are injected in the microfluidic device by pouring the batch culture in a 60 mL syringe similar to the ones used to media and sinks.
6. Once cells are trapped in the defined area (see [25] for details) Perfect Focus System is activated to assist autofocusing during the experiment and the acquisition routine of the microscope software is started to initiate image acquisition.

Calibration phase

The calibration phase is needed both to let cells adapt to the microfluidic environment, and to set the unit of measure of fluorescence, which may vary due to technical and biological variability in each experiment.

GAL1–GFP strain

1. Once cells are loaded in the microfluidic device, they are kept in a galactose enriched growing medium for 180 min by controlling the actuators in order to hold the syringe filled with galactose in a higher position with respect to the one carrying glucose.

Single copy SNCA_{A53T}–GFP strain

1. Once cells are loaded in the microfluidic device, they are kept either in a galactose enriched growing medium for 180 min or in a glucose enriched growing medium for 15 min, according the control task.

Multiple copy SNCA_{A53T}–GFP strain

1. Once cells are loaded in the microfluidic device, they are kept in a glucose enriched growing medium for 15 min.

A.7.2 Feedback control phase

Once cells are loaded in the microfluidic device, the user has to start a custom MATLAB script, that manages the entire experimental platform (controller implementation, actuation, image analysis), and has to set the reference signal for the control experiment.

Setpoint control experiments: the script is built to calculate the setpoint for the control as a percentage (indicated by the user at the beginning of the experiment) of the average of the fluorescence measured by the image processing algorithm during the calibration phase previously described. After this, the implemented script proceeds in executing all the code blocks necessary to reach and maintain the fluorescence reference.

Signal tracking control experiments: the length and the values of the steps of the staircase reference used in signal tracking control experiments is calculated by a custom MATLAB script that manages the entire experimental platform. The script is built to calculate the values of each of the steps as percentages (indicated by the user at the beginning of the experiment in the case of the staircase control reference) of the average of the fluorescence measured by the image processing algorithm during the calibration phase. At the end of the calibration, the implemented script proceeds in executing all the code blocks necessary to reach and maintain the fluorescence reference. The same procedure, with a different calculation for the control reference, applies to the case of signal tracking control with the ramp reference signal and the sinusoidal wave.

Appendix B

Table B.1: Model parameters for the cell ensemble employed to carry out the numerical simulations in Chapter 5.

Cell no.	α	β	γ	δ
1	0.02975	0.13054	0.00186	0.00691
2	0.16523	0.08070	0.00288	0.00139
3	0.00074	0.00106	0.01732	0.02601
4	0.04570	0.03330	0.00928	0.00902
5	0.02227	0.04788	0.00341	0.00918
6	0.00259	0.00157	0.01989	0.01027
7	0.00823	0.02174	0.00372	0.01025
8	0.02299	0.08750	0.00071	0.00284
9	1.66023	26.58133	0.00025	0.00362
10	0.56949	0.72304	0.00258	0.00400
11	0.00245	0.00020	0.03969	0.00772
12	0.80907	0.44750	0.03120	0.01386
13	0.03814	0.33104	0.00035	0.00253
14	0.01344	0.00469	0.00939	0.00636
15	0.03760	0.19493	0.00064	0.00146
16	0.01114	0.01844	0.00127	0.00213
17	0.01240	0.11476	0.00076	0.00357
18	0.10481	0.01236	0.02211	0.00345

(continued on next page)

Table B.1 (continued from previous page)

Cell no.	α	β	γ	δ
19	0.09420	0.11110	0.01884	0.02204
20	0.09522	0.02997	0.00989	0.00540
21	0.03551	2.34582	0.00005	0.00158
22	0.00296	0.01165	0.01162	0.02640
23	0.03773	0.33795	0.00279	0.01621
24	0.12622	0.04884	0.00812	0.00355
25	0.02789	0.02203	0.00152	0.00188
26	0.05741	0.06025	0.00310	0.00421
27	0.03821	0.23726	0.00095	0.00351
28	0.00978	0.00952	0.00868	0.00820
29	0.02155	0.07788	0.00148	0.00232
30	0.00516	0.00048	0.08400	0.01691
31	0.04731	0.04430	0.00810	0.00608
32	0.00320	0.00146	0.00422	0.00192
33	0.00355	0.00061	0.01309	0.00342
34	0.00501	0.01328	0.00146	0.00249
35	0.00030	0.00053	0.00385	0.00761
36	0.09793	0.15651	0.00227	0.00306
37	0.02246	0.00567	0.01895	0.01343
38	0.00538	0.03220	0.00007	0.00089
39	0.08950	0.21571	0.00141	0.00140
40	0.00152	0.00134	0.03212	0.03022
41	0.01276	0.01860	0.00155	0.00167
42	0.01062	0.01059	0.01622	0.01363
43	0.02229	0.00326	0.04245	0.00885
44	0.02210	0.02197	0.00562	0.00778
45	0.00465	0.00213	0.01672	0.00827
46	0.01404	0.00554	0.00285	0.00313
47	0.01175	0.00365	0.01575	0.00538

(continued on next page)

Table B.1 (continued from previous page)

Cell no.	α	β	γ	δ
48	0.03352	0.02448	0.03530	0.02461
49	0.06204	0.00679	0.03678	0.00735
50	0.06337	0.33655	0.00124	0.00548
51	0.00466	0.01253	0.00147	0.00567
52	0.01619	0.02251	0.00439	0.00699
53	0.00293	0.00375	0.00777	0.00765
54	0.00335	0.00158	0.02175	0.00623
55	0.01448	0.07803	0.00099	0.00903
56	0.11093	0.14324	0.00292	0.00318
57	0.00528	0.00283	0.09030	0.04282
58	0.02388	0.20272	0.00010	0.00110
59	0.01084	0.01136	0.05655	0.04543
60	0.06405	0.04461	0.00949	0.00505
61	0.00346	0.00317	0.02177	0.01585
62	0.01525	0.00717	0.00220	0.00157
63	0.03034	0.01025	0.00776	0.00607
64	0.06265	2.56571	0.00016	0.00296
65	0.11264	1.51085	0.00075	0.00873
66	0.01637	0.03496	0.00059	0.00177
67	0.00203	0.00052	0.04151	0.01217
68	0.00547	0.00242	0.00419	0.00192
69	0.00359	0.00384	0.00746	0.00772
70	0.32720	1.47635	0.00209	0.00670
71	0.00647	0.00156	0.03204	0.00588
72	0.03930	0.00275	0.15352	0.01521
73	0.01133	0.00240	0.06543	0.01194
74	0.04733	0.10892	0.00132	0.00305
75	0.00531	0.01213	0.00446	0.01051
76	0.00229	0.00547	0.00145	0.00284

(continued on next page)

Table B.1 (continued from previous page)

Cell no.	α	β	γ	δ
77	0.00223	0.00248	0.00192	0.00261
78	0.02787	0.05166	0.00826	0.01436
79	0.01156	0.00873	0.00800	0.00357
80	0.01127	0.04902	0.00169	0.00918
81	0.09549	0.02458	0.03745	0.02486
82	0.02149	0.05624	0.00458	0.01399
83	0.01898	0.00898	0.01585	0.00758
84	0.11931	0.11867	0.01596	0.01563
85	0.00504	0.01418	0.00735	0.01144
86	0.03671	0.21060	0.00152	0.01314
87	0.04409	0.01516	0.00656	0.00437
88	0.01058	0.07735	0.00060	0.00514
89	0.01943	0.06627	0.00672	0.01225
90	0.00313	0.00383	0.00127	0.00240
91	0.00320	0.00333	0.00710	0.00661
92	0.01679	0.01805	0.00080	0.00113
93	0.03798	0.03769	0.01489	0.01082
94	0.44651	0.74615	0.00578	0.00554
95	0.00605	0.00889	0.00495	0.00541
96	0.01872	0.07919	0.00159	0.00543
97	0.01310	0.13675	0.00007	0.00045
98	0.00113	0.00269	0.00738	0.01911
99	0.00818	0.00864	0.00070	0.00110
100	0.00136	0.00400	0.00033	0.00171

Bibliography

- [1] Karl Johan Åström and Richard M Murray. *Feedback systems: an introduction for scientists and engineers*. Princeton university press, 2010.
- [2] D Ewen Cameron, Caleb J Bashor, and James J Collins. “A brief history of synthetic biology.” In: *Nature reviews. Microbiology* 12.5 (2014), pp. 381–90.
- [3] Stefano Cardinale and Adam Paul Arkin. *Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems*. 2012.
- [4] Domitilla Del Vecchio, Aaron J Dy, and Yili Qian. “Control theory meets synthetic biology.” In: *Journal of the Royal Society, Interface / the Royal Society* 13.120 (Jan. 2016).
- [5] B. A. Foss, T. A. Johansen, and A. V. Sørensen. “Nonlinear predictive control using local models - applied to a batch fermentation process”. In: *Control Engineering Practice* 3.3 (1995), pp. 389–396.
- [6] Chongyang Liu, Zhaohua Gong, Bangyu Shen, and Enmin Feng. “Modelling and optimal control for a fed-batch fermentation process”. In: *Applied Mathematical Modelling* 37.3 (2013), pp. 695–706.
- [7] Andreas Miliadis-Argeitis, Sean Summers, Jacob Stewart-Ornstein, Ignacio Zuleta, David Pincus, Hana El-Samad, Mustafa Khammash, and John Lygeros. “In silico feedback for in vivo regulation of a gene expression circuit.” In: *Nature biotechnology* 29.12 (Jan. 2011), pp. 1114–1116.

- [8] Jared E Toettcher, Delquin Gong, Wendell Lim, and Orion D Weiner. “Light-based feedback for controlling intracellular signaling dynamics”. In: *Nature Methods* 8.10 (Jan. 2011), pp. 837–839.
- [9] Jannis Uhlenndorf, Agnès Miermont, Thierry Delaveau, Gilles Charvin, François Fages, Samuel Bottani, Gregory Batt, and Pascal Hersen. “Long-term model predictive control of gene expression at the population and single-cell levels.” In: *Proceedings of the National Academy of Sciences of the United States of America* 109.35 (2012), pp. 14271–14276.
- [10] Filippo Menolascina, Gianfranco Fiore, Emanuele Orabona, Luca De Stefano, Mike Ferry, Jeff Hasty, Mario di Bernardo, and Diego di Bernardo. “In-Vivo Real-Time Control of Protein Expression from Endogenous and Synthetic Gene Networks”. In: *PLoS Computational Biology* 10.5 (Jan. 2014), e1003625.
- [11] Justin Melendez, Michael Patel, Benjamin L Oakes, Ping Xu, Patrick Morton, and Megan N McClean. “Real-time optogenetic control of intracellular protein concentration in microbial cell cultures.” In: *Integrative biology : quantitative biosciences from nano to macro* 6.3 (2014), pp. 366–72.
- [12] Chiara Fracassi, Lorena Postiglione, Gianfranco Fiore, and Diego Di Bernardo. “Automatic Control of Gene Expression in Mammalian Cells”. In: *ACS Synthetic Biology* 5.4 (Jan. 2016), pp. 296–302.
- [13] Andreas Miliadis-Argeitis, Marc Rullan, Stephanie K Aoki, Peter Buchmann, and Mustafa Khammash. “Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth”. In: *Nature Communications* 7.May (Jan. 2016), p. 12546.
- [14] Evan J Olson, Lucas A Hartsough, Brian P Landry, Raghav Shroff, and Jeffrey J Tabor. “Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals.” In: *Nature methods* 11.4 (Jan. 2014), pp. 449–455.

-
- [15] Tal Danino, Octavio Mondrago, Octavio Mondragón-Palomino, Lev Tsimring, and Jeff Hasty. “A synchronized quorum of genetic clocks.” In: *Nature* 463.7279 (Jan. 2010), pp. 326–330.
- [16] Gianfranco Fiore, Giansimone Perrino, Mario Di Bernardo, and Diego Di Bernardo. “In Vivo Real-Time Control of Gene Expression: A Comparative Analysis of Feedback Control Strategies in Yeast”. In: *ACS Synthetic Biology* 5.2 (2016), pp. 154–162.
- [17] Giansimone Perrino and Diego Bernardo. “Modelling, simulation and control of single cell expression dynamics of the galactose-inducible promoter in yeast”. In: *2016 IEEE 55th Conference on Decision and Control (CDC)*. Cdc. 2016, pp. 3344–3349.
- [18] Giansimone Perrino, Cathal Wilson, Marco Santorelli, and Diego di Bernardo. “Control of gene expression for the study of neurodegenerative disorders: a proof-of-principle experimental study”. In: *IFAC-PapersOnLine* 49.26 (2016), pp. 8–13.
- [19] Drew Endy. “Foundations for engineering biology”. In: *Nature* 438.7067 (2005), pp. 449–453.
- [20] Gregory I. Lang and David Botstein. “A test of the coordinated expression hypothesis for the origin and maintenance of the GAL cluster in yeast”. In: *PLoS ONE* 6.9 (2011).
- [21] Irene Cantone, Lucia Marucci, Francesco Iorio, Maria Aurelia Ricci, Vincenzo Belcastro, Mukesh Bansal, Stefania Santini, Mario di Bernardo, Diego di Bernardo, and Maria Pia Cosma. “A Yeast Synthetic Network for In Vivo Assessment of Reverse-Engineering and Modeling Approaches”. In: *Cell* 137.1 (2009), pp. 172–181.
- [22] Filippo Menolascina. “Synthetic Gene Networks Identification and Control by means of Microfluidic Devices”. PhD thesis. 2011.
- [23] Gianfranco Fiore. “Identification and control of gene networks in living cells”. PhD thesis. Università degli Studi di Napoli Federico II, 2015.

- [24] George M Whitesides. “The origins and the future of microfluidics.” In: *Nature* 442.7101 (2006), pp. 368–73.
- [25] Mike S. Ferry, Ivan A. Razinkov, and Jeff Hasty. “Microfluidics for synthetic biology: From design to execution”. In: *Methods in Enzymology*. Methods in Enzymology 497 (2011). Ed. by Chris Voigt, pp. 295–372.
- [26] Raffaele La Brocca, Filippo Menolascina, Diego Bernardo, and Carlo Sansone. “Segmentation , tracking and lineage analysis of yeast cells in bright field microscopy images Time-lapse microscopy”. In: *1st International Workshop on Pattern Recognition in Proteomics, Structural Biology and Bioinformatics - PR PS BB* (2011), pp. 45–46.
- [27] Jean-Jacques E Slotine and Li Weiping. *Applied Nonlinear Control*. Vol. 199. 1. Prentice-Hall Englewood Cliffs, NJ, 1991, p. 461.
- [28] Matthew R Bennett, Wyming Lee Pang, Natalie A Ostroff, Bridget L Baumgartner, Sujata Nayak, Lev S Tsimring, and Jeff Hasty. “Metabolic gene regulation in a dynamically changing environment.” In: *Nature* 454.August (Jan. 2008), pp. 1119–1122.
- [29] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. “Stochastic gene expression in a single cell”. In: *Science* 297.5584 (Jan. 2002), pp. 1183–1186.
- [30] Peter S Swain, Michael B Elowitz, and Eric D Siggia. “Intrinsic and extrinsic contributions to stochasticity in gene expression.” In: *Proceedings of the National Academy of Sciences of the United States of America* 99.20 (Jan. 2002), pp. 12795–12800.
- [31] Gianfranco Fiore, Filippo Menolascina, Mario di Bernardo, and Diego di Bernardo. “An experimental approach to identify dynamical models of transcriptional regulation in living cells.” In: *Chaos* 23.2 (2013), p. 25106.

-
- [32] Manfred Morari and Jay H. Lee. “Model predictive control: Past, present and future”. In: *Computers and Chemical Engineering*. Vol. 23. 4-5. 1999, pp. 667–682.
- [33] David E Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. 1st. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2000.
- [34] Gene F Franklin, J David Powell, and Michael L Workman. *Digital Control of Dynamic Systems*. 3rd. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1990.
- [35] E Fossas, R Grinó, and D Biel. “Quasi-Sliding control based on pulse width modulation, zero averaged dynamics and the L2 norm”. In: *Advances in Variable Structure System, Analysis, Integration and Applications* (2001), pp. 335–344.
- [36] María Belén D’Amico and Fabiola Angulo. “Performance of a Zero Average Dynamics-controlled buck converter using different pulse-width modulation schemes”. In: *Int. J. Circuit Theory Appl.* 43.4 (2015), pp. 470–488.
- [37] Rafael R. Ramos, Domingo Biel, Enric Fossas, and Francesc Guinjoan. “A fixed-frequency quasi-sliding control algorithm: Application to power inverters design by means of FPGA implementation”. In: *IEEE Transactions on Power Electronics* 18.1 (2003), pp. 344–355.
- [38] H H McAdams and a Arkin. “Stochastic mechanisms in gene expression.” In: *Proceedings of the National Academy of Sciences of the United States of America* 94.3 (1997), pp. 814–819.
- [39] Jonathan M Raser and Erin K O’Shea. “Noise in gene expression: origins, consequences, and control.” In: *Science* 309.5743 (2005), pp. 2010–2013.
- [40] Mary J Dunlop, Robert Sidney Cox, Joseph H Levine, Richard M Murray, and Michael B Elowitz. “Regulatory activity revealed by dynamic

- correlations in gene expression noise.” In: *Nature genetics* 40.12 (2008), pp. 1493–1498.
- [41] Avigdor Eldar and Michael B Elowitz. “Functional roles for noise in genetic circuits.” In: *Nature* 467.7312 (2010), pp. 167–173.
- [42] Gábor Balázsi, Alexander Van Oudenaarden, and James J. Collins. *Cellular decision making and biological noise: From microbes to mammals*. 2011.
- [43] William J Blake, Mads KAern, Charles R Cantor, and J J Collins. “Noise in eukaryotic gene expression”. In: *Nature* 422.6932 (2003), pp. 633–637.
- [44] Ertugrul M Ozbudak, Mukund Thattai, Iren Kurtser, Alan D Grossman, and Alexander van Oudenaarden. “Regulation of noise in the expression of a single gene.” In: *Nature genetics* 31.1 (2002), pp. 69–73.
- [45] Llamosi Artémis, Andres M Gonzalez-Vargas, Versari Cristian, Cinquemani Eugenio, Ferrari-Trecate Giancarlo, Hersen Pascal, and Batt Gregory. “What Population Reveals about Individual Cell Identity: Single-Cell Parameter Estimation of Models of Gene Expression in Yeast”. In: *PLoS Comput Biol* 12.2 (Jan. 2016), pp. 1–18.
- [46] Joachim Almquist, Loubna Bendrioua, Caroline Beck Adiels, Mattias Goksör, Stefan Hohmann, and Mats Jirstrand. “A nonlinear mixed effects approach for modeling the cell-to-cell variability of Mig1 dynamics in yeast”. In: *PLoS ONE* 10.4 (2015).
- [47] Arren Bar-Even, Johan Paulsson, Narendra Maheshri, Miri Carmi, Erin K O’Shea, Yitzhak Pilpel, and Naama Barkai. “Noise in protein expression scales with natural protein abundance.” In: *Nature Genetics* 38.6 (2006), pp. 636–643.
- [48] Lennart Ljung. *System Identification (2nd Ed.): Theory for the User*. Ed. by Lennart Ljung. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1999, p. 519.

-
- [49] Lorraine V. Kalia and Anthony E. Lang. “Parkinson’s disease”. In: *The Lancet* 386.9996 (2015), pp. 896–912.
- [50] William Dauer and Serge Przedborski. *Parkinson’s disease: Mechanisms and models*. 2003.
- [51] Asa Abeliovich and Aaron D. Gitler. “Defects in trafficking bridge Parkinson’s disease pathology and genetics.” In: *Nature* 539.7628 (2016), pp. 207–216.
- [52] Tony Wyss-Coray. “Ageing, neurodegeneration and brain rejuvenation”. In: *Nature* 539.7628 (2016), pp. 180–186.
- [53] Mihael H Polymeropoulos, Christian Lavedan, Elisabeth Leroy, Susan E Ide, Anindya Dehejia, Amalia Dutra, Brian Pike, Holly Root, Jeffrey Rubenstein, Rebecca Boyer, Edward S Stenroos, Settara Chandrasekharappa, Aglaia Athanassiadou, Theodore Papapetropoulos, William G Johnson, Alice M Lazzarini, Roger C Duvoisin, Giuseppe Di Iorio, Lawrence I Golbe, and Robert L Nussbaum. “Mutation in the α -Synuclein Gene Identified in Families with Parkinson’s Disease”. In: *Science* 276.June (1997), pp. 2045–2047.
- [54] Michelle K Lin and Matthew J Farrer. “Genetics and genomics of Parkinson’s disease”. In: *Genome Medicine* 6.6 (2014), p. 48.
- [55] T Kitada, S Asakawa, N Hattori, H Matsumine, Y Yamamura, S Mizushima, M Yokochi, Y Mizuno, and N Shimizu. “Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism.” In: *Nature* 392.6676 (1998), pp. 605–608.
- [56] V. Bonifati. “Mutations in the DJ-1 Gene Associated with Autosomal Recessive Early-Onset Parkinsonism”. In: *Science* 299.5604 (2002), pp. 256–259.
- [57] Coro Paisán-Ruíz, Shushant Jain, E. Whitney Evans, William P. Gilks, Javier Simón, Marcel Van Der Brug, Adolfo López De Munain, Silvia Aparicio, Angel Martínez Gil, Naheed Khan, Janel Johnson, Javier Ruiz Martinez, David Nicholl, Itxaso Marti Carrera, Amets Saénz

- Peña, Rohan De Silva, Andrew Lees, José Félix Martí-Massó, Jordi Pérez-Tur, Nick W. Wood, and Andrew B. Singleton. “Cloning of the gene containing mutations that cause PARK8-linked Parkinson’s disease”. In: *Neuron* 44.4 (2004), pp. 595–600.
- [58] Alexander Zimprich, Saskia Biskup, Petra Leitner, Peter Lichtner, Matthew Farrer, Sarah Lincoln, Jennifer Kachergus, Mary Hulihan, Ryan J. Uitti, Donald B. Calne, A. Jon Stoessl, Ronald F. Pfeiffer, Nadja Patenge, Iria Carballo Carbajal, Peter Vieregge, Friedrich Asmus, Bertram Müller-Myhsok, Dennis W. Dickson, Thomas Meitinger, Tim M. Strom, Zbigniew K. Wszolek, and Thomas Gasser. “Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology”. In: *Neuron* 44.4 (2004), pp. 601–607.
- [59] Enza Maria. “Hereditary Early-Onset Parkinson ’ s Disease Caused by Mutations in PINK1”. In: *Science (New York, N. Y.)* 1158.2004 (2014), pp. 10–13.
- [60] Pavan K Auluck, Gabriela Caraveo, and Susan Lindquist. “ α -Synuclein: Membrane Interactions and Toxicity in Parkinson’s Disease”. In: *Annu. Rev. Cell Dev. Biol* 26.1 (2010), pp. 211–33.
- [61] F T Outeiro and Susan Lindquist. “Yeast cells provide insight into alpha-synuclein biology and pathobiology.” In: *Science* 302.5651 (2003), pp. 1772–1775.
- [62] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. 2006, XXII, 664 s. 85 illus.
- [63] Gabriele Pannocchia and James B. Rawlings. “Disturbance models for offset-free model-predictive control”. In: *AIChE Journal* 49.2 (2003), pp. 426–437.
- [64] William S. Cleveland. “Robust Locally Weighted Regression and Smoothing Scatterplots”. In: *Journal of the American Statistical Association* 74.368 (1979), pp. 829–836.

- [65] William Cleveland and S Devlin. “Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting”. In: *Journal of the American Statistical Association* 83.403 (1988), pp. 596–610.
- [66] Carsten Janke, Maria M Magiera, Nicole Rathfelder, Christof Taxis, and Simone Reber. “A versatile toolbox for PCR-based tagging of yeast genes : new fluorescent proteins , more”. In: *Yeast* 21.11 (2004), pp. 947–962.
- [67] Michal Breker, Melissa Gymrek, and Maya Schuldiner. “A novel single-cell screening platform reveals proteome plasticity during yeast stress responses”. In: *The Journal of Cell Biology* 200.6 (2013), pp. 839–850.
- [68] Luisa D’Amore, Giuliano Laccetti, Diego Romano, G. Scotti, and Almerico Murli. “Towards a parallel component in a GPU–CUDA environment: a case study with the L-BFGS Harwell routine”. In: *International Journal of Computer Mathematics* 92.1 (2015), pp. 59–76.