# Quantifying and Engineering Protein Dynamics in Bacteria

# Quantifying and engineering protein dynamics in bacteria

A DISSERTATION PRESENTED
BY
LAURENT POTVIN-TROTTIER
TO
THE COMMITTEE ON HIGHER DEGREES IN BIOPHYSICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
BIOPHYSICS

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
MAY 2017

Thesis advisor: Professor Johan Paulsson Laurent Potvin-Trottier

# Quantifying and engineering protein dynamics in bacteria

## Abstract

Genetically identical cells can display heterogeneity in their level of proteins due to stochastic gene expression. Such noise has been found to be widespread across biology, and can have a tremendous impact by allowing cells to access different phenotypes. However, the timescales of these fluctuations matter: slow fluctuations are potentially much more potent whereas if fluctuations are corrected rapidly they are relatively innocuous. Many simple mechanisms in cells – such as gene cascades, time-averaging to suppress fluctuations and epigenetic modifications – can create slow fluctuations. Quantifying protein dynamics remains a technical challenge as it requires measuring gene expression of long time series under constant growth conditions, and thus the timescales of fluctuations in gene expression remain largely unknown. By using a newly developed microfluidic device – to follow hundreds of cells for hundreds of cell divisions under constant growth conditions – we quantify the timescale of fluctuations of $\sim$ 50 transcriptional reporters and $\sim$10 translational reporters in *Escherichia coli*. All reporters show a strikingly similar and surprising behavior: an exponential decorrelation with a half-life of one generation. We show that the discrepancy with previous studies can be explained by artifacts arising from calculating autocorrelation functions with short time series or non-uniform growth conditions.

The general absence of slow fluctuations in cells opens up the question of how difficult it is to

create epigenetic memory with fluctuations in protein copy number. To answer this question, we revisit a synthetic oscillator, the repressilator, to engineer oscillations correlated over hundreds of generations. Synthetic gene circuits typically have much lower accuracy than their natural counterpart, and our hypothesis for this difference is that they are usually designed without considering stochastic gene expression. We used principles from stochastic chemistry in single cells to reduce error propagation and information loss. By simply removing features from the circuit rather than adding feedback loops, we created highly regular and robust oscillations, with circuits keeping their 14-generation periods in a wide range of growth conditions. The phase was kept for hundreds of generations such that flasks of cells and bacterial colonies displayed synchronous oscillations, even without coupling between cells.

Proteins cannot adapt to upstream changes on a scale faster than their lifetime. In order to respond to changes in their environment more quickly than their division time, cells must degrade their proteins. Fluctuations in the proteolytic machinery could therefore cause fluctuations in the half-lives of the degraded proteins. We develop and validate the first tool measuring instantaneous proteolysis rates in single bacterial cells. By measuring the saturation curve of *ssrA*-tagged proteins, we show that these substrates exhibit Michaelis-Menten kinetics, with very high affinity and a half-life of $\sim$45 seconds. We show that the SspB adapter protein increases the affinity of the ClpXP protease to *ssrA*-tagged substrates, which was previously demonstrated in vitro but not in vivo. However, depending on the substrate's local structure, degradation can be faster without the adapter. By measuring changes in degradation rates over time, we discover a proteolytic response: at high concentration of *ssrA*-tagged substrates, cells compensate by producing more proteases. We connect

this phenomenon to the heat shock response and characterize the associated toxicity. We discuss the limitations of using fluorescent proteins as reporters for degradation – fluorescent proteins targeted with natural tags can be partially degraded such that there is no decay in fluorescence – and suggest potential alternative methods for measuring degradation in single cells.

# Contents

To Kathy, mon amour, ma raison d'être.

# Acknowledgments

This work would not have been possible without the contribution and support of many people that I would like to thank here.

I would like to thank Johan "the beacon of light" Paulsson for being such a fantastic advisor during these years. His endless energy and passion for science were contagious, and he created and maintained a great open and collaborative work environment which enabled my research to blossom. Members of the lab were always isolated from all non-science external factors, such as ups and down of grants. I deeply appreciated his generous support during job applications, whether it was a typical career path or not. I enjoyed being bombarded by an endless stream of ideas, and last but not least, I come out enlightened with a new perspective, which is not simply about stochastic noise, but on how to approach problems differently and fundamentally.

I would like to thank all members of the Paulsson lab, past and present, for keeping a fun environment and for always being generous with their time. Thank you Andreas Hilfinger, for "being Andy", for so many coffee breaks, for working hard to keep the levels of coffee up, for your endless enthusiasm in fixing small problems, for showing how to keep your head up and for your moral support. Thank you Somenath Bakshi for your endless excitement about science and for all the interesting discussions; Silvia Cañas-Duarte for making the lab a little more normal; Charles Baker

for holding up the fort in the morning and for many lunchtimes; Charlotte Strandqvist for being always so positive; Emanuele Leoncini for keeping the Italian touch in the lab; Shlomi Reuveni for thoughtful discussions and advices; Ruoshi Yuan for being the nicest cubicle-mate; Janelle Vultaggio for somehow enduring the lab *and* making it run smoothly. Thank you Stephan Uphoff for setting the standard for the perfect lab citizen, for your incredible generosity, positivity and enthusiasm for science, it was an exceptional opportunity to work with you. Thank you Rishi Jajoo for your amazing patience, your endless enthusiasm for science, and for always being there to teach me biology; Martin Loose for insightful discussions and comments on the manuscript; Dirk Landgraf for being so generous with his time and somehow having made half of the strains that I needed; Nathan Lord for developing and sharing the experimental setup, his patience when I had so many questions at the wrong time and for always being positive and excited about science; Shay Tal for initiating the 12 o'clock lunch that I worked hard to perpetuate, Per Malkus for teaching me some biology; and many other members who contributed to this great environment: Yoonseok Jung, Burak Okumus, Jiawei Yan, Sadik Yildiz, Ghee Chuan Lai, Thomas Norman.

I would also like to thank my colleague and friend Scott Luro, or as some people refer to my "work wife". It was an exceptional opportunity to work with such a talented experimentalist (i.e. the "magic hands"). I appreciated your endless energy and having such an extremely hard-working collaborator (even when you were "working at the library"). I always had to cut down by half my suggestions of things to do, because I knew you were going to make three times as much. You know a collaboration is fruitful when the results are more than the sum of the parts; I think that in this case, the parts were multiplied.

I would like to acknowledge and thank people who contributed to other projects not included in this dissertation: Stephan Uphoff, Andy Yuan and Ann Hochschild, Genevieve Dobihal and David Rudner, David Riglar, Alexander Naydich and Pamela Silver. Sharing and discussing exciting new ideas is what I love the most about science. It was an extraordinary opportunity to work with so many talented people.

I would also like to express my deepest gratitude to the people of the Systems Biology department. You would be hard-pressed to find a department with a more open, collaborative and supportive environment. Everyone was always there to help and share their time. This exceptional community is what made this work possible.

Thank you Jim Hogle and Michele Jakoulov, for managing the Biophysics PhD program, aka "the best" PhD program. From Michele color-coded notes to Jim attending all the qualifying exams, dissertation advisory committee meetings and defenses, students were always well-supported and taken care of. There was never any technical hurdles in my way, and the program's flexibility is what allowed my PhD to be such a great experience.

I would like to thank my previous advisors and mentors, Caroline Boudoux, Paul Wiseman and Rick Horwitz. Thank you for supporting me, encouraging me and for being always so enthusiastically positive.

I would also like to thank the members of my qualifying examination committee as well as members of my dissertation advisory committee. Thank you Richard Losick, Pamela Silver, David Rudner, Peter Chien, Andrew Murray, for your generosity, insightful comments, and support.

I would like to thank my long-term friends, Amélie St-George-Robillard, Benoit Bourassa-Moreau

and Jean-Phillipe Coutu, for your support throughout these years, for sharing my excitement about random things, for your enthusiasm about my science and for so many good times that kept me going.

I am deeply grateful to my parents, François Trottier and Jasmine Potvin, for their continuous support along whatever path I follow. My dad gave me a passion for the outdoors, as well as basic scientific skills by showing me how to do experiments (i.e. blowing up glassware). My mom has always inspired me with her tremendous energy, joie de vivre, and courage even in the most difficult times. Thank you Michel Beauchemin, for teaching me patience as a kid, a vital quality for all scientists.

I would like to acknowledge financial support from Natural Sciences and Engineering Research Council of Canada (NSERC) and Fonds de Recherche du Québec - Nature et technologies (FRQNT).

I would like to end by thanking, Kathy Beaudette, my fiancée, partner and life companion. Thank you for always being there, whether through the ups and downs of graduate school or just anytime, and for following me in my (numerous) crazy pursuits and adventures. You are what keeps me going no matter what. I would simply not be where I am today without you. This work is for you.

# 1

# Introduction

Protein levels can vary greatly between genetically identical cells in the same environment. Such 'noise' can arise from random expression events or from uneven partitioning of molecules between the two daughters at cell division. At high abundances, the noise magnitude can become small compared to the average, but it is also possible for noise to propagate between network components in a way that is independent of abundances. For example, mRNAs are often present in such low num-

bers per cell that fluctuations arise spontaneously, and cells that have twice as much of an mRNA may produce the protein at twice the rate, transmitting the fluctuations in mRNA abundances to proteins. Because molecules such as mRNAs and proteins largely determine the identity of the cells, genetically identical cells exposed to the same environment will exhibit different behaviors. Cells have evolved to tolerate, control or exploit this noise.

Fluctuations in the number of molecules per cell are often quantified by estimating the statistical distributions over a population. However, the timescales on which numbers vary around the mean (i.e. how rapidly deviations are corrected) are also crucial, but have not been quantified as accurately or as systematically. Specifically, fluctuations that are rapidly corrected cannot substantially affect downstream processes because their levels change before those processes have had a chance to respond. Slow fluctuations can by contrast create long-lasting, possibly even epigenetic, phenotypes.

Several high-throughput studies have surveyed protein distributions and revealed widespread variability between cells. Protein dynamics in single cells have in turn primarily been studied in special cases. One reason is that quantification of protein dynamics is technically more challenging than taking snapshots. In fact, statistically quantifying the timescales can require long-term imaging of cells growing under precisely controlled growth conditions, as any changes in the conditions would appear as slow fluctuations. Protein time series in bacteria are typically acquired by imaging cells as they grow and divide on an agar pad. However, the growth conditions are then changing as the bacteria form a micro-colony and the cells start competing for limited resources. Additionally, the cells quickly become crowded in the field of view and the fluorescence in one cell is affected by its neighbors simply due to the point-spread function of light.

Recent developments in microfluidics have yielded devices that enable researchers to follow multiple well-separated cell lineages as they grow and divide for long periods of time, with fresh medium washing away the newborn cells. The growth conditions – whether rich or poor – can then be kept precisely constant over time, making the results easier to interpret.

We leveraged this high-throughput quantitative setup to both measure and engineer protein dynamics on multiple timescales in bacteria. We will start by surveying the timescales of fluctuations for natural genes in *E. coli*. This will show a striking universality: all measured genes decay on a timescale of one cell division. By using a synthetic biology approach, we will then engineer very slow protein dynamics. Specifically, by using the quantitative platform and insights from stochastic theory we could modify a classic synthetic circuit – the repressilator – to create oscillators that keep phase for more than 200 generations before drifting out phase by half a period. Finally, we will quantify the kinetics of protein degradation, which allows cells to adapt to their environment on a timescale faster than one generation. The half-life of the substrates targeted for degradation will be as low as 0.03 generation.

The dissertation is structured as follows. In chapter 2, we will introduce stochastic gene expression and present theoretical results in order to develop an intuition for how noise transmission in chemical cascades affect the timescales of the fluctuations. We will then survey those timescales for ∼60 native genes in *Escherichia coli* and mention our efforts to describe similar properties in fission yeast. We then describe some interpretation artifacts that have arisen from analyzing short time series.

In chapter 3, we will show how we can engineer gene synthetic circuits with correlations lasting

over dozens of generations by using insights from stochastic theory. The chapter is composed of a reformatted, published article, reproduced here with permission from Nature.

After observing extremely long timescales, we will characterize how quickly can cells respond to their environment in chapter 4. We will introduce proteolysis in bacteria, with a focus on *E. coli*. We will then present the development and validation of the first method allowing measurement of instantaneous degradation rates in single cells. This tool will be used to measure the saturation curve of degradation tagged-substrate, which led us to discover a new proteolytic feedback response: when cells are challenged with a high concentration of substrates, they appear to respond by producing more proteases.

Because many details of the experimental methods are shared between the chapters, we present them side-by-side in appendix B.

*The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' but 'That's funny...'*

Isaac Asimov

# 2

# Fluctuation timescales in micro-organisms

The objective of this chapter is to provide a global survey of the fluctuation timescales in *E. coli*, by quantifying the protein dynamics of many natural genes under precisely controlled growth conditions.

We will start by briefly introducing stochastic gene expression, and the statistical metric we will use to quantify the timescales – the autocorrelation function. We will then introduce a toy model

in order to develop an intuition for how the timescales of fluctuations change when they are transmitted from one component to another in chemical cascades. After showing how plasmid-based reporters can bias the measurements, we quantify the fluctuation timescales of $\sim 60$ native genes. We will also discuss how using short time series for calculating autocorrelations can introduce artifacts that have completely confounded some conclusions in the field.

### Contributions

The results presented here are part of a large collaboration. The project started as Nathan Lord's Ph.D. project, during which he developed the microfluidic device, built the transcriptional reporters and imaged them. Yoonseok Jung in turn developed the microfluidic device for *Schizosaccharomyces pombe*, built the yeast reporters and imaged them. I built the *E. coli* translation reporters and imaged them. Due to space constraints, I will briefly summarize the results of Lord and Jung, and then focus on my results.

## 2.1 Introduction

It is now well established that stochastic fluctuations are ubiquitous in cells (for reviews, see references [1–4]). Such fluctuations arise from the stochastic chemistry inside the cells, as molecules present in low numbers randomly collide with each other. These fluctuating molecules, such as mRNAs and proteins, play important regulatory functions inside the cells, and thus, genetically identical cells exposed to the same environment can behave differently. Those fluctuations can be detrimental: accessing a wide range of phenotypes can make cells grow more slowly overall. How-

ever, cells can also exploit this non-genetic heterogeneity in many ways. For example, bacteria can stochastically switch to a slow-growing persister state where they tolerate antibiotics [5, 6]. This is different from resistance: after an antibiotic treatment, persisters will grow into a population where only a fraction of the population is resistant by chance. Noise can also be exploited by stem cells to differentiate probabilistically [7] or by cancer cells in their response to drug treatment [8].

Noise can be intuitively quantified in two ways: by the distribution of protein numbers in the population, and by the timescale of these fluctuations (i.e. how long does a cell maintain a state of the distribution). The timescale is important, as slow fluctuations are potentially much more potent. Take for example the case of the bacterial persisters. If the bacteria switched into the antibiotic resistant state for only a few seconds, any meaningful antibiotic treatment would eliminate the cells, making persistance ineffective.

Theoretical frameworks have been developed to understand stochastic gene expression and the experimental results obtained over the past 15 years [1, 9–15]. There are three main points which we will focus on.

First, the noise observed in cells is often attributed to transcription [16–23]. Messenger RNAs are typically present in low numbers in cells and are therefore more prone to low number stochastic effects. Genes can also stochastically switch between states that are effectively on or off in terms of expression, possibly creating bursts of mRNA production and creating substantial fluctuations even at higher numbers.

Secondly, this noise can be propagated, because the rates of chemical reactions often depend on the concentration of another species. If a protein is translated from an mRNA with fluctuating

levels, its own levels will also fluctuate, even if the protein is present in large numbers. The noise in the protein can be further propagated if it affects the expression of other genes.

Thirdly, noise can be time-averaged. For example, if the average lifetime of the protein is much longer than that of the mRNA (which is often the case), the protein will not have time to respond to mRNA fluctuations, and will instead effectively time-average its concentration and thereby partly prevent noise propagation. Time-averaging therefore appears as a trivial way to suppress noise. However, it comes at a cost: time-averaging prevents but also slows fluctuations. This trade-off between noise suppression and slow timescales suggests that slow protein dynamics could be common in cells.

In addition to time-averaging slowing down fluctuations, there are multiple reasons to expect fluctuations to be long-lived. Simple schemes such as genes affecting each other in a cascade will effectively put fluctuations through low-pass filters, suppressing them while also slowing them down. Epigenetic modifications [24–26], positive feedback loops [27], bistability [28] and zero-order ultrasensitivity [29] can also create memories over multiple generations.

While the distribution of native proteins in *E. coli* has been measured on a global scale [30], the timescales of these fluctuations remain largely unknown. In bacteria, synthetic constructs have been shown to decorrelate on a timescale of one generation [31–33]. However, this might not be representative of the behavior of natural genes. Taniguchi et al. measured extremely long correlation times of 8 to 21 generations in three native genes [30]. It was even reported that natural genes exhibit slow, transient oscillations with periods of up to 20 generations [34]. In human cells, a survey has shown most protein fluctuations to be corrected on the order of 1 to 5 generations [35].

Surveying the timescales of fluctuations has so far been challenging due to technical limitations. Quantifying slow fluctuations requires extraordinarily long time series under precisely constant growth conditions. Using short time series can introduce artifacts, as we will show in section 2.3.3. Minor changes in growth conditions over time can result in apparent slow timescales. Moreover, the fluorescence from other cells can affect the measurement if the cells are crowded in the field of view.

By using a newly developed microfluidic device that addresses all of these challenges, we sought to quantify the timescales of native genes in *E. coli*.

## 2.2 Quantifying protein dynamics

To quantify the fluctuation timescales, we will use the autocorrelation function:

$$A(T) = \frac{\langle (x(t) - \langle x(t) \rangle_t)(x(t+T) - \langle x(t) \rangle_t) \rangle_t}{\langle x(t)^2 \rangle_t} \tag{2.1}$$

where $x(t)$ is the concentration of molecules at time $t$ and $T$ is the time lag. The autocorrelation function measures the correlation coefficient between two time points separated by a time lag $T$. It intuitively represents the "memory" of the system: how long does it take before values at different time points become uncorrelated.

In order to build some intuition for the behavior of autocorrelation functions, we will start by looking at a simple toy model. A molecule $x$ is produced at a constant rate ($\lambda_x$) and eliminated at a constant rate per molecule ($\beta_x$). A second molecule $y$ is produced at a rate proportional to the number of $x$ ($\lambda_y$) and also eliminated at a constant rate per molecule ($\beta_y$). This may describe the

mRNAs ($x$) and the proteins ($y$):

$$\begin{aligned} x &\xrightarrow{\lambda_x} x+1 \\ x &\xrightarrow{\beta_x x} x-1 \\ y &\xrightarrow{\lambda_y x} y+1 \\ y &\xrightarrow{\beta_y y} y-1 \end{aligned} \tag{2.2}$$

Note that $\beta_y$ typically represents exponential dilution due to cell growth (e.g. $\beta = \ln 2/\tau_{div}$), but may also include additional elimination such as degradation, as long as the total rate is proportional to the number of molecules (e.g. $\beta = \beta_{dil} + \beta_{deg} + \beta_{bleach}$). In this case, because the rates are linear functions of the number of molecules, the chemical master equation can be solved exactly for the moments such as the averages and the covariances [10], giving the well-known results for the mRNA-protein model:

$$\begin{aligned} \eta_{xx} &= 1/\langle x \rangle \\ \eta_{xy} &= \frac{1}{\langle x \rangle} \frac{\tau_x}{\tau_x + \tau_y} \\ \eta_{yy} &= \frac{1}{\langle y \rangle} + \frac{1}{\langle x \rangle} \frac{\tau_x}{\tau_x + \tau_y} \end{aligned} \tag{2.3}$$

where we used the normalized covariances ($\eta_{xy} = \sigma^2_{xy}/(\langle x \rangle \langle y \rangle)$) and the molecules lifetimes ($\tau_i = 1/\beta_i$). The effects of time-averaging appear obvious: if $\tau_y \gg \tau_x$, then the second term of $\eta_{yy}$ vanishes and fluctuations are not effectively transmitted from X to Y. In other words, the noise

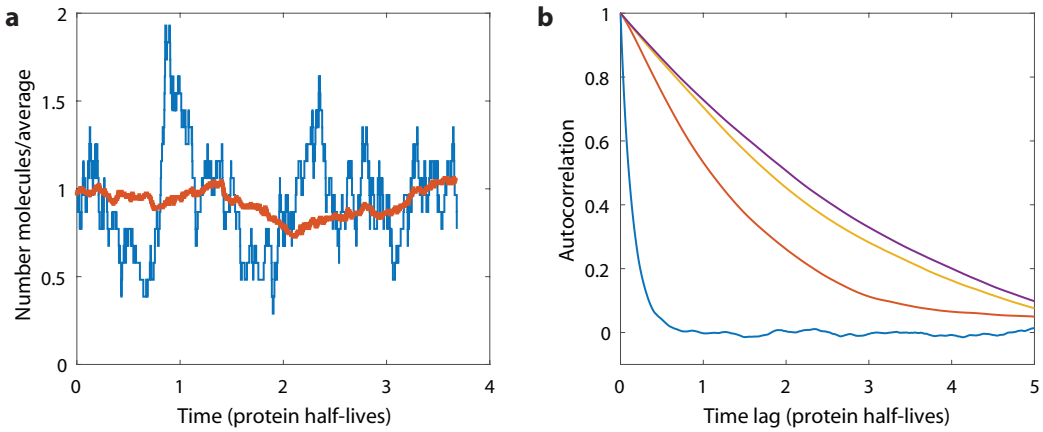



**Figure 2.1:** Stochastic simulations of simple models show the effects of time-averaging and reaction cascades. a) Simulated time trace of the mRNA-protein model. The normalized number of molecules (numbers/average numbers) is shown varying over time. The mRNA numbers fluctuate more (with respect to their mean, blue trace) and faster, with the protein time-averaging the fluctuations (red trace). The simulations were performed using equations (2.2), with ($\lambda_x = 100, \beta_x = 10, \lambda_y = 10, \beta_y = 1$). The autocorrelation functions show exponential decay with the half-life of the mRNA (blue) and the protein (red). The effects of cascades are shown, with the purple protein depending on the yellow which depends on the red, with progressively slower time scales.

would be the same if the rate of production was replaced by a constant.

We can also calculate the autocorrelation matrix using $\mathbb{A}(t) = \exp(-\mathbb{J}t)\mathbb{C}$, with $\mathbb{J}$ the jacobian matrix ($\mathbb{J}_{i,j} = \frac{\partial}{\partial x_j}\frac{d\langle x_i\rangle}{dt}$) and $\mathbb{C}$ the covariance matrix ($\mathbb{C}_{i,j} = \eta_{ij}$). We obtain:

$$\begin{aligned} A_{xx}(t) &= e^{-t/\tau_x} \\ A_{yy}(t) &= e^{-t/\tau_y} + \frac{\eta_{xy}}{\eta_{yy}}\frac{\tau_x}{\tau_y - \tau_x}\left(e^{-t/\tau_y} - e^{-t/\tau_x}\right) \end{aligned} \tag{2.4}$$

Note that we have normalized the autocorrelation function such that $A(0) = 1$ per convention. The autocorrelation for molecule $x$ simply decays exponentially with the time constant of the molecule's lifetime (Figure 2.1b, blue trace).

For molecule $y$, the picture is less clear. The first important point to note is that the function is

*always* greater than $e^{-t/\tau_y}$ (because the second term is always positive). This has important consequences: the autocorrelation cannot decay faster than the molecule's lifetime. Therefore, proteins cannot respond to changes faster than the division time unless they are degraded, a process which we will explore in chapter 4.

Let us now consider $\tau_y \gg \tau_x$, which is an approximation for the mRNA-protein model, as the lifetime of the proteins is typically much greater than the mRNA's. In that case, $\eta_{xy}$ goes to zero and the second term of (2.4) vanishes such that $A_{yy}$ is now equal to:

$$A_{dil}(t) = e^{-t/\tau_y} \tag{2.5}$$

where we renamed the autocorrelation dilution limited exponential. For stable molecules that are only diluted, the autocorrelation function will then half every cell division (Figure 2.1b, red trace).

To illustrate the effects of cascades, we now set $\tau_y = \tau_x = \tau$. We obtain:

$$A_{cas}(t) = e^{-t/\tau} + \frac{\eta_{xy}}{\eta_{yy}} e^{-t/\tau} \frac{t}{\tau} \tag{2.6}$$

In this case, the decay will be non-single-exponential, which will be more apparent for lag values on the order of $\tau$ (e.g. Figures 2.1b, yellow trace, and 2.2). The deviation from single exponential will depend on the $\eta_{xy}/\eta_{yy}$ ratio. This ratio goes to one if most noise arises from $x$ (i.e. $\langle y \rangle \gg \langle x \rangle$ ). Cascades of genes thus slow down fluctuations.

Finally, let us consider the case where fluctuations of the upstream component are much slower

($\tau_x \gg \tau_y$). This could represent $x$ being the number of transcription factors and $y$ now being the mRNAs.

$$A_{inh}(t) = \left(1 - \frac{\eta_{xx}}{\eta_{yy}}\right) e^{-t/\tau_y} + \frac{\eta_{xx}}{\eta_{yy}} e^{-t/\tau_x} \tag{2.7}$$

The inheritance of the timescale depends on the ratio between the upstream fluctuations and the total fluctuations ($\eta_{yy} = 1/\langle y \rangle + \eta_{xx}$). If most of the noise comes from the upstream fluctuations, the autocorrelation will decay with the timescale of their lifetime ($\tau_x$). However, if the noise comes from the fluctuations of $y$, most of the slow timescale will not appear and the autocorrelation will decay with timescale $\tau_y$. However, because $\tau_x \gg \tau_y$, the decay will have a long tail, where the small contribution from the upstream component will decay slowly.

## 2.3 Experimental results

### 2.3.1 Gene cascade

In order to illustrate to effects of noise transmission, we introduced the same transcriptional reporter ($P_{rpsL}$) on the chromosome and on a plasmid in the same cells (Figure 2.2). The autocorrelation of the chromosomal reporter showed exponential decay with a half-life of one generation, as expected if it does not inherit slow fluctuations because the reporter is stable. However, the reporter on the plasmid exhibited slower fluctuations. This is consistent with this reporter inheriting plasmid copy number fluctuations, as plasmid copy numbers are expected to vary on the timescale of one generation (similar to equation (2.6)).

There are two important points to note here. First, fluctuations can be transmitted from up-

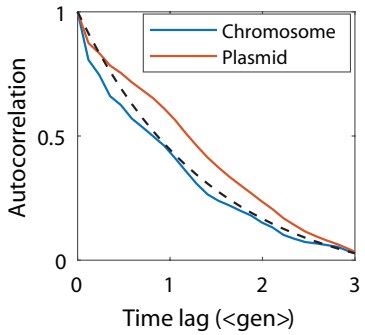



**Figure 2.2:** Autocorrelation functions of experimental system showing cascade inheritance. The transcriptional reporter ($P_{rpsL}$) is introduced simultaneously on the chromosome and on a low copy mini-R1 plasmid with compatible fluorophores (LPT40). The autocorrelation function for the chromosomal reporter (mVenus, blue trace) shows exponential decay with the half-life of one cell division, while the plasmid reporter (CFP, red trace) shows slower fluctuations, suggesting that it inherits plasmid copy number fluctuations.

stream components in a way that can be straightforwardly measured. Secondly, the use of plasmid vectors can greatly affect the dynamics of the transcriptional reporter by transmitting slow fluctuations. It is therefore necessary to use chromosomal reporters to quantify natural protein dynamics.

### 2.3.2 Survey of genes in *E. coli*

We sought to measure the timescales of fluctuations of native genes on a larger scale in *E. coli*. Because measuring autocorrelations requires long time series, and because any variation in the growth conditions can cause changes that appear as slow fluctuations, we used a microfluidic device [36–38] which allow us to image hundreds of cells for hundreds of generations under constant growth conditions (Figure 2.3). The bacteria are trapped in short channels and the newborn cells are washed away by the constantly supplied fresh medium. We verified that the properties of the cells in the device were uniform spatially and stable for multiple days (Figure 3.12).

We started by measuring the properties of ∼50 transcriptional reporters – integrated into the

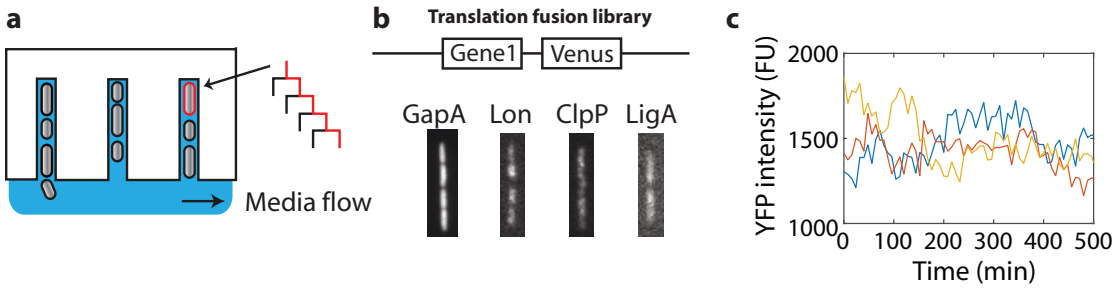



**Figure 2.3:** Experimental setup for measuring autocorrelations. a) Microfluidic device in which the bacteria are trapped in short channels and the newborn cells are washed away by the constantly supplied fresh medium. We follow the cell at the end of the channel -- the ``mother'' cell -- as it grows and divide for hundreds of generations. b) Schematic of the translation fusion library. Genes are fused at the C-terminus with a Venus at the native locus. Representative images of a cell trench for four reporters, from left to right, GapA-Venus (NDL208), Lon-Venus (NDL205), ClpP-Venus (LPT76) and LigA-Venus (NDL207). c) Three representative time traces for the GapA-Venus (NDL208) reporter.

chromosome to avoid artifacts from plasmid-based reporters (e.g. Figure 2.2). This library was created by integrating a previously generated plasmid-based GFP promoter library [39] at one neutral locus in the chromosome (attTn7), although similar results were obtained by using a different locus [36].

Strikingly, all reporters showed the absence of slow fluctuations, with the autocorrelation decaying with the half-life of one generation (Figure 2.4). This is consistent with the stable fluorescent proteins time-averaging fast promoter fluctuations. Indeed, because the reporters are not degraded, they cannot report on timescales faster than one generation.

Because slow fluctuations could arise from epigenetic mechanisms acting locally on the chromosome, we generated a library of ∼10 translation fusion reporters at their native loci (Figure 2.4b) by transferring a pre-existing Venus library [30] into our experimental strain. This could also report timescales faster than one generation. Four of these reporters had a low signal and the SulA reporter

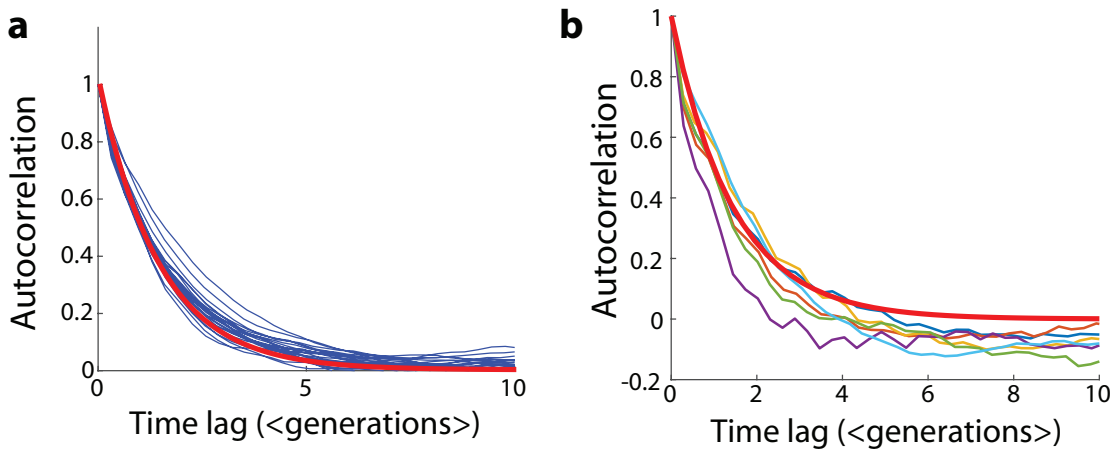



**Figure 2.4:** Autocorrelation functions show the absence of slow fluctuations for all reporters. a) Transcriptional reporters (blue traces) decay with the timescale of one generation, consistent with dilution limited exponential decay. Data from reference [36]. b) The translation reporter fusions also decay with a timescale of one generation (blue, SeqA-Venus, LPT74, red, ClpP-Venus, LPT76, orange, AtpD-Venus, LPT75, green, LigA-Venus, NDL207, baby blue, LigA-Venus, NDL208). The Lon-Venus (purple, NDL205) decays slightly faster, suggesting that it could be targeted for degradation. An exponential with a half-life of one generation is indicated in red in a) and b).

had a filamentous phenotype, most likely because the C-terminal fusion interfered with its degradation (i.e. similar to a *lon* mutant). Strikingly, the remaining reporters again decorrelated on the timescale of one generation. After correction for uncorrelated white noise, the reporters with low signal also appeared to decay on a similar timescale (Figure A.1). Only the Lon fusion had a slightly shorter half-life, suggesting that it might be targeted for degradation, which is consistent with the fact that it was identified in a screen for substrates of the ClpXP protease [40].

### 2.3.3 Artifacts from sub-sampling the autocorrelation function

In the previous section, we observed the absence of slow fluctuations in all of our reporters, which is in contrast with some other studies [30, 34]. Because small changes in the conditions can create apparent long-term correlations, we suggest that the discrepancy could be explained by the non-

homogeneous conditions of the previous experiments. However, one study used a similar microfluidic device and found slow oscillations in newborn cell length as well as natural genes in *E. coli* [34].

Tanouchi et al. report slow, transient oscillations – with periods of up to 20 generations – in the size of newborn *E. coli* cells. They found that ∼30-40 % of the cells display the oscillations in the autocorrelation function, only transiently, and that each such autocorrelation estimate oscillates with a different frequency. By observing the oscillations in different growth conditions, at different temperatures, in different organisms (*Bacillus subtilis* and *Schizosaccharomyces pombe*) and in the concentration of natural proteins, they conclude that this is a general mechanism in microbes. Finally, the authors also propose a simple mathematical model that fits the experimental data.

We suggest that these oscillations in fact represent an artifact of extreme undersampling of the autocorrelation function. The authors calculated the autocorrelation function for short time traces, and selected the fraction of estimates that looked oscillatory. For such short time traces ($\sim$ 70 time points instead of the $\geq$5,000 we used in the previous section), the autocorrelation can look oscillatory even if the time traces themselves do not oscillate. For example, by repeating their methodology, we could find a fraction of the time traces in our data set with autocorrelations that looked oscillatory, but we also found the same for a simple series of simulated dice rolls, which clearly cannot oscillate (Figure 2.5). For these handpicked autocorrelations, the time traces do not look oscillatory, and the autocorrelation is within the large lag error estimate [41].

The model that the authors used to recapitulate their data is called the auto-regressive 1 model (AR(1)) [42], which is solved and cannot produce oscillations for any positive rates. They used the power spectrum to represent the "probability of transient oscillations at a given frequency". How-

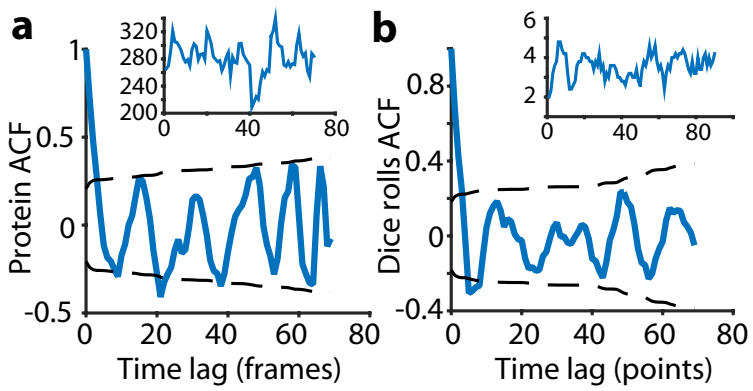



**Figure 2.5:** Low sampling of the autocorrelation function (ACF) can create autocorrelation estimates that look oscillatory. a) Example of autocorrelation estimate using a time trace of 70 points of the YFP concentration (SeqA-Venus, LPT74). The estimate can look oscillatory (but within the large lag error estimate, dashed lines, $\pm 1$ STE), while the time trace (inset) doest not. One time point represents 8 minutes. b) Example of autocorrelation estimate using the time series of a smoothed average of dice rolls. The estimate can look oscillatory even in this non-oscillatory system. The time trace (inset) does not appear to oscillate.

ever, using this definition would imply that every possible stochastic process would oscillate.

The *average* autocorrelation that the authors measure is the same we report: an exponential decay with a half-life of one generation. While it is possible to imagine a complicated process that switches stochastically between exponential dilution and oscillations, in the absence of a mechanism that could create such oscillations, we conclude that there are no evidence of slow, transient oscillations in these data sets or ours.

## 2.4 Discussion and conclusion

There are many reasons to expect that cells would exhibit slow fluctuations. The densely connected cellular network would suggest that proteins would inherit fluctuations from transcription factors, polymerases, ribosomes, etc. Epigenetic modifications could create memories lasting multiple gen-

erations. We measured the protein dynamics of transcriptional and translational reporters systematically in *E. coli* and observed no slow fluctuations. All the reporters showed a striking similarity in their autocorrelation functions: a simple exponential dilution with a half-life of one generation. That behavior can be well captured by a simple toy model of constant production rate (or fluctuating quickly) and exponential dilution (equation (2.5)) and rules out numerous network models where fluctuations are stabilized.

Our translational fusions also indicated that most proteins are not degraded, as previously reported [43]. Of the eleven fusions observed, one was degraded (SulA) and one had a slightly faster timescale that suggested degradation (Lon). However, we must point to a major caveat that we will investigate in chapter 4: fluorescent proteins can greatly interfere with protein degradation. As ClpAP is the only protease that can degrade fluorescent proteins from their N-terminus, we hypothesize that Lon is degraded by ClpAP. This hypothesis could easily be tested by mutating *clpA*. Making a good reporter for protein degradation is an exciting challenge that will be discussed in more details in chapter 4.

The experiments have also been repeated in a variety of growth conditions [36], which showed exactly the same behavior. Even more surprising, our group measured protein dynamics systematically in the fission yeast *Schizosaccharomyces pombe* and showed the absence of slow timescale, with fluctuations diluted out exponentially [44]. Taken together, the results point to a widespread mechanism of passive correction of fluctuations and the absence of slow timescales in most genes of *E. coli* and *S. pombe*. The results contrast with some previous reports that showed slow fluctuations. We suggest that these discrepancies were either due to non-uniform growth conditions or artifacts of

undersampling the autocorrelation function.

Given the complex networks in the cell, it is surprising that fluctuations are not transmitted in cascades. It suggests a fast, large and local source as the main contributor of fluctuations. Messenger RNAs are short-lived and present in low copies, which make them a plausible candidate. However, we still do not understand why mRNAs do not inherit slow fluctuations from upstream components such as transcription factors. One possibility is that they could operate at saturation, making them insensitive to changes in upstream components. This comes at a price of not being sensitive to *any* change in conditions, whether real or noise. Another possibility is that global upstream influences are tightly controlled so that they have low noise or shorter half-lives. In fact, many ribosomal protein subunits have been found to be degraded [40], even though producing ribosomes is one of the major limitation on growth. It could be that cells have taken precautions to avoid transmittal of slow fluctuations that are potentially much more potent.

*Dans les champs de l'observation le hasard ne favorise*

*que les esprits préparés.*

*In the fields of observation chance favors only the pre-*

*pared mind.*

Louis Pasteur, 1854

# 3

# Synchronous long-term oscillations in a synthetic gene circuit

In the previous chapter, we saw that typical genes in *E. coli* are corrected on the timescale of one generation and do not exhibit long term memory. These results bring up a new question: how slow can fluctuations be, or how long can we create epigenetic memory from fluctuations in gene expression?

This will take us through a different approach, where we will attempt to *engineer* long timescales in bacteria. Synthetic biology, or the engineering of biology for specific purposes, was kick-started in 2000 by the publication of the first synthetic oscillator, the repressilator [45], and the genetic toggle switch [46]. Here, we will revisit the repressilator by leveraging insights from stochastic gene expression to engineer accurate oscillators that can keep memories over long timescales.

Powerful theoretical frameworks have been developed to analyze gene expression, with exact solutions for linear systems (i.e. where the rates of the chemical reaction depends linearly on the concentration) and a good estimate for non-linear systems using linear noise approximation [10]. Further work has shown fundamental constraints on noise suppression [11, 15] – regardless of control mechanisms – that identify what cells absolutely cannot do. Tools for calculating noise from stochastic partitioning of molecules at cell division [47, 48], for separating extrinsic from intrinsic sources of noise [12, 13] and for strictly testing hypotheses in sparsely characterized gene networks [14] have also been developed. However, frameworks describing dynamical systems, such as oscillations, have remained elusive. By "building to understand" – using a synthetic circuit as a simple toy model where we know all the chemical interactions – we aim to also gain insights and intuition on the role of noise in dynamical systems.

## Contributions

This chapter is a published article reproduced here with permission from Nature [49]. Johan Paulsson and I conceived the study. Nathan D. Lord developed the microfluidic device and helped me with some of the experiments. Johan Paulsson, Glenn Vinnicombe and I performed the theoretical

analysis. Glenn Vinnicombe did other theoretical derivation not included in this thesis. I prepared the strains, performed the experiments and the data analysis.

## 3.1 Abstract

Synthetically engineered genetic circuits can perform a wide range of tasks but generally with lower accuracy than natural systems. Here we revisited the first synthetic genetic oscillator, the repressilator [45], and modified it based on principles from stochastic chemistry in single cells. Specifically, we sought to reduce error propagation and information losses, not by adding control loops, but by simply removing existing features. This created highly regular and robust oscillations. Some streamlined circuits kept 14 generation periods over a range of growth conditions and kept phase for hundreds of generations in single cells, allowing cells in flasks and colonies to oscillate synchronously without any coupling between them. Our results show that even the simplest synthetic genetic networks can achieve a precision that rivals natural systems, and emphasize the importance of noise analyses for circuit design in synthetic biology.

## 3.2 Main text

Many biological systems show remarkably precise and robust dynamics. For example, the circadian clock in cyanobacteria uses a combination of transcriptional and post-translational control mechanisms [50, 51] to keep phase for weeks without entrainment, while displaying robustness to changes in temperature and growth rate [51–53]. Synthetic circuits built from well-characterized parts can also exhibit a wide range of dynamical features – including arithmetic computations [54, 55], oscilla-

tions [45, 56–61], logic gates [62] and edge detection [63] – but often with lower accuracy. For example, the repressilator [45], a now iconic device that helped jump-start the field of synthetic biology 15 years ago, showed clear signs of oscillations using a simple design where three genes inhibit each other's production in a single loop (A⊣ B⊣ C⊣ A). However, only about 40% of cells were found to support oscillations, and those oscillations were quite irregular. Subsequent synthetic oscillators evaluated different control topologies or repression mechanisms [56–61], but most were again quite irregular in both phase and amplitude despite being mathematically designed to display sustained oscillations in a broad range of parameters.

The challenge when designing synthetic circuits to operate reliably in single cells is that biochemical noise can do more than just create different rate constants in different cells. On one hand, simple intrinsic noise can in principle enhance control [64] and even create high-quality oscillations in systems that could not display limit cycles for any rate constants in the absence of noise [65, 66]. On the other hand, any component present in low numbers can in principle randomize behavior of the whole system, and a single stochastic signaling step can introduce fundamental constraints [11] that cannot be overcome by any control system. This suggests that simplicity could even help achieve accurate oscillations as long as stochastic effects are accounted for in the design, and that minimal control topologies may not only be elegant and interesting but also very effective. We therefore revisited the original repressilator to reduce error propagation from the reporter system, from core cellular processes, and from within the circuit itself.

The repressilator consists of three genes – *tetR* from the Tn10 transposon, *cI* from bacteriophage $\lambda$ and *lacI* from the lactose operon – and each repressor has a C-terminal *ssrA* tag [67] that targets it

for degradation (Figure 3.1a). The whole circuit was encoded on a low-copy pSC101 plasmid in an *Escherichia coli* strain lacking *lacI*, and a second high-copy ColE1 reporter plasmid encoded GFP under the control of TetR with a modified degradation tag [45, 68]. We first reevaluated this circuit using a microfluidic device in which cells are trapped in short channels and newborn cells are washed away by fresh medium [37, 38] (Figure 3.1a). Tracking reporter levels under the microscope for hundreds of consecutive generations across hundreds of single-cell traces (Methods) revealed clear oscillatory dynamics in all cells (Figure 3.1b, Extended Data Figure 3.5a), particularly in the *rate* of production (Extended Data Figure 3.5a). This shows that the simple design was sound and that some of the erratic behavior originally reported was due to the limited imaging platforms available at the time.

We next evaluated how much of the noise reflected error propagation from the reporter system. Mathematical predictions have suggested that high-copy ColE1 cloning vectors fluctuate substantially and slowly, due to poorly controlled self-replication, and therefore effectively transmit fluctuations to encoded proteins [69]. Moving the YFP reporter onto the low-copy repressilator plasmid indeed reduced the relative standard deviation in amplitude greatly, from 78% to 36% (Figure 3.1d, Extended Data Figure 3.6a-b). Degradation tagged reporter proteins have also been predicted to potentially have 'retroactivity' effects on oscillations [70] due to competition for shared proteases. Protease competition has in fact been cleverly exploited for improved control in synthetic circuits [61], but stochastic theory [69, 71] suggests that saturated degradation enzymes also can create effects related to dynamic instability of microtubules, with large random fluctuations in single cells. Comparing a range of constructs indeed showed that the synthetic degradation tag caused fluctuations to propagate from the reporter proteins to the repressors via the proteolysis system (§3.5.1). Sur-

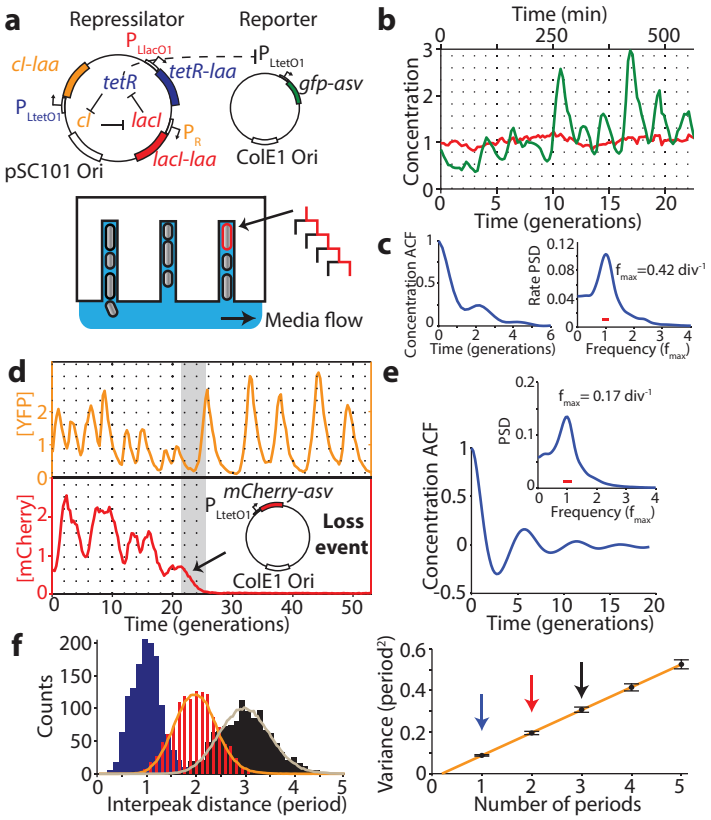



**Figure 3.1:** Reducing reporter interference. a) Schematics of the original repressilator plasmids as described in text and microfluidic device where *E. coli* cells are diffusively fed in growth channels and daughters eventually are washed away. b) Typical time trace of a single cell for original repressilator (NDL332). The GFP concentration (green trace) oscillates noisily while a constantly expressed RFP (red trace) stays constant. Both traces were normalized to their means. c) Autocorrelation functions (ACF) and power spectral densities (PSD) were calculated over the whole population (2,706 generations) and demonstrate oscillations with a mean period of 2.4 average division time. d) Top: oscillations are more regular when the reporter is expressed on the repressilator plasmid rather than on a separate high-copy plasmid (Extended Data Figure 3.6). Some cells irreversible shift period from ∼2.5 to ∼5.5 generations. Bottom: The period change was invariably connected to a loss of the separate mCherry-ASV-expressing reporter plasmid. Analysis of e.g. empty plasmid vectors, various reporter proteins and reporter degradation tags, and circuits with and without repressor degradation (§3.5.1 and 3.5.3) show that the interference was caused by the reporter *ssrA* degradation tag where the last three amino acids were substituted to ASV. e) ACF and PSD for the YFP expressing repressilator without separate reporter plasmid (LPT25), calculated over all 8,694 total cell divisions observed. Average period was 5.6 generations. Reporter protein close to fluorescence detection limit at troughs, and the actively degraded repressors should be much lower yet. The PSD was normalized by peak frequency, with width of the window function indicated by red line. f) Histograms of interpeak distances for one, two and three periods in blue, red and black respectively. Orange and grey lines were obtained by summing two or three samples (respectively) from the blue distribution. Consecutive periods are thus independent. Panel on right shows that the variance in period grows linearly with the number of periods elapsed (LPT25).

prisingly, however, the 'competing' reporter proteins *accelerated* the degradation of ssrA-tagged substrates (Extended Data Figure 3.7). Removing this interference created very regular oscillations, with periods increasing from ∼2.4 to ∼5.7 generations (Figure 3.1d-e, Extended Data Figure 3.5b, Extended Data Figure 3.6c-e). Characterizing the phase drift – the statistical tendency of oscillations in individual cells to go out of phase with each other – showed that on average this circuit oscillates for ∼5.5 periods before accumulating half a period of drift (Methods).

In cell-free extracts, i.e., without the low-copy noise of single cells, the repressilator has been shown to display the sinusoidal curves expected for harmonic oscillators [72]. However, analyzing the highly asymmetric shape of the time traces in single cells shows that it effectively operates as a *relaxation* oscillator [73], i.e., with a characteristic build-up phase sharply followed by an almost pure dilution and degradation phase until concentrations reach very low levels (Figure 3.1d, Extended Data Figure 3.8 and Box 1). The mathematical conditions for sustained harmonic oscillations – cooperative repression and similar mRNA and protein half-lives [45] – are then less relevant, but it becomes key to reduce the heterogeneity in the build-up and dilution phase (Box 1, §3.6.2) since stochastic effects otherwise fundamentally compromise the system's ability to keep track of time. Specifically, if the production phase for each of the repressors involves a low number of stochastic production events, statistical variation in that number will cause heterogeneity in peak amplitude, which then to some extent creates heterogeneity in the subsequent dilution and decay period. If peak protein abundances are low, random degradation events or uneven partitioning of molecules at cell division will in turn cause heterogeneity in the decay and dilution process. However, increasing peak abundances should only help marginally unless the repression thresholds are also increased

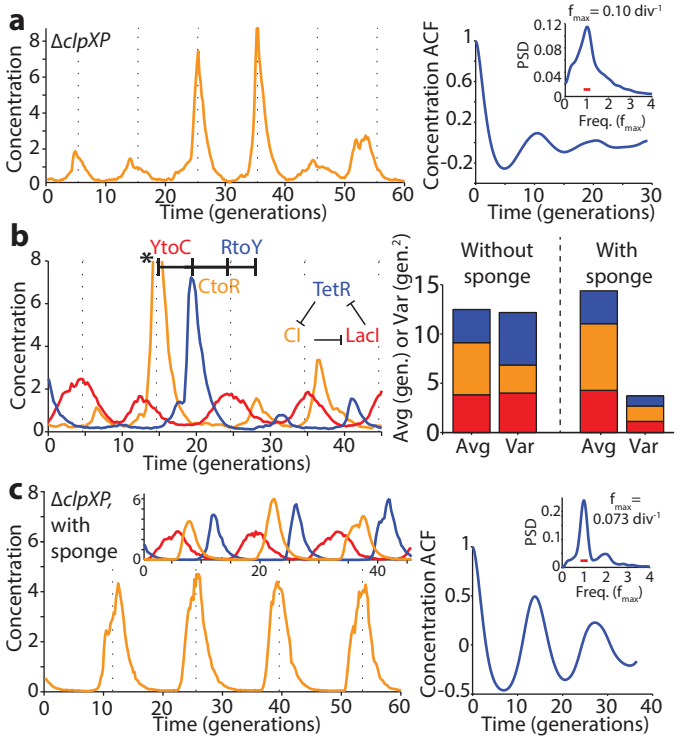



**Figure 3.2:** Identifying and eliminating inherent sources of error. a) Typical time trace in $\Delta$*clpXP* cells (LPT61) where repressors are not degraded. ACF and PSD calculated over 5,356 cell divisions. The average period was 10 generations, and the correlation coefficient was 0.1. Dashed vertical lines are separated by an average period to illustrate periodicity in a-c. b) (Left) Time trace of multi-reporter repressilator ($\Delta$*clpXP*, LPT113). TetR represses the production of YFP (yellow trace), LacI inhibits the production of CFP (blue trace) and CI represses the production of RFP (red trace). Peak indicated by asterisk not shown due to its high amplitude of 11.5 units. (Right) Interpeak distances evaluated for YFP to CFP (YtoC, red), CFP to RFP (CtoR, yellow) and RFP to YFP (RtoY, blue) without (LPT113, $n = 163, 150$ and 173) and with the titration sponge (plasmid with $P_{Ltet}$ binding sites, LPT117 and LPT127, combined, $n = 109, 86$ and 116). Respective contributions to the average and variance shown by bar plot. The RtoY part of the oscillation (induction of YFP, low TetR levels) represents 27% of the period, but contributes 44% of the variance. Addition of the $P_{Ltet}$ titration sponge brings down the variance almost fourfold. c) Example time trace of single reporter repressilator with $P_{Ltet}$-*mCherry-asv* ($\Delta$*clpXP*, LPT64), along with ACF and PSD calculated over 3,695 generations. Oscillations have an average period of 14 generations and a correlation coefficient of 0.5 after one period. Inset shows a time trace from the triple reporter repressilator without degradation and with titration sponge (LPT127, color scheme as in b).

appropriately (Box 1), since the last few steps contribute disproportionally to the variance (Box 1, §3.6.2).

Motivated by these results, we eliminated repressor degradation by removing the *ssrA* degradation tags from the repressors, by using a Δ*clpXP* strain, or both (§3.5.3). These circuits indeed oscillated in all cells, with a period of ∼10 generations. However, as predicted the noise in the period was only slightly reduced (Figure 3.2a, Extended Data Figure 3.9c, Extended Data Figure 3.10). To pinpoint the reason we built a circuit with compatible fluorescent reporter proteins for each repressor. Analyzing the variance in the three interpeak distances showed that the noisiest phase was when TetR levels were low (Figure 3.2b). We then estimated the protein abundances from the partitioning errors at cell division (§3.5.5), and found that the derepression of the TetR controlled promoter occurs at an extremely low threshold.

The theory suggests that the regularity could be greatly improved if this threshold was raised, e.g. using a 'sponge' of repressor binding sites that soaks up small numbers of TetR molecules. The high-copy reporter plasmid included in the original repressilator design in fact already carried binding sites for TetR, and simply reintroducing it greatly reduced the noise in all steps (Figure 3.2b) whereas similar sponges for CI and LacI had minor effects (Extended Data Figure 3.11d) as expected. Titration may in fact be a particularly useful way to increase the thresholds because it can also help create sharp switches [74] (§3.6.3), which may or may not be necessary for oscillations in single cells but generally should increase accuracy.

These changes created a streamlined repressilator with highly regular oscillations that peak every ∼14 generations (Figure 3.2c). Each repressor spends several generations at virtually undetectable

concentrations (§3.5.5) followed by several generations at concentrations that completely saturate repression. The amplitude still displays some variation (Extended Data Figure 3.12d), but because the time it takes to dilute levels from a peak amplitude of $N$ to a threshold of $S$ depends logarithmically on $N/S$, little variation in amplitude is transmitted to the timing (Box 1, Extended Data Figure 3.12e). Indeed the phase drift was only ∼14% per period (Extended Data Figure 3.9d), i.e., on average the circuit should oscillate for ∼18 periods before accumulating half a period of drift (Methods). The theory shows that similar accuracy should be possible in systems where dilution in growing cells is replaced by first order degradation, and that it is not the slowness itself that creates accuracy, but the absolute number of proteins at peaks and troughs.

The periods of circadian clocks, as measured in hours, are often robust to changes in growth conditions. However, other intracellular oscillators may need periods that instead are robust relative to internal physiological time scales such as the generation times. Synthetic circuits generally do not display either type of quantitative robustness because periods depend on so many different parameters that change with conditions. That is in principle also true for the circuits above: as conditions change, plasmid copy numbers, RNA degradation, gene expression, cell volume, etc. change in non-trivial ways. However, the logarithmic dampening that makes individual periods insensitive to fluctuations in peak amplitudes (Box 1) is also predicted to make the number of generations per period insensitive to growth conditions (Box 1). Indeed we found that the circuit retained the 14 generation period under all conditions tested (Figure 3.3a), including growth at 25-37°C and in conditioned medium from early stationary phase culture where cells become much smaller and almost spherical. The combination of robustness to conditions and great inherent precision suggest that cells could

display macroscopic, population-scale oscillations without cell-cell communication. We therefore synchronized (Extended Data Figure 3.13) a liquid culture and maintained it in early exponential phase (Methods). We indeed found that whole flasks oscillated autonomously, with a period of the expected ~14 generations (Figure 3.3b). We also imaged the growth of large colonies originating from single cells containing the triple reporter repressilator. Because only the cells at the edges of the colonies grow significantly, cells in the interior were arrested in different repressilator phases, creating ring-like expression profiles much like the seasonal growth rings seen in tree stumps (Figure 3.3c, Extended Data Figure 3.11a). The regularity originated in the autonomous behavior of single cells – no connections were introduced and cells kept their own phase when merging into areas where the neighboring cells had a different phase (Extended Data Figure 3.11c).

The results above (summarized in Extended Data Figure 3.14) illustrate the importance of understanding genetic networks at the level of stochastic chemistry, particularly for synthetic circuits where the noise has not been shaped by natural selection and where the heterologous components and reporters used may interfere with cells. We hypothesize that if statistical properties are systematically measured and the mechanisms are iteratively redesigned based on general stochastic principles, the next generation of synthetic circuits could rival or even surpass the precision of natural systems.

## 3.3 Box 1 | Relaxation oscillations of the repressilator

Depending on parameters, simple repressilators can produce traditional harmonic oscillations with sinusoidal trajectories, as observed *in vitro*, or relaxation oscillations with separate build-up and relaxation phases (panels below) as we observe experimentally in single cells (middle panel). Due to

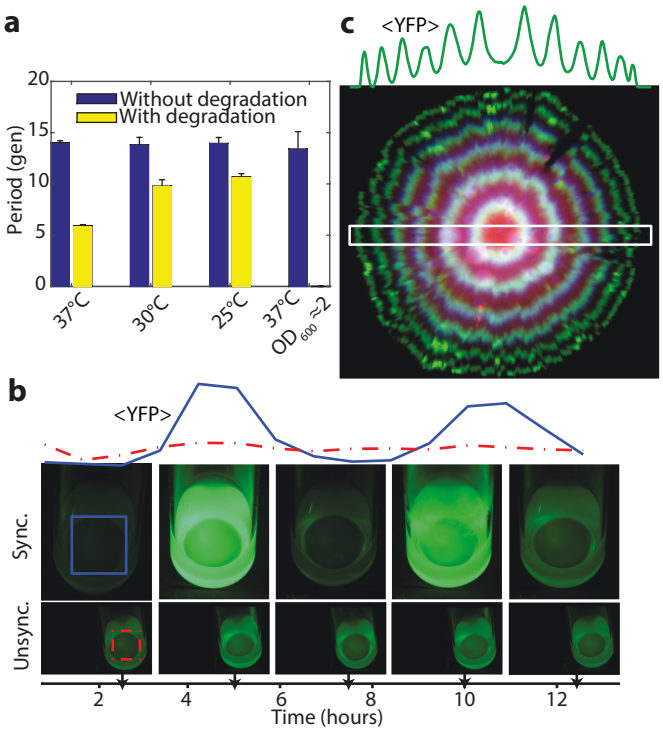



**Figure 3.3:** The modified repressilator shows great robustness to growth conditions. a) The repressilator without degradation and with titration sponge (LPT64) has a period of 14 generations at different temperatures (blue bars, division time of 27, 40 and 59 min for 37°C, 30°C and 25°C respectively) and in conditioned media ($OD_{600}$ ∼2, doubling time of 44 min). Repressilator with repressor degradation (LPT25) shows a varying period (yellow bars, doubling time of 26, 34 and 52 min for 37°C, 30°C and 25°C respectively). Error bars indicate STD on the first maximum of the ACF obtained by bootstrapping. b) Cells containing multi-reporter repressilator without repressor degradation and with $P_{Ltet}$-*peptide-asv* plasmid ($\Delta$*clpXP*, LPT117) were grown in liquid culture in 25mL flasks. After the culture was initially synchronized with IPTG, it was kept in exponential phase via dilution. Average YFP intensity shown for colored square area, with unsychronized culture for comparison. c) A ∼5 mm diameter colony of cells with the triple reporter repressilator (LPT117) reveals tree-like ring patterns in FP levels. The average YFP intensity is reported for the slice in the white rectangle. The decrease in RFP levels towards the edge of the colony is likely due to different response to stationary phase of its promoter.

low abundance fluctuation effects, single cells can in principle also achieve stable oscillations without the traditional requirements of cooperative repression and feedback delays (simulation example in left panel, details in §3.6). However, having low abundances introduces other constraints. For example, using Poisson communications theory [75] we demonstrate (SI §A[1]) hard limits on the ability of such systems to keep track of time, in terms of the average number of molecules $N$ at the peak of the oscillations, even if a repression control system could remember time series of decay events. For circuits like the actual repressilator, where repression is set by the current repressor level, the constraints are more severe yet, and limited both by heterogeneity in $N$ and the inherent noise of the first-order elimination process until levels reach the repression threshold $S$. Variation in $N$ can be reduced if repressors approach a steady state where production is balanced by elimination (left panel). Noise in $N$ also only has a damped effect on the time to reach a threshold because that time approximately depends *logarithmically* on $N/S$: doubling $N$ only adds one more half-life before reaching $S$. However, substantial noise can arise towards the end of each decay phase if the repression threshold $S$ is too low, as the last few steps then dominate the total decay time (right panel and inset). Specifically, the average length of the decay phase is $\sum_{i=S}^{N} 1/i \approx \log(N/S)$ while the variance is $\sum_{i=S}^{N} 1/i^2 \approx 1/S$ in units of protein half-lives (for details see §3.6.2), creating an optimal threshold $S_{opt}$ that minimizes the CV in the decay time. Increasing $N$ should only significantly reduce this CV if $S > S_{opt}$, have virtually no effect if $S < S_{opt}$, and decrease the CV in proportion to $\sqrt{N}$ if $S$ is close to $S_{opt}$ (Extended Data Figure 3.8e). Finally, the exponential nature of the decay phase allows the repressilator to be robust to growth conditions: changes in $N/S$ are logarithmically

[1]This appendix is not included in the dissertation.

damped in their effects on the average period. Though $N$ and $S$ change with conditions, for many repressors they may change with similar factors that cancel out in the ratio, e.g. because conditions with less gene expression (lowering $N$) also tend to produce smaller cell volumes (lowering $S$).

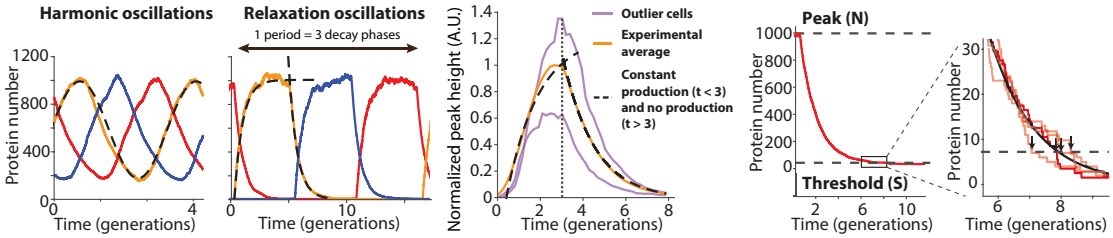
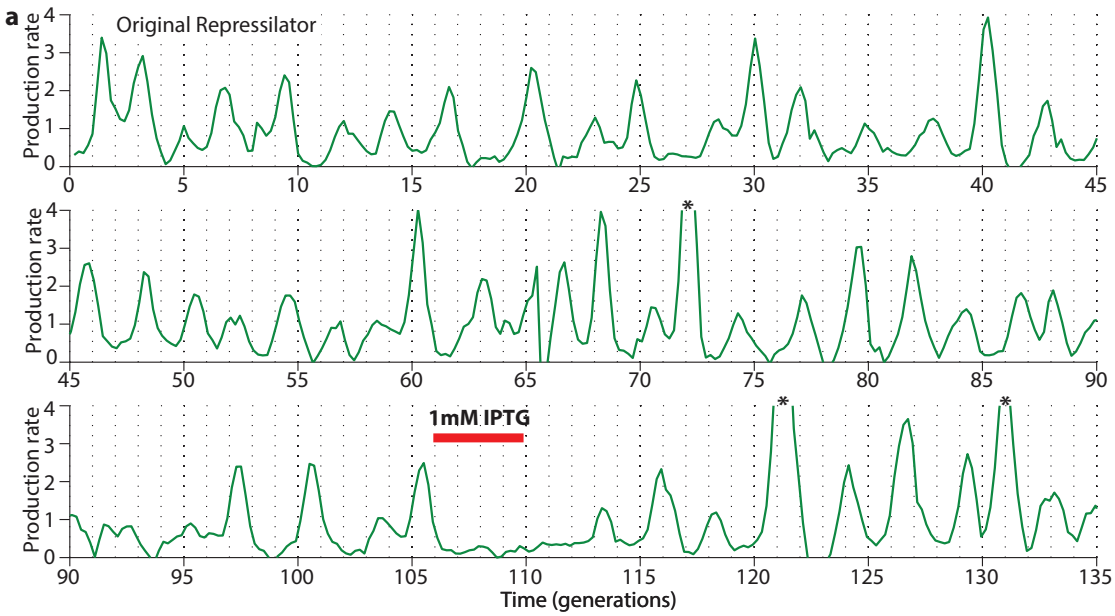


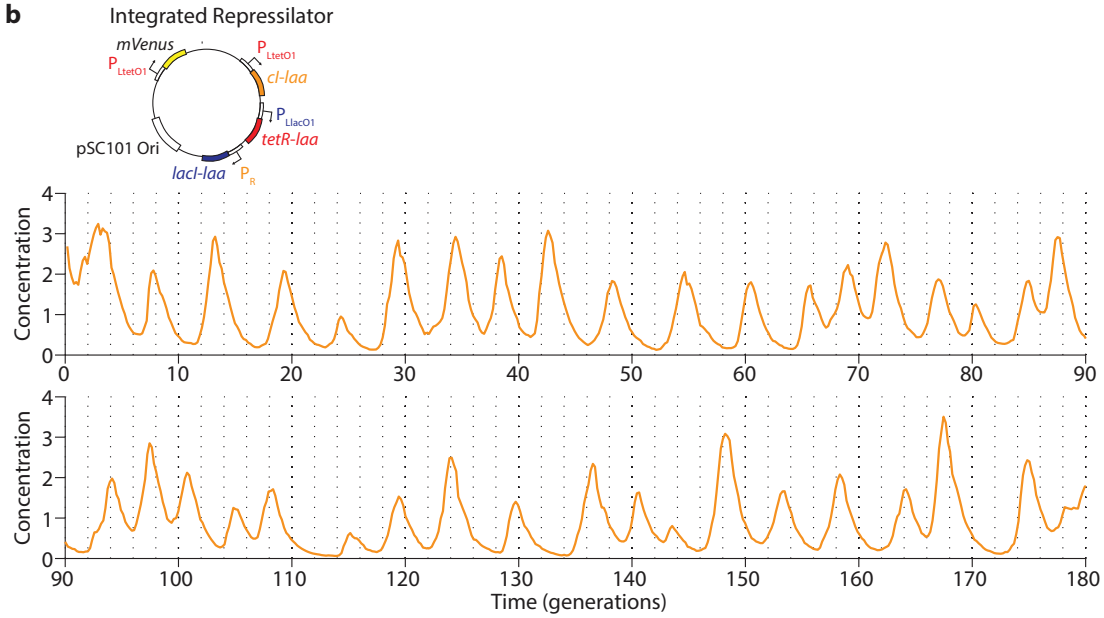

**Figure 3.4:** Left panel: Simulated time traces showing that the repressilator can produce harmonic oscillations, with shapes very close to a sine wave (dashed lines), or relaxation oscillations, depending on the parameters. Relaxation oscillations have two distincts phases: a build-up with constant production and dilution and a decay phase, with only dilution (dashed lines). The period can then be decomposed as a sum of three decay phases (one for each of the repressor, blue, red and yellow traces). Middle panel: The experimental data shows that the repressilator displays relaxation oscillations. The peaks were aligned and averaged (yellow trace) and fit well to the build-up and decay phases (dashed lines). Outlier cell traces (magenta) show how fluctuations are dampened by the exponential dilution. Right panel: The length of the decay phase depends on the log of the ratio between the peak height $N$ and the threshold $S$. Even for a large ratio, there can be significant heterogeneity in the timing if the threshold is low (inset).

## 3.4 Extended Data Figures

**Figure 3.5** ***(following page)*****:** Oscillations in the original and integrated repressilator circuits. a) Original (NDL332, GFP production rate) and b) integrated repressilator (LPT25, YFP concentration) oscillations are sustained for more than one hundred generations. The two time traces were normalized to their respective means. Three peaks in a) indicated by asterisks have been clipped due to their high amplitude (5.9, 7.1 and 4.8) to allow better visualization of the oscillations. IPTG was added to the media for the time period indicated by the red bar (in a) in order to synchronize the cells in the device.

**Figure 3.5:** (continued)



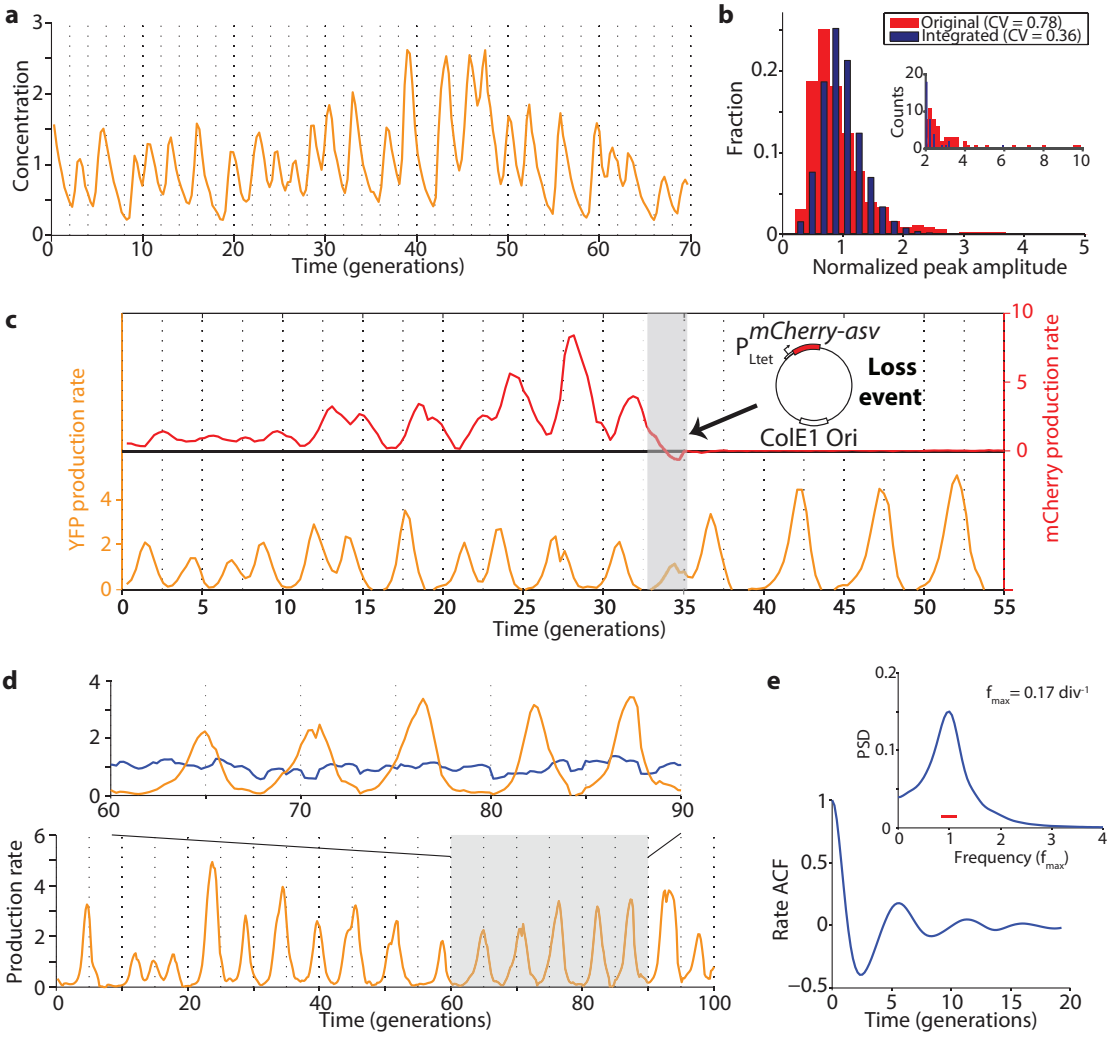
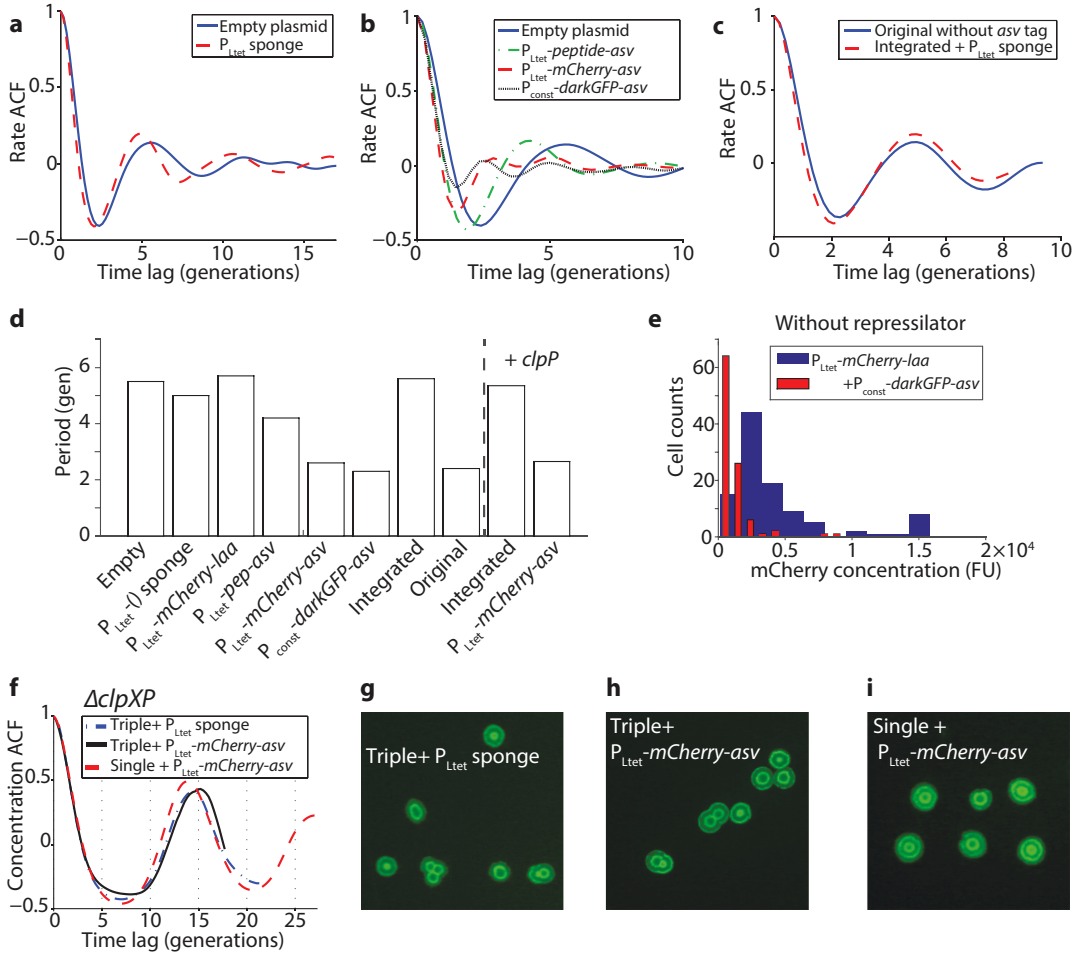



**Figure 3.6:** Interference from the reporter plasmid. a) Oscillations of the integrated repressilator with the $P_{Ltet}$-*mCherry-asv* plasmid (LPT54) have a more constant peak amplitude compared to the original repressilator, indicated by the CV of the peak amplitude decreasing from 0.78 to 0.36 (b). The inset in b) zooms in on the tails of the distributions. c) Additional plasmid loss event of integrated repressilator with $P_{Ltet}$-*mCherry-asv* reporter. The reporter plasmid is lost around generation 34, as evidenced by the loss of red fluorescence. The oscillation period shifts from $\sim$2 to $\sim$5 generations quickly after the plasmid loss event. d) Example time trace of the integrated repressilator, without the reporter plasmid (LPT25). The YFP production rate oscillates (yellow trace), while the segmentation marker (blue trace) stays relatively constant (close-up of the shaded region on top). Both traces were normalized to their respective means. e) Autocorrelation function (ACF) and power spectral density (PSD) were calculated over the whole population (8,694 total generations) and demonstrate strong oscillatory behavior, with an average period of 5.6 generations. The width of the window function used for calculating the power spectrum is indicated by a red line.

**Figure 3.7** ***(following page)*****:** Summary of results explaining the difference in period between the original and integrated repressilator. a) The $P_{Ltet}$ sponge (LPT44) makes the oscillation slightly shorter and more regular compared to the empty plasmid (LPT45), but cannot explain the change in period. b) Increasing the expression of `competing' substrates tagged with the *asv* tag makes the oscillations *faster*. The period gradually decreases from 5.5 generations for the empty plasmid (LPT45) to 4.2 with $P_{Ltet}$-*pep-asv* (LPT46), to 2.6 with $P_{Ltet}$-*mCherry-asv* (LPT54) and to 2.3 with $P_{const}$-*darkGFP-asv* (LPT53). c) Removing the degradation tag on the reporter of the original repressilator (LPT60) produces oscillations very similar to the integrated repressilator with the sponge (LPT44). d) Summary of the period of the different construct presented in this figure, compared to the original (NDL332) and integrated repressilator (LPT25). Introduction of ASV-tagged molecules is sufficient to explain the change in period, whereas introduction of LAA-tagged molecules slows down the oscillations (LPT55). Overexpressing a functional ClpP-mGFPmut3 fusion makes the period slightly faster (5.4 gen, LPT159), but does not rescue the effect of the ASV-tagged proteins (2.8 gen, LPT165). e) Expressing ASV-tagged molecules in the absence of the repressilator lowers the mean abundances of ssrA-tagged molecules $\sim$ four-fold, suggesting that presence of ASV-tagged molecules cause faster degradation rates. f) In the $\Delta$*clpXP* background, the oscillations are not affected by the presence of ASV-tagged molecules or additional reporter. Triple reporter with $P_{Ltet}$ sponge (LPT127), triple with $P_{Ltet}$-*mCherry-asv* (LPT118) and single with $P_{Ltet}$-*mCherry-asv* (LPT64) have very similar autocorrelation functions and ring patterns (g-h-i). There were slight variations in the imaging conditions due to manual focusing and non-uniformity of the LED illumination.

**Figure 3.7:** (continued)



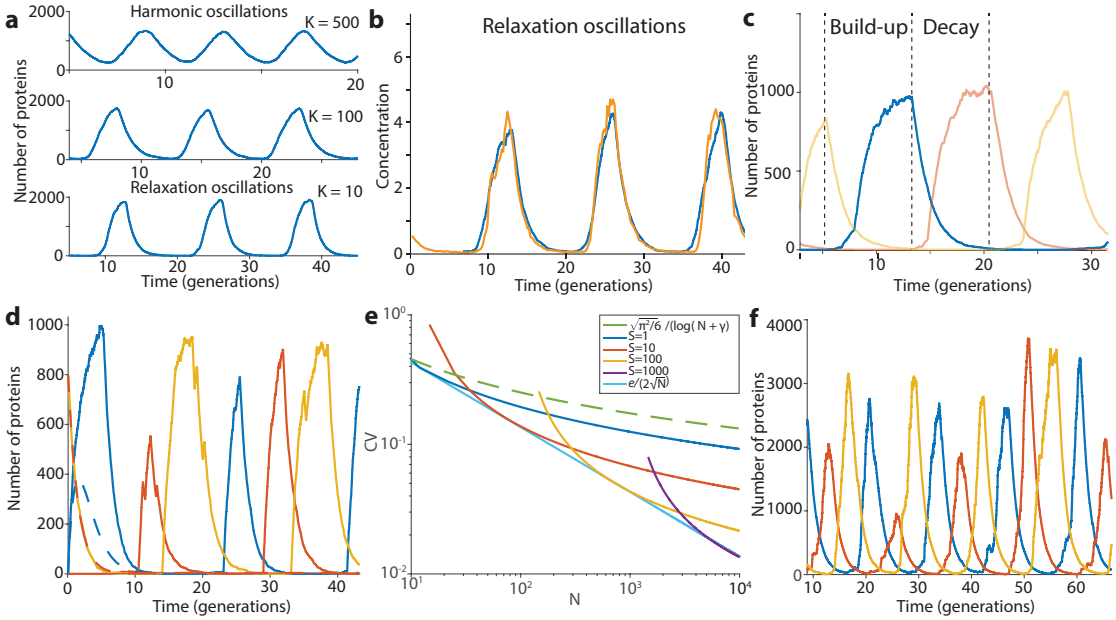



**Figure 3.8:** Modeling results. a) The repressilator can display harmonic or relaxation oscillations. The gradual transition between the regimes is shown here by varying the parameter $K$ in the minimal model ($\lambda = 2000$ and $n = 4$, §3.6.1). b) The experimental data suggest that the repressilator oscillates in the relaxation regime. Simulated time trace (blue, $K = 13, \lambda = 10^3$, and $n = 3$) is overlayed with time trace of experimental data (from Figure 3.2c, LPT64, yellow). c) Close-up of a simulated time trace (minimal model, $K = 0.2, \lambda = 10^3$, and $n = 2$, §3.6.1) in the relaxation regime showing the three different repressors (blue, red and yellow). The oscillations can be separated in two distinct phases: an accumulation phase during which the protein (blue) is completely derepressed (red below threshold) and starts at very low numbers, and a decay phase that starts when the repressor is completely repressed (red above threshold) and ends when it goes below the repression threshold of the next component (yellow starts to accumulate). d) Relaxation oscillators have different parameter requirements for oscillations. Simulated time traces (solid lines) show oscillations without biochemical cooperativity or phase shift due to the presence of mRNA (minimal model, $K = 0.01, \lambda = 10^3$, and $n = 1$, §3.6.1). The deterministic differential equations with the same parameters show damped oscillations (dashed lines with flipped colors). e) Even for perfect threshold mechanism, significant noise comes from the decay phase if the threshold ($S$) is too low (or too high) with respect to the peak value ($N$). If $S << N$, then the CV in one decay step goes down very slowly (1/log$(N/(S+1))$). However, if it is reasonably close to its optimal value (e.g. $0.05 < S/N < 0.3$), it goes down much faster ($1/\sqrt{N}$). The CV is shown for different combination of $S$ and $N$, as well as the asymptotic traces. f) Simulated time trace of the model of §3.6.3 shows oscillations of similar shape, peak amplitude numbers, period and phase drift as the experimental data by using reasonable parameters ($\lambda = 60, K = (5, 10, 10)$ for the three repressors, $n = 1.5$ for all repressors, $\langle b \rangle = 10, \langle N_o \rangle = 10$ $\langle N_t \rangle = 40$).

**Figure 3.9 *(following page)*:** Period histograms and kymographs of selected strains. Peak-to-peak distance of the oscillations was calculated as described in Methods, and the average period, as well as the CV (standard deviation divided by the mean) are reported in the figure panels. a) Original repressilator (NDL332). b) Integrated repressilator (LPT25). c) Integrated repressilator in $\Delta$*clpXP* (LPT61). d) Integrated repressilator in $\Delta$*clpXP* with $P_{Ltet}$-*mCherry-asv* (LPT64). e) Integrated repressilator with $P_{Ltet}$ sponge (LPT44). f) Integrated repressilator with $P_{Ltet}$-*mCherry-asv* (LPT54) g) Kymograph ($xy$-$t$ montage) of the raw data is presented for three strains. The image of a single growth channel is presented every 1, 2 and 7 frames (5 min/frame) for the top, middle and bottom panel respectively. The oscillations in concentration are difficult to see in the fast oscillator (although clear when looking at production rate), but they can be clearly seen in the slow oscillators. The growth channels are open towards the bottom of the images, where media is supplied.

**Figure 3.9:** (continued)

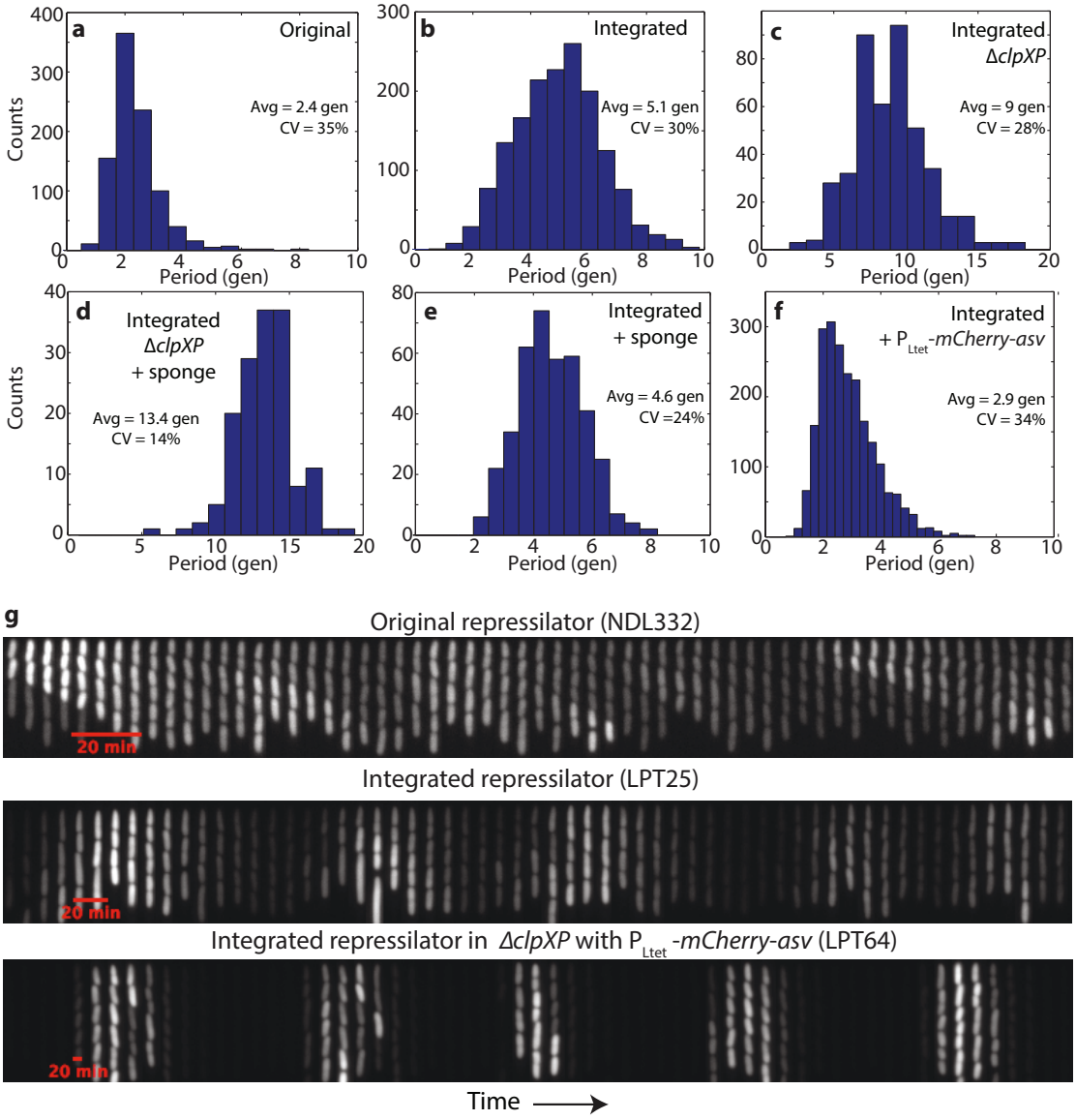


**Figure 3.10** ***(following page)*****:** Oscillations of the repressilator without degradation tags. a) Schematic of integrated repressilator without degradation tags, with or without the $P_{Ltet}$ titration sponge. b) Without the titration sponge (LPT120), the oscillations are erratic in amplitude, with a correlation coefficient of $\sim$0.1 after one period. c) Addition of the sponge (LPT124) makes the oscillations much more regular, with a correlation coefficient of $\sim$0.25 after one period. d) Time trace and autocorrelation of integrated repressilator without degradation tags in $\Delta$*clpXP* (LPT128). Introduction of the mutation did not change the oscillations substantially (compared to c). e) The colonies of integrated repressilator without degradation tags with $P_{Ltet}$ sponge (LPT124) exhibit spatio-temporal ring patterns in the YFP images. f) Close-up of the colonies show that the spatio-temporal patterns were similar if the titration sponge contained only the promoter (LPT124) or expressed an ASV-tagged peptide (LPT125), suggesting that these strains have similar oscillations.

**Figure 3.10:** (continued)

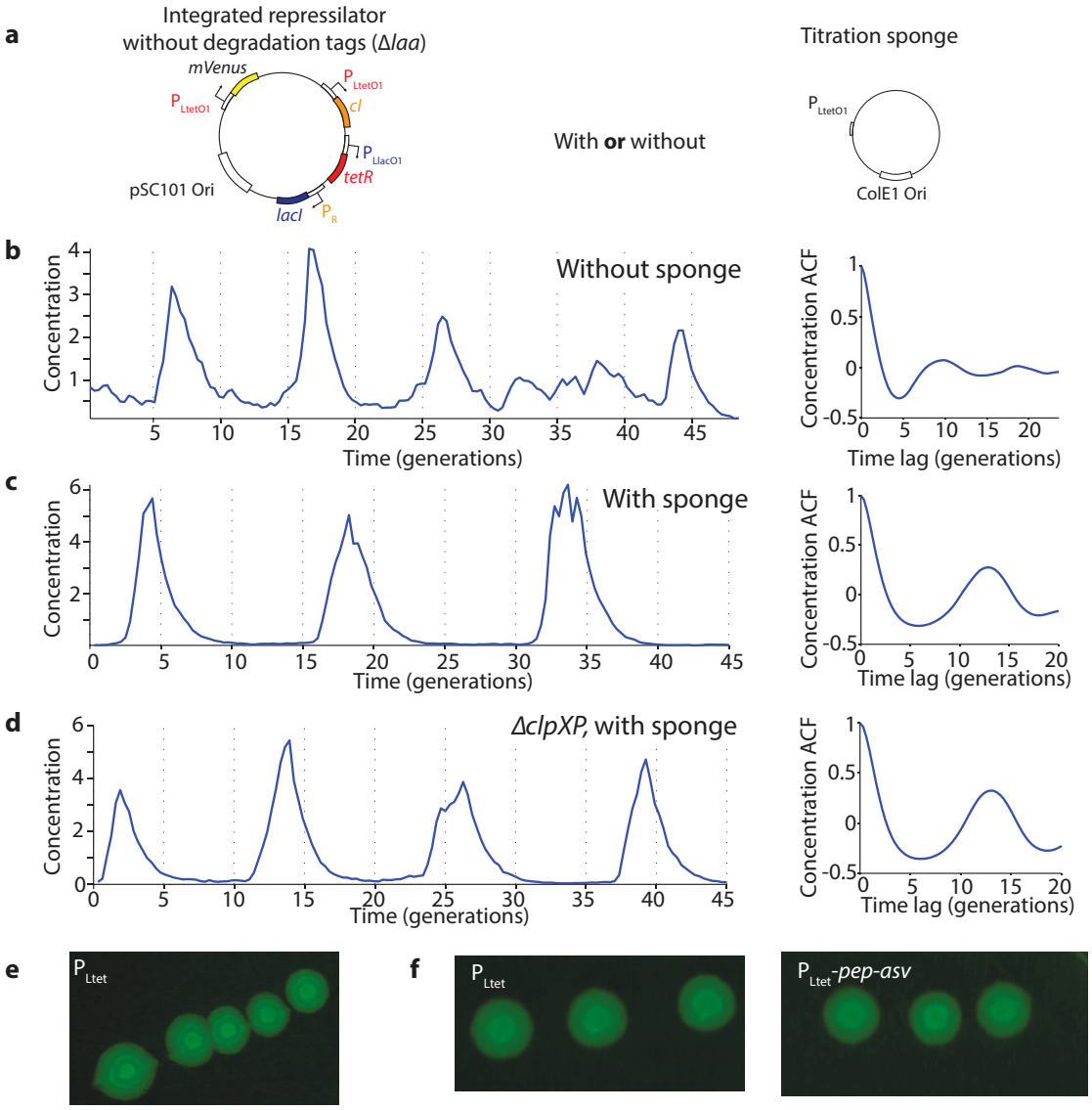


**Figure 3.11** ***(following page)*****:** Macroscopic spatial patterns of the repressilator. a) Time course growth of a single colony grown from a $\Delta$*clpXP* mutant cell containing the integrated repressilator and titration sponge (LPT64, $P_{Ltet}$-*mCherry-asv*). Oscillations in YFP levels produce macroscopic, ringed structures in the YFP channel (bottom row), while such patterns are absent in the constitutive segmentation marker (CFP, middle row) and gross colony morphology (bright field, top row). The bottom and left white spots in the brightfield images are reflections from the white LEDs. b) On the left, unsynchronized cells were plated and different phases of the oscillators are represented by different ring patterns (dark or bright center of different sizes). Synchronization of the cells with IPTG makes the patterns similar, with a dark center of the same size. c) The ring patterns do not synchronize when adjacent colonies merge into each other. d) Only the presence of the $P_{Ltet}$ sponge is required for macroscopic oscillations, while titration of the other repressors do not affect greatly the oscillations. From left to right: LPT153, LPT154, LPT157, LPT143, LPT155, LPT156 and LPT152. Several strains were also evaluated in the microfluidic device.

**Figure 3.11:** (continued)

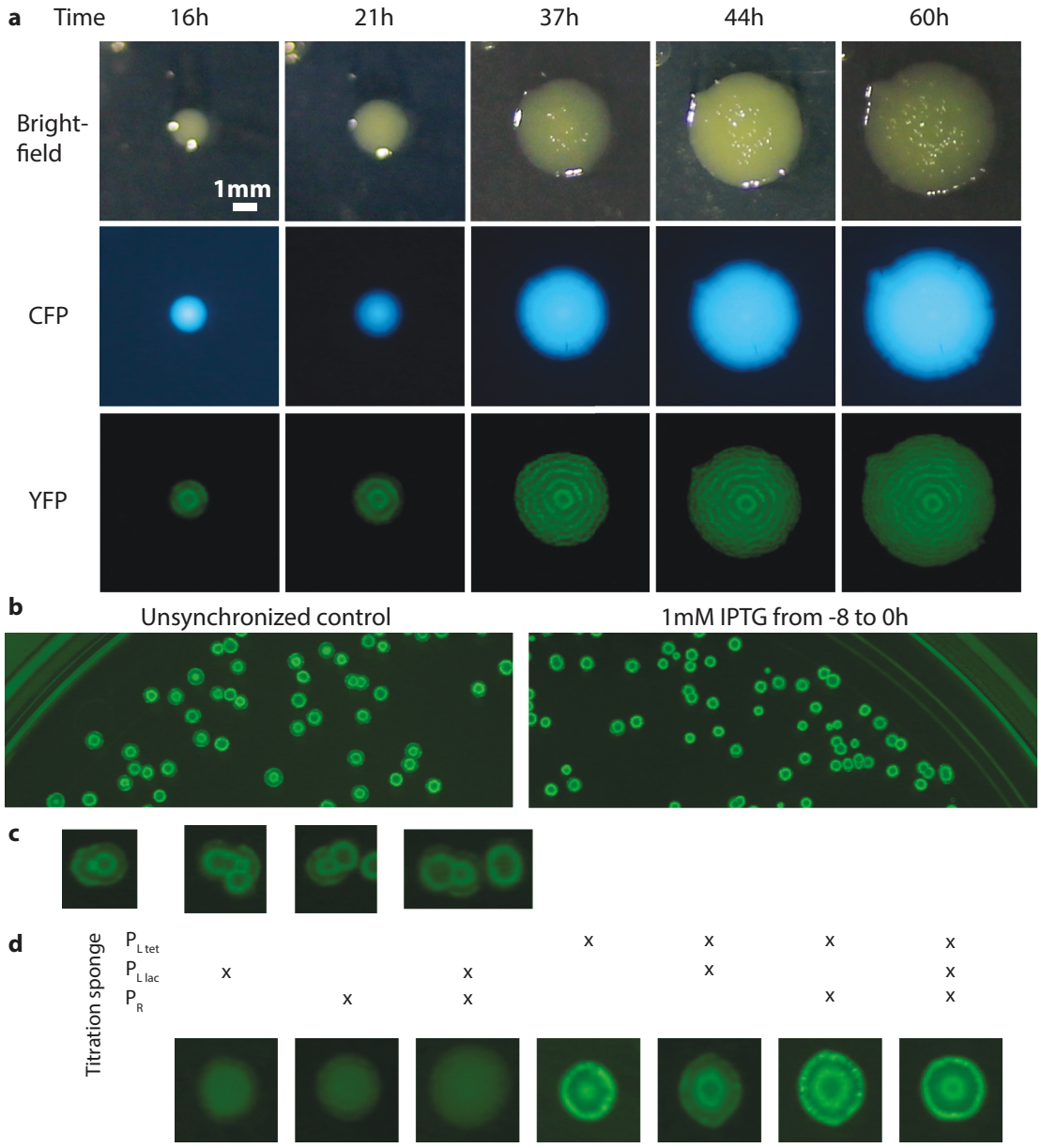
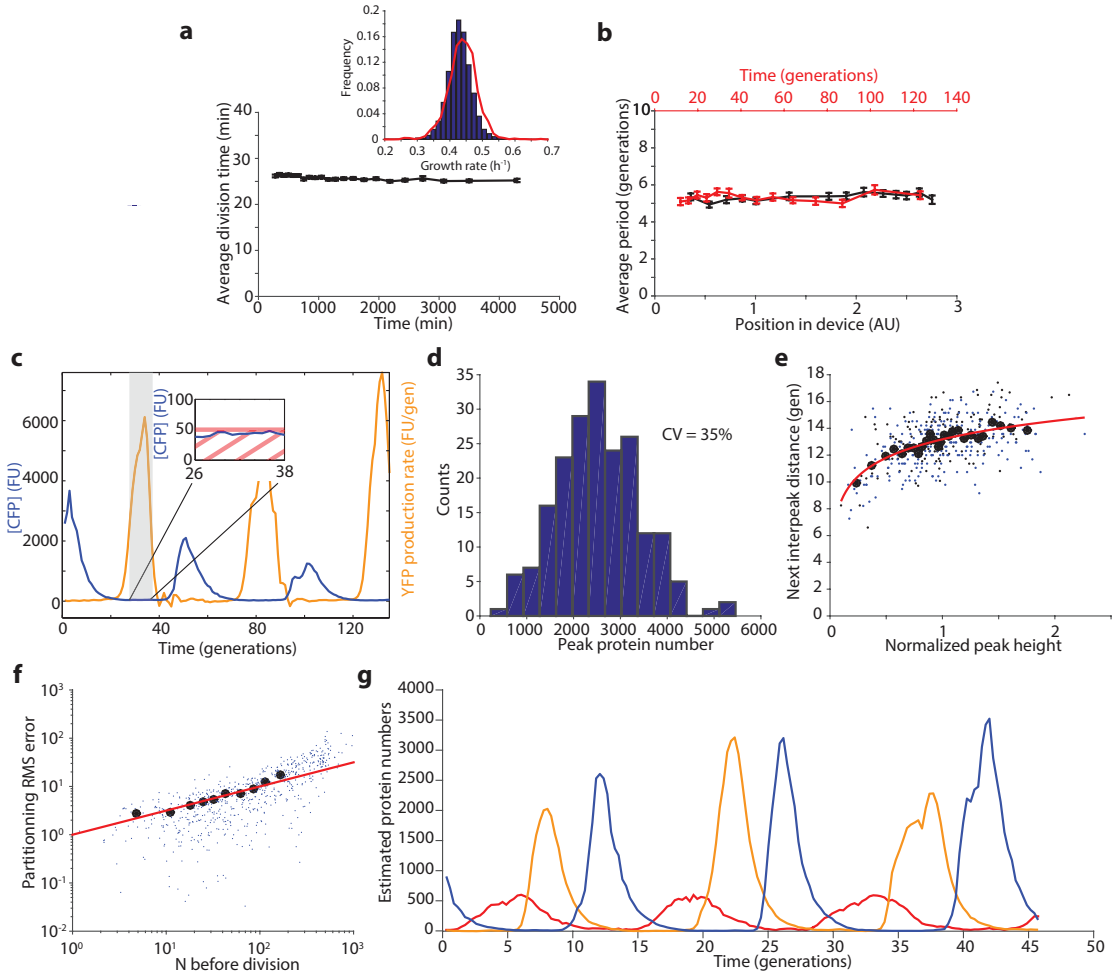


**Figure 3.12** ***(following page)*****:** Characterization of the microfluidic device and of the oscillations. a) The average division time of the integrated repressilator (LPT25) is constant over time. The inset shows the distribution of growth rates of two independent experiments (45,828 and 9,135 points are shown in the blue and red distributions, respectively), with a slight difference in the mean ($\sim$ 1%). b) The period of the oscillations is constant in space (position in field of view, distance to inlets and outlets and different media channels) and time. Each point represents a bin of 400 (a) or 100 (b) points, with the error bars indicating standard error of the mean. c) The induction/repression switch of CI (reported by YFP) occurs when the transcriptional reporter for TetR (CFP) is below the detection limit. Typical time trace of multireporter repressilator without repressor degradation and with $P_{Ltet}$-*peptide-asv* plasmid ($\Delta$*clpXP*, LPT117). The production rate of YFP is shown alongside the CFP concentration. The inset shows that the switch from induction to repression occurs below the detection limit of $\sim$ 50 FU. d) The distribution of peak amplitude of the repressilator without degradation but with titration sponge shows significant heterogeneity (LPT64, CV of 35%). e) The peak amplitude has a small influence on the next period, due to exponential dilution. The red line shows a fit to $y = 1.99 * log(x) + 13.18$ and explains 25% of the variance in the periods. Black circles are bins of 15 points of black dots (LPT64) and blue dots (LPT156). f) Estimating partitioning root mean squared (RMS) errors at cell division during the dilution phase showed that it scaled binomially, and allowed us to roughly estimate a fluorescence units to protein scaling factor. Black circles are bins of 50 blue dots (LPT64). The red line shows the fit (after conversion) to $\sqrt{\langle (n_1 - n_2)^2 \rangle} = \sqrt{N}$, where $n_i$ is the number of proteins in the daughters right after division, and $N = n_1 + n_2$ the number in the mother cell. g) Typical time trace of triple reporter repressilator without degradation with titration sponge (LPT127) in estimated protein numbers (concentration $\times$ average cell size).

**Figure 3.12:** (continued)



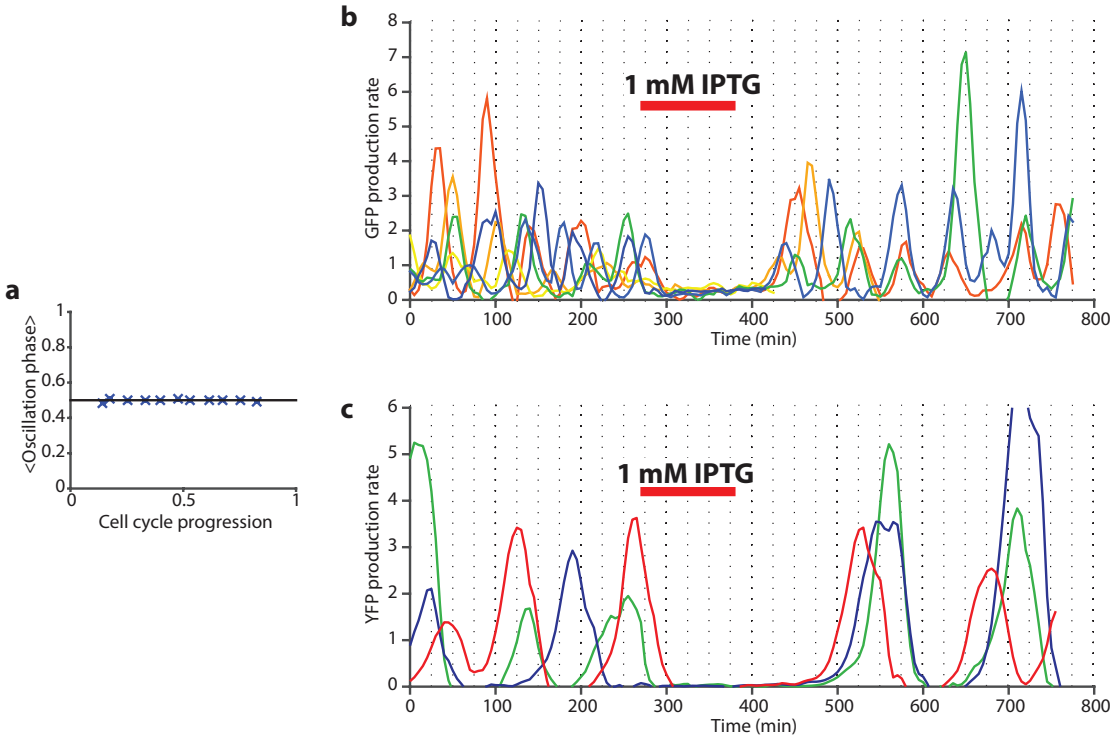
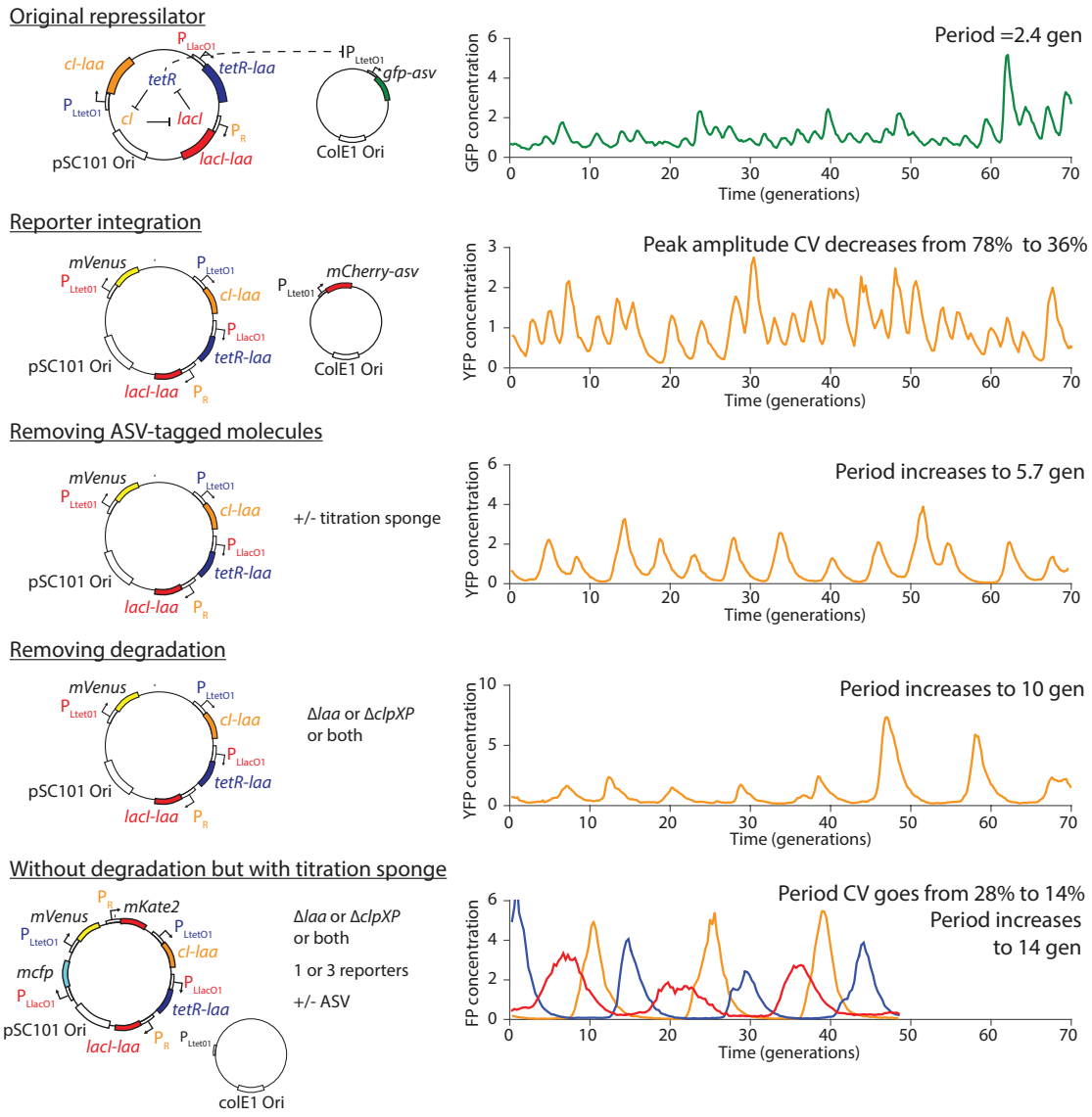



**Figure 3.13:** Robustness and synchronization of the oscillations. a) The phase of the oscillations is independent of the phase of the cell cycle. The average phase of the oscillation phase is shown as a function of the position in the cell cycle. Each point represents a bin of 3,000 data points, which have been average in *x* and *y* after being sorted on their *x* values. The error bars represent standard error on the mean and are of similar size to the symbols. Similar results were obtained for different strains, but here are shown for the integrated repressilator (LPT25). b) Synchronization of different cells in the microfluidic device was done by introducing 1mM IPTG. The original repressilator (NDL332) shows a modest level of synchrony in the oscillations of the GFP production rate. c) The integrated repressilator shows a more robust synchronization in the YFP production rates, but takes more time to recover from the pertubation.

**Figure 3.14** ***(following page):*** Schematic of the major changes to the repressilator and resulting effects on the oscillations. The original repressilator displays sustained oscillation with a period of 2.4 generations, albeit with a variable amplitude. Integrating the reporter on the pSC101 plasmid decreases the peak amplitude CV from 78% to 36%. Then, removing the presence of ASV-tagged molecules increases the period to 5.7 generations, due to the interference with degradation of the repressors in the former case. Removing degradation entirely increases the period to 10 generations, but significant amplitude and phase drift subsist. Reintroducing a sponge of binding site for the TetR repressors raises the repression threshold and enables the repressilator to exhibit precise oscillations (as well as macroscopic oscillations), by decreasing the period CV from 28% to 14% and increasing the period to 14 generations. Typical time traces are shown from top to bottom of NDL332, LPT54, LPT25, LPT61 and LPT127.

**Figure 3.14:** (continued)



## 3.5 Supplementary results and discussion

### 3.5.1 Interference of the reporter plasmid

#### ASV-tagged molecules

The results for the period change due to the introduction of ASV-tagged molecules are summarized in Extended Data Figure 3.7. The addition of substrates tagged with the native LAA tag slowed down the oscillations (compared to the titration sponge, Extended Data Figure 3.7d), as expected if it serves as a competitive substrate of ClpXP. Surprisingly, however, the ASV version of the degradation tag made the periods faster, and the effect became more pronounced the more numerous and stable the ASV tags were. The identity of the ASV-tagged substrate changed the magnitude of the effect: expressing a small peptide-ASV, repressilator-controlled mCherry-ASV or constitutive dark GFP (Y66L), reduced the period from 5.5 to 4.2, 2.6 and 2.4 generations, respectively. Overexpressing a functional ClpP-mGFPmut3 fusion [76] decreased the period slightly, but could not compensate for the presence of ASV-tagged substrates.

To confirm that the proteolysis of the natural, *ssrA*-tagged molecules was faster in the presence of ASV-tagged molecules, we expressed both in the absence of the repressilator (Extended Data Figure 3.7e). For the same production rate, the mean abundances of *ssrA*-tagged molecules was reduced four-fold in the presence of ASV-tagged molecules. This response was not observed in the $\Delta clpXP$ strain (Extended Data Figure 3.7f-h): it did not make a significant difference if the titration sponge – in the $\Delta clpXP$ strain – expressed an ASV-tagged protein or contained only the promoter. Further-

more, the response was also not observed in the repressilator that did not contain degradation tags on the repressors (Extended Data Figure 3.10f).

### Titration sponge

Addition of the titration sponge had only a minor effect on the fast and medium oscillators: the period became slightly shorter (from 5.5 to 5 generations), and the oscillations slightly better, with a phase drift of 24% per period (Extended Data Figure 3.9) and a slight increase in correlation (Extended Data Figure 3.7a). This is consistent with expectations since many proteins are not degraded, or degraded more slowly, when bound to DNA [77]. Thus in the strains without repressor degradation, the proteins and repressor sites are diluted out and replaced on the same time-scale, but when repressors are degraded, the repressors turn over after the titration sponges fills up, and the abundance of the free component becomes almost independent of the titration sites (see §3.6.3).

Trying all possible combinations of titration sponges for the different repressors (Extended Data Figure 3.11d) showed that only the sponges containing binding sites for the TetR repressor displayed macroscopic oscillations, and that titration of the other repressors did not influence greatly the oscillations. This suggests that the activity threshold is higher for these repressors, as further supported by our observations of absolute abundances and the amount of dilution required to trigger production in the next step.

### 3.5.2 Robustness of the oscillations

The phase of the repressilator oscillations were independent of the cell cycle progression (Extended Data Figure 3.13a), as expected. The period of the modified repressilator with degradation (LPT25) increased when grown at lower temperature (up to 10 generations at 25°C). This was expected since degradation of many components ceases during slow growth. The repressilator without degradation (including the titration sponge) was remarkably robust (Figure 3.3a). It oscillated at 13.8±0.2 generations at three different temperatures, as well as with conditioned media from early stationary phase culture ($OD_{600}$ ∼2), where the cell physiology was radically different. Over these conditions, the standard deviation in the average period was ∼2%. This robustness across growth conditions is in fact expected for oscillators relying only on dilution, for a combination of reasons.

First, though the total expression rates vary greatly between conditions, *E. coli* to a large extent compensate with changes in size that keep concentrations more constant. For example, under conditions where they make twice as much proteins, they also roughly have twice the volume since much of the mass is in protein form. If this applies to the repressors that form the repressilator, it means that the amplitude peaks may be twice as high, but so is the derepression threshold which is set by the concentration. The number of generations required for dilution, from the peak to the next derepresion, is then the same.

Second, the exponential nature of dilution means that there is a logarithmic dependence between the fold-reduction (from the peak to the derepression threshold) and the time that reduction takes. For example, a two-fold change in the required dilution factor only changes the dilution time with

one cell generation, which is much less than a two-fold change when oscillations are slow. Likewise, since the build-up phase approaches a steady state, the height of the peak is insensitive to further increases in the dilution time of the repressor controlling the height of that peak.

Third, protein degradation changes greatly between growth conditions (without a compensating change in concentration because so few proteins are actively degraded), but the 14-generation oscillator does not rely on protein degradation in any way.

### 3.5.3 Removing the degradation tags of the repressors

Removing the degradation tags from the repressors in cells that were wild-type for ClpXP produced similar results to having degradation-tagged repressors in the $\Delta clpXP$ strain (Extended Data Figure 3.10). The oscillations were again quite irregular, but the addition of the titration sponge again made the oscillations much more regular, and e.g. enabled the bacterial colony to exhibit ring patterns. Combining the repressilator without degradation tags on the repressors with the $\Delta clpXP$ strain also gave similar results.

### 3.5.4 Stability of the conditions in the microfluidics device

Characterization of the microfluidic device has shown that the conditions experienced by the cells were very uniform both across the device and over time. Division times were stable over multiple days, and growth distributions very similar across different experiments (Extended Data Figure 3.12a). The period of the oscillations was uniform spatially (different cells, position in the field of view, distance to the inlet/outlet, media channel) and temporally (Extended Data Figure 3.12b).

### 3.5.5 Characterization of the oscillations

#### Using binomial partitioning to estimate absolute protein abundances

To roughly estimate the average abundances of the repressor proteins in the strains without repressor degradation, we considered the statistical partitioning error at cell division of the corresponding fluorescent proteins, which are expressed under the same promoters from the same plasmid. This allows to approximately convert fluorescence units into number of proteins, as previously demonstrated [31].

Specifically, we measured the root mean squared (RMS) errors between daughters at cell division during the dilution phase of the repressilator oscillation, and confirmed that it scaled binomially with total abundance (Extended Data Figure 3.12f), as expected for fluorescent proteins diffusing freely in the cytoplasm. We then used the magnitude of the error to identify the absolute number of molecules. One caveat of this approach, in addition to the fact that the fluorescent proteins are just a proxy for the actual repressors, is that any aggregation of the proteins could increase the error but still produce binomial scaling. For example, if all proteins formed dimers and these segregated independently of each other, the errors would still be binomial but the average abundance would be off by a factor of two. Fluorescent proteins primarily multimerize when tagged to native proteins that multimerize [76] (increasing the local concentration), but since all red fluorescent proteins that we know of have severe multimerization tendencies as well as very slow maturation, including the mCherry used here, we only estimate a lower bound for the abundance of this component.

The results suggest that the YFP and CFP transcriptional reporters peaked at around 2,000-

3,000 copies (Extended Data Figure 3.12d,g) per cell, while for mCherry we can only say the levels are higher than about 500 copies, which gives a rough estimate of the peak values of the repressors. These numbers should only be seen as a very rough approximation, since some of the repressors form dimers which affects the estimated numbers, but are at least consistent with the dilution times observed, and the fact that the titration sponges had a substantial effect.

### Repression by TetR occurs at very low levels

We attempted to measure input-output relations for the fast-maturing CFP and YFP transcriptional reporters (CFP reports levels of TetR and YFP reports on CI), but not for mCherry due to the longer maturation time. We established a detection limit of at least $\sim$ 50 FU for CFP in our imaging conditions, due to combined contributions of auto-fluorescence and shot noise. As shown in Extended Data Figure 3.12c, the switch between induction and repression occurs below the detection limit of our transcriptional reporter.

Even if the repression switch occurs below the detection limits, it is possible to get a rough estimate for the location of the derepression threshold by using the protein number at the peak and the duration of the dilution phase obtained from the triple reporter analysis. This suggests a switching point of approximately 5 molecules for the oscillator without titration sponge (LPT61), and approximately 20 molecules with titration sponge (LPT64). Note that the errors in estimation of the dilution phase and the conversion of fluorescent units to molecules are exponentially amplified, and thus again this should only be considered a rough estimate.

### Peak amplitude heterogeneity and dampening

Even in the very precise oscillators, we observe significant peak amplitude heterogeneity, with a CV of about 35% (Extended Data Figure 3.12d). However, this heterogeneity is dampened by the exponential dilution, which results in a dependence of the next period in the log of the peak height (e.g. for a peak twice higher than average, the period will be only one generation longer). Using the law of total variance, the fit to a log of the peak height explains 25% of the variance in the period (Extended Data Figure 3.12e).

## 3.6 Theory

Because key properties of the single-cell kinetics – like the cooperativity or the noise in plasmid copy numbers – are unknown even for the repressilator, we use theory not to validate detailed models, but to identify potential design challenges. First we show that a simple repressilator model can behave either as a harmonic oscillator or a relaxation oscillator, depending on parameters, and that the data show that it operates in the latter regime. The criteria for stable oscillations are then very different, and we show that relaxation repressilators can display regular and sustained oscillations with minimal phase lags or cooperativity. We then analyze how noise arises in the accumulation and decay phases of the oscillations, and how titration sponges affect the dynamics. Finally, in Appendix A we derive exact limits on the accuracy of repressilator-type systems – regardless of control mechanisms – due to information loss and finite numbers of control molecules.

### 3.6.1 Harmonic versus relaxation oscillators

We start by considering a minimal model of the repressilator:

$$p_i \xrightarrow{\lambda \frac{K^n}{K^n + p_j^n}} p_i + 1$$

$$p_i \xrightarrow{p_i} p_i - 1$$

where $(i, j) = (1, 3), (2, 1), (3, 2)$, $\lambda$ is the production rate, $K$ the protein level at half-repression, $n$ the Hill coefficient and $p_i$ the number of repressor proteins. We have set the protein degradation rate constants to 1 without loss of generality, and thus express time in units of average protein lifetimes. Note that the production rates are therefore in units of protein produced per protein lifetime.

Depending on the parameters, this system can behave as a harmonic oscillator, where stability properties are captured by linearizations, or as a relaxation oscillator [73] (Extended Data Figure 3.8a) with self-sustained periodic repetitions of an aperiodic phenomenon with a period set by a relaxation time. As opposed to sinusoidal/harmonic oscillators, relaxation oscillators exhibit discontinuous jumps and all-or-none behavior. This well describes the repressilator dynamics we observe: as shown in the three-color analysis, the production of a repressor starts abruptly and remains constant until it suddenly turns off. In fact, even the seemingly gradual reduction in expression before the turn-off observed for the 14 generation oscillator is a part of the relaxation path rather than a result of control exerted: measuring the production rate and dilution rate shows that the concentration only starts to level off because dilution effects become increasingly important at high numbers.

The strongly non-sinusoidal shape thus comes from thresholding, where promoters spend most of their time in either a completely on-state or completely off-state. The mathematical requirements for oscillations are then very different. Classical stability analysis for harmonic oscillators emphasizes the importance of phase lags and high cooperativity, but in the relaxation regime it is possible to oscillate without much of either – in fact for extreme parameters, stochastic models show that oscillations are possible without any phase lag or cooperativity (Extended Data Figure 3.8d). Due to the clear separation of the two phases – a build-up phase with constant production and a pure decay phase with no production (Extended Data Figure 3.8c) – we next analyze how much noise arises in each phase, with a particular focus on noise in timing rather than in amplitude. The period can either be seen as the sum of three decay phases, or as the sum of three production phases, since the beginning of a production phase of one repressor marks the beginning of a decay phase of another. We find that a key architectural feature of the repressilator – in the observed relaxation mode – is that variation in the build-up occurs in a way has little effect on the overall timing, while timing noise from the decay phase is hard to avoid given the low promoter derepression thresholds observed. Specifically we iterate some known results about noise in gene expression to show how removing degradation tags reduces the noise in the starting value of the decay phase, and then analyze the inherent variation in the decay phase as a function of the starting value and the derepression threshold. Finally we consider the propagation of noise between the build-up and decay phases, i.e., how variation in peak amplitude affect the decay phase and how variation in the decay phase affects the build-up phase.

### 3.6.2 Heterogeneity in the accumulation and decay phases

#### Accumulation phase

Motivated by the experimental observation that the oscillation cycle has a build-up phase with almost constant production rate, we consider the variation in the number of proteins made during a time window $T_a$ during which gene expression is unrepressed. Most of the results presented here are well known from the extensive literature on stochastic gene expression and are only briefly repeated here for convenience.

First, consider a process in which proteins are made constitutively in bursts of size $b$:

$$p_i \xrightarrow{\lambda} p_i + b$$

The CV (standard deviation over mean) in the number of proteins produced is then $1/\sqrt{\lambda T_a}$, and even if we further allow variation in $b$, the CV is still inversely proportional to the square root of $\lambda T_a$. Statistical accuracy can thus be increased either by increasing the production rate, or by integrating over longer time periods by increasing $T_a$. The latter can be achieved by eliminating repressor degradation and thereby extending the periods.

If the build-up phase is short, so that dilution/degradation is negligible, the number of proteins produced is similar to the peak amplitude (the troughs of the oscillations are virtually at zero so proteins from the previous build-up phase contribute very little). However, for long enough times $T_a$, decay and dilution eventually become non-negligible, and the system approaches a steady-state

(this was directly observed in the data for the slowest oscillator but is likely the case also for the faster oscillator where the actively degraded repressors should approach a steady state much faster than the stable reporter observed). Increasing $T_a$ further thus does not reduce the CV in peak amplitude. Specifically we consider:

$$p_i \xrightarrow{\lambda \mathbf{z}} p_i + b$$

$$p_i \xrightarrow{p_i/\tau_p} p_i - 1$$

where $\mathbf{z}$ is some extrinsic noise variable, $b$ is a geometrically distributed burst size and $\tau_p$ the lifetime of the protein. For simplicity, if we assume $\mathbf{z}$ is following its own birth-death Poisson process, then at the end of the build-up phase [1, 9, 69]:

$$CV_p^2 = (1 + \langle b \rangle) \frac{1}{\langle p_i \rangle} + \frac{1}{\langle z \rangle} \frac{\tau_z}{\tau_p + \tau_z}$$

where $\tau_z$ is the lifetime of $z$. Therefore, increasing the proteins' lifetimes still reduces noise by increasing the average protein abundance and time-averaging over extrinsic influences.

### Decay phase

Motivated by the experimental observation that the decrease in repressor concentration is indistinguishable from exponential decay, we here model the variation expected in the time until the repressor reaches low enough concentrations that expression of the next component in the cycle is derepressed. Specifically, we are interested in the average and standard deviation of the decay time

$T_d$, from an initial peak of $N$ to a threshold $S$. This time contributes one third of the period, and determines the build-up time of another repressor in the system. We ignore two biological complications. First, for systems with actual degradation of proteins, fluctuations in the proteolysis machinery could produce fluctuating degradation rate constants, i.e., essentially contributing an extrinsic source of noise. However, this noise is unknown, and our primary purpose with this theory is to explain the effects observed for the circuits without active degradation, when molecules are diluted out during the cell cycle and then randomly partitioned between daughters at cell division. For binomial partitioning (as we directly observe, see section 3.5.5), this can be modeled by explicitly accounting for cell growth and division, as we and others have done in other studies [69]. However, because our goal here is to produce insights into experimental design rather than to capture the exact dynamics, we use a common approximation where binomial segregation and dilution are approximated with first order decay, i.e., where the protein decays in individual random steps:

$$p_i \xrightarrow{p_i} p_i - 1$$

Here we again have set the degradation rate constant to one, to express time in units of average protein lifetimes.

Each step is thus probabilistic with an exponentially distributed waiting time, and the total time to decay from $N$ to $S$ molecules is a sum of the random variables for the individual steps. For very high thresholds, where $S$ is almost as high as $N$, the standard deviation in the total time is substantial compared to the average total time, since few steps are summed. For very low thresholds there

are large numbers of steps, but the last few steps are rate limiting and dominate the sum. For example, the average waiting time for going from $N$ to $N-1$ molecules is $N$ times shorter than the average waiting time for going from 1 to 0 molecules. If the derepression threshold is reached when the last molecule is eliminated ($S$=0), the total variance in the time equals $\sum_{i=1}^{N} 1/i^2$, which is 1 for $N = 1$ and quickly approaches $\pi^2/6$ for higher $N$. The average time is in turn $\sum_{i=1}^{N} 1/i$, which is approximately $\log(N) + \gamma$ where $\gamma = 0.57721\ldots$ is Euler's constant. More generally, for any threshold $S < N$, the variance is $\sum_{k=S+1}^{N} \frac{1}{k^2}$, which approaches $\frac{1}{S+1} - \frac{1}{N}$ for large $N$ and fixed $S/N$[2] (which in turn is approximately $1/(S+1)$ if, in addition, the ratio $N/S$ is sufficiently large[3].) The average follows $\sum_{k=S+1}^{N} \frac{1}{k} \approx \log(N/(S+1))$ under the same conditions. Thus, for a thresholding scheme of this type

$$CV \approx \frac{\sqrt{1/(S+1)}}{\log(N/(S+1))}$$

If $N/S$ is large the CV increases very slowly with $N$, because the logarithm becomes insensitive to changes in $N/S$. For example, doubling $N/S$ from $1,000$ to $2,000$ only increases $\log(N/S)$ by about 10%. For any given $N$, there is also an optimal threshold that minimizes the CV at $(S+1)/N = 1/e^2 = 0.13534$, giving a mean of 2 and a variance of $e^2/N$. Therefore:

$$CV \approx \frac{\sqrt{N/(S+1)}}{\log(N/(S+1))} \frac{1}{\sqrt{N}} \geq \frac{e/2}{\sqrt{N}} = \frac{1.359}{\sqrt{N}}$$

Extended Data Figure 3.8e in turn shows the exact CV as a function of $N$ for various values of $S$,

[2] specifically $N \sum_{k=S+1}^{N} \frac{1}{k^2} \to \left(\frac{N}{S+1} - 1\right)$ as $N \to \infty$.

[3] Note that we sometimes use the approximate $N/S$ ratio (instead of $N/(S+1)$) for simplicity as it is good approximation for the $S$ and $N$ values considered.

which closely follows these approximations since $N \gg 1$ throughout. For any fixed $S$, the decrease in CV with $N$ is thus logarithmic, and it is necessary to increase $S$ along with $N$ in order to reduce the CV effectively. However, the CV remains close to its minimum for a rather wide range of $S/N$ values (e.g. between 0.05 and 0.3, see Extended Data Figure 3.8e). Thus if $S$ is increased along with $N$, it should be possible to decrease the CV by the inverse square root of $N$. These simple theoretical findings could directly explain our experimental observations. The estimate of the TetR threshold suggests that it was lower than optimal even before eliminating repressor degradation. Thus $N/S$ was already very high and we should only expect a very minor reduction in the heterogeneity of the timing, as we indeed observed. However, when the threshold is also increased (Figure 3.2b) by adding a titration sponge, the variation in the timing should be more drastically reduced by the increased $N$, as we also observed.

### Transmission of noise between phases

Fluctuations in the decay time will also affect and be affected by fluctuations in amplitude. Here we show how this noise propagation to some extent is naturally damped by the mechanisms of the system. First, a doubling in the peak amplitude will only result in an increase of one generation of the decay phase, which is small given that the oscillator has a period of 14 generations and each repressor thus spends 4-5 generations in the decay phase. More generally, the time it takes to exponentially dilute out levels from $N$ molecules to $S$ molecules on average approximately equals $\log(N/(S+1))$. The additional variance in the decay period due to the variation in peak amplitude can be estimated using the rule for error propagation for functions of random variables (a small deviations approach),

which immediately shows that for $\log(N/(S+1))$ the additional variance in timing is given by $\sigma_N^2/\langle N\rangle^2$, which equals $1/\langle N\rangle$ if the initial number of molecules is Poisson distributed. If the repressor molecules are produced in geometrically distributed bursts of size $b$ this variance increases to $(1+\langle b\rangle)/\langle N\rangle$. In either case, as long as $N/S$ is reasonably high this term will be small in comparison to $\langle 1/(S+1)-1/N\rangle$ [4], the expected variance in the decay steps, and so little noise is transferred from the peaks to the periods, as we also observe experimentally (Extended Data Figure 3.12e).

Noise is also transmitted from the decay phase to the accumulation phase, as the decay time $T_d$ of one repressor determines the accumulation time $T_a$ of another. In the parameter regime in which the repressor in question does not have time to approach a steady-state before production is turned off again, significant noise is propagated because a doubling in the accumulation time will result approximately in a doubling of the amplitude peak. Even in this regime, little noise propagates back to affect the next decay phase due to the logarithmic principle above. Furthermore, in a regime where the repressor reaches a (quasi-)steady-state before production stops, little noise is transmitted since fluctuations in the accumulation time then only change the time the repressor spends around its steady-state value.

### 3.6.3 The role of the titration sponge

The theoretical analysis above shows that low repression thresholds $S$ create large noise in the decay phase of the oscillation even for perfect threshold mechanisms. This seems to explain why titration sponges improve the regularity of oscillations so much. In the next section we include titration in

[4] specifically $N/(S+1) >> 1+b$. If the first term is 10 times larger, there will then be a 10% contribution to the variance.

a simulation model, and here we first discuss non-trivial aspects of titration such as the effect of fluctuating thresholds, changes in the effective cooperativity, and the impact of titration when the repressors are degraded.

The perhaps most obvious complication is that the titration sponge not only raises the threshold but also randomizes it because it is encoded on a plasmid that fluctuates in single cells. The effect of such fluctuations is included in the simulation model where we also account for the fact that only some of the sites will have repressor bound at any given time. However, for approximate but simpler results consider a situation where the threshold $S$ varies around an average $\langle S \rangle$, and fluctuations are so slow in time that $S$ can be considered to be fixed during each derepression window. This creates additional variance in the decay period due to variation in the conditional average (in the sense of the Law of Total Variance). As for the discussion of uncertainty in $N$ in the previous section, the rule for error propagation for functions of random variables immediately shows that for $\log(N/(S+1))$ the additional variance in timing is approximately given by $\sigma_S^2/\langle S \rangle^2$, which equals $1/\langle S \rangle$ if the threshold is Poisson distributed. Fluctuations in $S$ then contribute about as much to the variance in timing as the inherent randomness of the decay process (this is again a small noise approximation since the 'intrinsic' noise due to random decay events is $\langle 1/(S+1) \rangle$ but not $1/\langle S \rangle$). The experimentally observed improvement of the oscillations with the addition of the titration sponge still makes sense though, because it appears to raise the threshold by an order of magnitude. Intuitively, it may randomize the threshold to some extent, but it also avoids the extreme variation that comes from low thresholds, where the last few decay events greatly randomize control.

The experimental data also shows that the titration sponge only has a small effect when the re-

pressors are degraded (§3.5.1). This is also easily explained by the model above. In fact, none of the conditions for having an effective titration sponge seem satisfied when the repressors are degraded. First, we showed above that there is an optimal threshold $S$ for every peak abundance $N$ that minimizes the relative standard deviation in the decay window. With repressor degradation, $N$ is much lower because of the short repressor half-lives. Based on the numbers observed ($N \approx 600$), we believe that $S$ is still too low to be optimal even under those conditions, but raising $S$ only has a large effect on the noise if $S$ is far below the optimum for a given $N$ (Extended Data Figure 3.8e). Second, it is unlikely that repressors are effectively titrated out when the repressors are tagged for degradation. Many proteins appear not to be degraded, or to be degraded much more slowly, when bound to DNA (see for example Ref. [77]). For example imagine a threshold of 50. If there are 51 repressor molecules present but only one available for degradation because the others are titrated out, then the noise in this step is the same as if there were no titration and just one repressor molecule present. Another way of looking at this is that if the repressors are degraded, taking out some of them by titration makes little difference, because they quickly re-adjust to steady state. The fact that both repressor molecules and binding sites are diluted at the same rate is key to ensure effective titration. Finally, for repressors that are actively degraded, the noise in the time needed to decay below some threshold concentration may not be dominated by the last few reaction steps, as in the model above, but by 'extrinsic' noise from fluctuating proteases. Raising the threshold to eliminate the last few steps may then have a marginal impact.

As described above, given the experimental observations of the low TetR threshold, in this context the main effect of the TetR titration sponge seems to be to simply increase the threshold. Most

previous models have instead emphasized that titration (and other schemes for zero-order elimination of molecules) create ultra-sensitive switches similar to increasing the effective cooperativity (see e.g. references [78–82] ). It is possible that this also helps in this system, but several system-specific features suggest it may not be the case. First, relaxation oscillators where the independent relaxation path is long compared to the time-window for switching are less sensitive to cooperativity – even if the decision to repress was gradual, it is still brief compared to the rest of the dynamics. Indeed, we see that high cooperativity is not needed for the stochastic models to oscillate with high regularity (Extended Data Figure 3.8d). Second, we introduced the titration sites on plasmids that fluctuate in copy numbers, and because the binding sites are no different from the regular promoters, not all sites will be filled at the time of the derepression. Thus the titration remove a variable number of molecules, and may not increase the effective cooperativity much.

### Simple computational model

We will now show that a simple computational model combining these features can not merely oscillate, but reproduce the amplitude, period and shape of the peaks for the estimated abundances of proteins and plasmids. In fact, we found that many versions of the model and parameter combinations can reproduce the data, and thus our purpose is not to promote one such model over others but merely to demonstrate that the types of models we have in mind are not inconsistent with the key known facts. For example, the absence of definitive data on e.g. binding affinities, repressor dimerization, and repression kinetics, prevents us from usefully modeling the switches in terms of elementary reactions. However, to make the models more realistic than perfectly sharp thresholds we

assume Hill functions with reasonably small Hill coefficients (i.e. $\sim 1.5 - 2$). We further account for translational bursting and for fluctuations in both the repressilator plasmid and the plasmid for the titration sponge, to show that reasonably accurate oscillations (e.g. correlation coefficient after one period > ∼0.2) are possible despite these sources of randomness. Specifically we assume

$$p_{\text{Tot,i}} \xrightarrow{\lambda N_{\text{o}} \frac{K^n}{K^n + p_{\text{Free,j}}^n}} p_{\text{Tot,i}} + b$$

$$p_{\text{Tot,i}} \xrightarrow{p_{\text{Tot,i}}} p_{\text{Tot,i}} - 1$$

where $(i, j) = (1, 3), (2, 1), (3, 2)$. $N_{\text{o}}$ is the number of promoters, $p_{\text{Free,j}}$ is the number of unbound repressor molecules, $p_{\text{Tot,j}}$ is the total number of repressor molecules and $b$ is a geometrically distributed random variable with mean $\langle b \rangle$, representing the number of proteins produced from each mRNA, where the mRNA lifetime is assumed to be short compared to the protein lifetime, which is the case for the oscillators with a $\sim 14$ generation period. For the number of titration sponge plasmids $N_{\text{t}}$ and the number of repressilator plasmids $N_{\text{o}}$, we simply assume

$$N_{\text{o}} \xrightarrow{\langle N_{\text{o}} \rangle} N_{\text{o}} + 1, \quad N_{\text{o}} \xrightarrow{N_{\text{o}}} N_{\text{o}} - 1$$

$$N_{\text{t}} \xrightarrow{\langle N_{\text{t}} \rangle} N_{\text{t}} + 1, \quad N_{\text{t}} \xrightarrow{N_{\text{t}}} N_{\text{t}} - 1$$

This is in fact not so different from some specific plasmid models, where self-replication and non-cooperative negative feedback combine into a constant production rate. The decay of plasmids again

approximates the dilution and segregation, and thus has the same 'degradation' rate constant of 1 as the proteins.

Titration effects are harder to model due to several complications. For example, simply removing $N_{\mathrm{t}}$ copies from the pool of free proteins (which we found was enough to reproduce key experimental observations) does not account for the fact that repressor binding is not complete, or for the fact that the actual binding sites on the repressilator plasmid are of the same type as on the decoy plasmid. Since we do not know the binding details and only use the Hill function as a generic way to capture nonlinear effects, there is no straightforward way to model titration through elementary reactions. However, motivated by the fact that the dynamics we model are very slow, we assume that the number of proteins bound to promoter sites is in quasi-equilibrium due to fast binding and unbinding. The fraction of promoters which are unbound is $\frac{K^n}{p_{\mathrm{Free,j}}^n+K^n}$. The number of bound repressor molecules will then be proportional to

$$N_{pro}\frac{p_{\mathrm{Free,j}}^n}{p_{\mathrm{Free,j}}^n+K^n}$$

where $N_{pro}$ is the total number of promoters for the respective repressors

$$N_{pro}=\begin{cases} N_{\mathrm{o}}+N_{\mathrm{t}}, & j=1 \\ N_{\mathrm{o}}, & j=2,3 \end{cases}$$

where TetR is repressor number 1. For integer $n$ this is consistent with modeling a single elementary

reaction of $n$ repressors binding with the promoter in a reaction that comes to equilibrium quickly, where the total number of repressor molecules would then be given as $p_{\text{Free,j}} + nN_{pro}\frac{p^n_{\text{Free,j}}}{p^n_{\text{Free,j}}+K^n}$. However, since we use the Hill function as a generic way to account for cooperativity we instead consider

$$p_{\text{Tot,j}} = p_{\text{Free,j}} + 2N_{pro}\frac{p^n_{\text{Free,j}}}{p^n_{\text{Free,j}} + K^n}$$

where the factor of two comes from the two binding sites per promoter. This equation is then solved at each step of the Gillespie algorithm [83] to update the free number of proteins. This is an approximate scheme, but one that accounts for several of the important complications, and we find that many ways of accounting for titration can explain the data.

For the simulation of Extended Data Figure 3.8f, the parameters used are $\lambda = 60$, $K = (5, 10, 10)$ for the three repressors, $n = 1.5$ for all repressors, $\langle b \rangle = 10$, $\langle N_{\text{o}} \rangle = 10$ $\langle N_{\text{t}} \rangle = 40$. The mean period for this model is approximately 9 protein lifetimes, or 13 generations, and the CV of the cycle to cycle variation in period (e.g. phase drift) is approximately $10\%$, which is similar to the experimental data. Extended Data Figure 3.8d in turn shows that oscillations are possible as described above, without bursts, biochemical cooperativity ($n$=1) titration or phase lag contribution from mRNAs (we consider infinitely fast mRNA dynamics). Deterministic linear theory would require a Hill coefficient of over 3 in each component for oscillations in this case.

*During the time that [Karl] Landsteiner gave me an education in the field of immunology, I discovered that he and I were thinking about the serologic problem in very different ways. He would ask, What do these experiments force us to believe about the nature of the world? I would ask, What is the most simple and general picture of the world that we can formulate that is not ruled out by these experiments? I realized that medical and biological investigators were not attacking their problems the same way that theoretical physicists do, the way I had been in the habit of doing.*

Linus Pauling

# 4

# Protein degradation kinetics in single bacterial cells

In the previous chapter, we saw the importance of protein degradation on the repressilator. It drastically changed the period of the oscillations from 2.5 to 14 generations and it prevented the most accurate oscillations by lowering the number of molecules. Some effects of degradation were also

counter-intuitive: the ASV-tagged reporter for the oscillations acted as an *anti-competitor* for degradation, accelerating rather than slowing the oscillations.

We saw in chapter 2 that proteins cannot respond to changes faster than their half-life. Protein degradation is therefore necessary for response faster than one generation, which could be vital if cells need to respond quickly to a changing environment. The main objective of this chapter is to develop a tool to measure degradation rates in single cells. This tool would allow us to quantify the impact of fluctuations in degradation rates on the cells.

We will start by briefly introducing proteolysis and its the main actors in *E. coli*. We will then explore the counter-intuitive effects of the reporter in the repressilator, which will further motivate the development of a tool for measuring degradation rates in single cells. Our approach will then be described and validated. A comparison with the typical pulse-decay approach will showcase the limitations of that particular approach. This will lead us to describe the issues associated with using fluorescent proteins as reporters for degradation.

We will then characterize what we term a degradation response: a high concentration of substrate targeted for degradation can induce production of proteases. We will show its association with the heat shock response as well as pleiotropic effects such as an increase in intracellular ATP concentration. Finally, the interference effects due to transcriptional reporters as well as the toxicity of our particular construct will be discussed.

In this chapter, we will also describe "negative results", as we believe the dissemination of these results is important for scientific progress. Indeed, a better access to such results would have greatly accelerated the progression of this project.

## Contributions

The work presented in this chapter was performed in full collaboration with Scott Luro.

## 4.1 Introduction

In *E. coli*, most proteins are not actively degraded but instead diluted out as cells grow and divide [43]. This limits how quickly cells can approach lower steady state protein concentrations, since even if production is entirely turned off the concentration can only be halved each doubling (as seen in chapter 2). Similar principles in fact apply when cells adjust to higher steady states: for many kinetic schemes the time it takes to move e.g. half the distance to a new steady state is set entirely by the half-life of the component and not by the production rate. For dynamics faster than the cell cycle, proteins thus need active proteolysis. Indeed, many stress response pathways in *E. coli*, where rapid responses may be crucial, are regulated via proteolysis, such as the heat shock response (RpoH [84]), the envelope stress response (RpoE [85]) and the general stress response (RpoS [86]).

Protein degradation is performed by a variety of proteases, chaperones and unfoldases. These proteins are also part of the protein quality control network (PQC), which is reponsible for the refolding and/or degradation of misfolded proteins, aggregated proteins and damaged proteins. This network is well conserved across bacteria. In fact, the ClpXP protease can even been found in human mitochondria [87]. Proteases have also been recently suggested as a novel target for antibiotics, which could potentially be very useful to kill even non-growing cells which are recalcitrant to many treatment. Specifically, new molecules (e.g. acyl depsipeptides) have been found to activate

the peptidase ClpP for unregulated degradation of peptides and unfolded proteins, making the cells effectively degrade themselves [88–92].

In the following section, we will briefly introduce the main players of protein degradation in *E. coli*, with a particular focus on ClpXP and the *ssrA* tag. For an in-depth review, see reference [93].

### 4.1.1 Proteolysis in *E. coli*: an overview

In *E. coli*, there are five different AAA+ (ATPase associated with diverse cellular activities) proteases: ClpAP, ClpXP, HslUV, Lon and FtsH. The proteases are composed of two parts, an unfoldase and a peptidase. We will start by describing ClpXP, arguably the most studied bacterial protease [94], as a model for proteolysis in bacteria.

ClpX is the unfoldase and forms hexameric ring shape structure that align concentrically with the ClpP peptidase, which in turn forms a barrel structure of two stacked heptameric rings (i.e. a 14mer, Figure 4.1). The $ClpX_6$ ring can bind either or both ends of the barrel, forming single- (1:1) or double-headed complexes (2:1), although the two heads appear to operate independently [95]. Degradation proceeds when the unfoldase recognizes a degradation tag, called the degron, which is typically an amino acid sequence located at the N- or C-terminus of the proteins (although some reside within the protein). These tags can be either added to the proteins during translation (e.g. the *ssrA* tag [67]), genomically encoded on the protein sequence (e.g. Crl [40]]) or exposed by protein processing (e.g. N-end rule [96]) or changes in the protein structure (e.g. RpoH [97, 98]). The unfoldase then unfolds the substrate and translocates it through the peptidase, where the peptide is hydrolyzed into small fragments of 3-8 amino acids (Figure 4.1). Unfoldases therefore provide

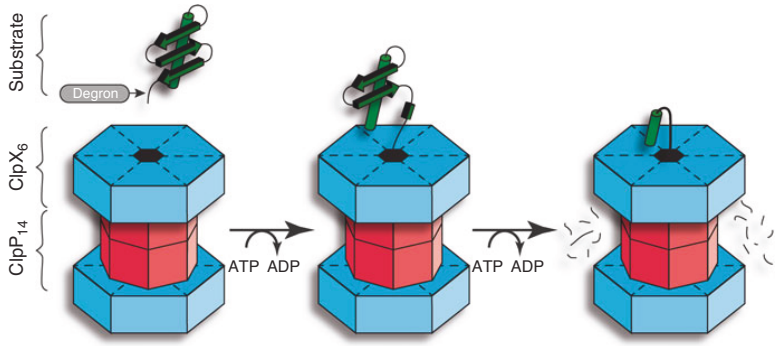



**Figure 4.1:** Cartoon schematic of degradation by the protease ClpXP. The substrate, containing a degradation tag (degron), is recognized by the unfoldase (ClpX). It is then unfolded by pulling on its recognition sequence and translocated through the peptidase (ClpP), processively unravelling the substrate. The peptidase cuts the peptides in fragments 3-8 amino acids long. Figure reproduced with permission from reference [93]

substrate specificity for the associated peptidase.

### The *ssrA* tag and trans-translation

Messenger RNAs that lack a stop codon, due to truncations or errors in transcriptions, cause the ribosome to stall on the 3' end, as it cannot dissociate without the stop codon. Because of the vital role of ribosomes for growth, bacteria have evolved different mechanisms for ribosome rescue [99]. The main pathway involves the *ssrA* RNA [67, 100–102], which has been termed a tmRNA because of its function as both a tRNA and an mRNA [67, 103]. It is charged with an alanine while the rest of the RNA encodes for a short peptide sequence followed by a stop codon. The tmRNA can then dock to the empty A-site of the stalled ribosome, resume translation of the *ssrA* mRNA and terminate translation, releasing the ribosome. It has been estimated that about 1 protein for every 200 translated proteins is naturally *ssrA*-tagged [100, 102] – emphasizing the importance of this mechanism.

The *ssrA* tag, inserted during translation of damaged mRNAs, target the proteins for rapid degradation via the added peptide sequence (AANDENYALAA), because of the potential toxicity of these aberrant proteins. In *E. coli*, *ssrA*-tagged proteins are recognized by ClpXP, ClpAP, Lon and FtsH [102, 104], although more than 90% of the degradation is mediated by ClpXP [102]. Because this tag can be added to any protein, it has been used extensively as a model for unfolding and adapter-mediated substrate delivery [105–108].

### Protease specificity and adapters

Roughly 100 genomic targets for ClpXP have been identified by using a mutant ClpP lacking the peptidase activity ($ClpP^{trap}$ [40, 109]). These substrates include transcriptional regulators (Crl, LexA, RpoS), proteins involved in translation (RplJ, TufB), protection against stress (Dps), cell division, motility, metabolism, chaperones and proteases. Recognition motifs have been proposed at both the N- and C- termini for recognition by ClpXP [40].

Unfoldase specificity can be influenced by proteins termed adapters. The adapter SspB docks on ClpX and greatly increase the specificity of the complex to *ssrA*-tagged proteins [110] (and the anti-sigma factor RseA [85]). While ClpXP can still recognize and degrade *ssrA*-tagged substrate efficiently without SspB [110], other adapters are essential for recognition, such as RssB (also known as SprE) for degradation of RpoS by ClpXP [86, 111, 112].

ClpP can also bind to ClpA to form the ClpAP protease (or even double-headed ClpAPX hybrid [113]). The substrates recognized by ClpAP have been less well characterized than the ClpXP substrates. Interestingly, ClpAP can degrade ClpA, albeit with lower affinity than the *ssrA* tag [114,

115], suggesting feedback control. ClpAP recognition of N-end rule substrates – a highly conserved pathway originally found in eukaryotes – is greatly increased with the adapter ClpS [116], though also possible without the adapter [117]). These substrates require pre-processing in order to expose their recognition sequence. For example, putrescine aminotransferase (PATase) is modified post-translationally by leucyl/phenylalanyltRNA-protein transferase (LFTR) in order to be targeted for degradation [96]. The substrate Dps can be cleaved by the methionine aminopeptidase (MetAP) to form $Dps_{2-167}$, recognized by ClpXP [40], or by an unknown endopeptidase to form $Dps_{6-167}$, recognized by ClpAPS [96, 118].

The Lon protease has a similar structure to ClpXP, except that the unfoldase and the peptidase are encoded on the same gene and forms hexamers. Lon is the main protease responsible for protein quality control by degrading misfolded proteins [119, 120]. Lon recognizes misfolded proteins using short stretch of amino acids rich in aromatic residues normally hidden in folded proteins [121]. Lon can also recognize native proteins where the degron is genomically encoded, such as the cell division inhibitor SulA [122] (suppressor of *lon*).

FtsH is an evolutionarily conserved inner-membrane anchored protease. It can degrade unstructured substrates as well as folded proteins. It plays an important role in heat-shock by degrading the master regulator of heat-shock response, RpoH [84, 123, 124]. Interestingly, FtsH cannot degrade RpoH in vitro, probably due to its weak unfoldase activity [125]. It has been proposed that the chaperones DnaK-DnaJ-GrpE and GroEL/ES bind and unfold RpoH so that it can be processed by FtsH [97, 98]. DnaK binds unfolded proteins and therefore acts as a sensor: during heat shock, a high number of misfolded proteins are present, titrating DnaK away from RpoH and stabilizing the

sigma factor [97, 126]. Although multiple proteases contribute to the degradation of RpoH, FtsH is thought to be the main contributor [84, 123, 124]. It also plays an important role in the quality control of membrane proteins and the regulation of lipid synthesis. FtsH is the only essential protease in *E. coli*, although it can be rescued by a point mutation in the *fabZ* gene to restore the balance of lipids [127]. An FtsH$^{trap}$ approach has been used to identify potential FtsH substrates [128]. Few putative substrates were confirmed in vitro, which might be due to the weak unfoldase activity and the need for co-factors for degradation.

### Partial degradation

When a protease engages a domain resilient to degradation, it can release its substrates. This phenomenon has been termed partial degradation, or substrate processing, as the protein now lacks a degradation tag and is no longer targeted for degradation. Proteases possess different unfoldase strengths and are expected to yield different amount of partial degradation (see Table 4.1). It has been shown that the local structure (i.e. where the protease is pulling) is important for partial degradation, not the overall stability of the protein[1] [107, 130]. This substrate processing can be used by the cell: the ATP-binding clamp loader subunit DnaX in *Caulobacter crescentus* is partially degraded by ClpXP to form different stable fragments of different physiological function [131]. Partial degradation is further enhanced when the stable domain is preceded by a glycine rich region, although the exact requirements for this sequence remain to be elucidated [132–134]. The glycine rich region impairs the unfolding properties of the protease [135]. Partial degradation has also been observed in

[1] It has been argued that it is the interaction between the specific protease and the local structure; however, the evidence is not convincing [129]

**Table 4.1:** Processivity (complete/partial degradation ratio) of different proteases degrading *E. coli* DHFR through the C-terminus. Note that this data represents processivity of one particular domain, and it is included as an approximate guide to compare unfoldase strength. Data from reference [137].

| Protease | Processivity |
|---|---|
| 26S proteasome | >9 |
| ClpAP | 1.25 |
| ClpXP | 0.4 |
| HslUV | 0.1 |
| Lon | <0.05 |
| FtsH | 0 |

vitro, for example by targeting titin-I27 to ClpXP [108] or GFP to FtsH [136].

### 4.1.2 Current methods for measuring proteolysis

#### Pulse-decay

Most in vivo measurements use some variation on the broader theme of the pulse-decay method. The idea is to pulse the production of the substrate (e.g. with an inducible promoters), label the substrate (e.g. with radio-labeled amino acids) and chase to watch the decay by running a SDS-PAGE at multiple time points. Although this method typically does not allow measuring the full kinetics, it has been used to determine substrate half-lives (though half-lives are less relevant if the decay is not exponential). This can also be done with fluorescent proteins, with the decay measured by the decrease in fluorescence signal.

One study in particular has used this method to measure degradation rates in single cells of *ssrA*-tagged GFP [81]. The authors induced substrate production in bulk cultures, washed the cells, treated with chloramphenicol to stop protein production and transferred to agar pads for imaging

the decay. Degradation was reported to follow zero-order kinetics, i.e., the enzymes were shown to be completely saturated by the substrate of interest, eliminating a constant amount of substrate per unit time. Potential issues with this particular method that will be discussed in section 4.4.

#### Measuring what is missing

An interesting way to measure degradation has been used by Farrell et al. [138]. By expressing a GFP with and without a degradation tag, they could measure what is "missing" from the strain with the degradation tag (assuming production of the two proteins is indentical). This has allowed measurement of steady-state degradation rates (proteins/generation) in a variety of mutants using bulk fluorescence assays. However, as the authors noted, these rates also include degradation of substrate before protein folding, which can be 20 times faster than processing fully folded substrates [108].

#### In vitro

In vitro measurements have been incredibly useful to identify specific interactions as well the molecular mechanisms of proteolysis. They also enable measurement of the substrate-enzyme kinetics. However, they do not capture the complex in vivo environment and they cannot answer the questions of the following section, concerning time dynamics and cell-to-cell heterogeneity.

### 4.1.3 Motivation

A lot has been learned about proteolysis in bacteria, but many basic questions remain unanswered. For example, what are the degradation kinetics of substrates in vivo? Do they obey Michaelis-Menten

kinetics? Do they commonly operate at saturation, as recently reported [81]? Measurement of the in vivo kinetics can also be useful for the design of synthetic circuits, which commonly use proteolysis to obtain faster protein responses [45, 56–61]. The kinetics are also non trivial because of the many different unfoldases, peptidases and adapters with overlapping specificity for many different substrates, forming degradation "queues" (Figure 4.2).This arrangement of multiple queues for degradation is in striking contrast with eukaryotes, where degradation is performed by a single enzyme, the proteasome[2]. Have the queues evolved any properties that prevent fluctuations in one substrate concentration from propagating to other substrates? How noisy are degradation rates for natural substrates? Given that proteolysis regulates many important stress sigma factors (e.g. RpoH, RpoS and RpoE), fluctuations in degradation could have a major impact on the cells. Such fluctuations could even be exploited by the cells, as many toxin-anti-toxins modules linked to persisters are regulated by proteolysis [139–142]. For example, heterogeneity in degradation rates of an anti-toxin could result in cells entering a persister state. In contrast to other basic mechanisms such as transcription [16], protein degradation has not been well characterized in single cells, and therefore the dynamics in cells and the extent of the heterogeneity remain largely unknown.

If keeping degradation rates constant is important, there could be feedback control on the rates of degradation. Although feedback control for noise suppression is notoriously difficult to achieve [11], protease could use post-translational feedback to minimize delays, by degrading themselves for example. In fact, many proteases process themselves autocatalytically (e.g. ClpP, FtsH) or are known

[2]Interestingly, bacteria from actinobacteria and nitrospira phyla have a proteasome replacing one or more proteases

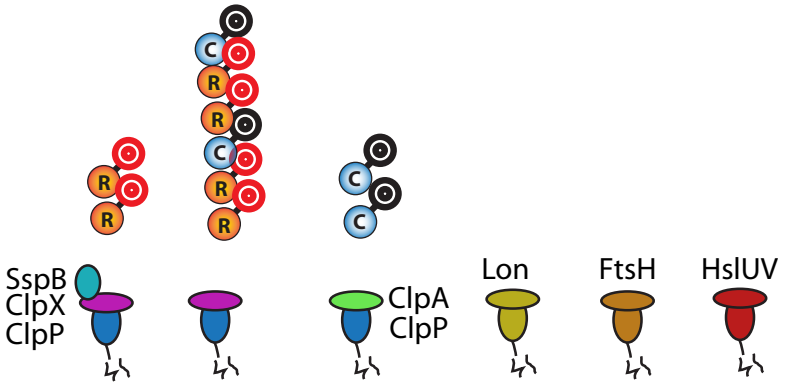



**Figure 4.2:** Cartoon schematic of degradation queues formed by proteases and adapters. Proteases have overlapping specificity of some substrates. Note that these queues differ from the traditional first-in-first-out customer serving ordering. They are more appropriately described with priority ordering, with the binding affinity of the substrate to the protease determining the priority.

to degrade themselves (e.g. ClpAP degrades ClpA). In addition, there is evidence for transcriptional control: ClgR is a substrate of ClpXP and up-regulates ClpXP in some bacteria [143–145], forming a potential feedback loop. The next section will further motivate the development of tools to study these processes in vivo in single cells.

## 4.2 Counter-intuitive effect of an adapter protein on the repressilator

In chapter 3, we observed a counter-intuitive effect where a GFP tagged for degradation did not compete with other substrates, but instead acted as an "anti-competitor". Briefly, the presence of *ssrA*-(ASV) tagged molecules made the oscillations faster rather than slower, whereas the presence of natural *ssrA*-(LAA) tagged molecules had the expected effect of slowing down the oscillations (explained in more details in section 3.5.1). Here, we continue to investigate this effect to further pinpoint the underlying mechanism. Specifically, we measured the period of the repressilator as a proxy for the degradation rate of the *ssrA*-(LAA)-tagged repressor in many different deletion mutants

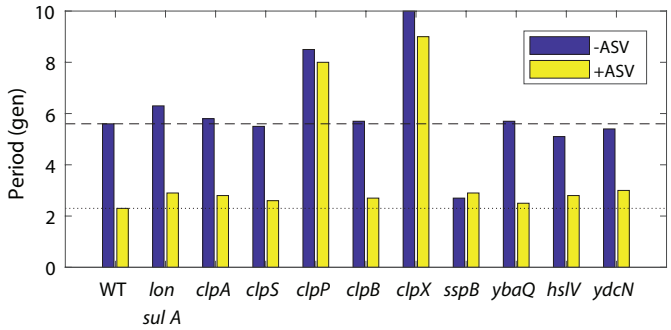



**Figure 4.3:** The ASV-effect is caused by titration of the SspB adapter proteins. Period of the repressilator -- measured using the first local maximum of the autocorrelation function -- in deletion mutants in the presence or absence of ASV-tagged molecules. Deletion of SspB recapitulates the effect of the presence of *ssr*A-ASV tagged molecules on the repressilator. The strains are, from left to right ((LPT25, LPT53), (LPT323, LPT324), (LPT189, LPT202), (LPT185, LPT198), (LPT190, LPT203), (LPT186, LPT199), (LPT191, LPT204), (LPT187, LPT200), (LPT139, LPT133), (LPT188, LPT201), (LPT140, LPT134). Note that half of the strains have an additional RFP, which allowed us to look at two distinguishable strains per lane in the microfluidic device.

related to proteolysis in the presence or absence of *ssrA*-(ASV) tagged molecules (Figure 4.3).

The results showed that deleting the SspB adapter protein was sufficient to recapitulate the effect of *ssrA*-(ASV) tagged molecules, whereas combining the two does not change the effect. This result was surprising because the SspB adapter protein has been shown to increase the affinity of the ClpXP protease to *ssrA*-(LAA) tagged molecules and therefore *increase* their degradation rate by the ClpXP protease [110]. Here, removing SspB increases the frequency of the oscillations, most likely by increasing degradation rates. It is straightforward to see how *ssrA*-(ASV) tagged molecules can recapitulate the effect of deleting SspB. As shown in Figure 4.4, the ASV mutation of the tag contains the intact SspB binding sequence of the tag, while the ClpX binding sequence is mutated [146]. Therefore, the SspB proteins can be effectively titrated away by the presence of the *ssrA*-(ASV) tagged molecules, while these proteins are probably not degraded by ClpXP because of the mutated tag.


**Figure 4.4:** Schema showing how the *ssrA*-(ASV) tagged molecules can compete for binding to the SspB adapter. The SspB binding region of the tag is intact, while the sequence that binds ClpX has been mutated. Mutation of one of the last three amino acids is sufficient to inhibit ClpX binding [146], and therefore the molecules are most likely not recognized and degraded by the protease.

Our first hypothesis was that removing the SspB adaptor, or titrating it out with a GFP that cannot be delivered to ClpXP, would prevent all SspB dependent substrates in the cell from being degraded by the protease, and that the other substrates therefore have less competition. However, the adaptors are thought to primarily recognize the degron rather than the rest of the protein. When we measure the degradation rate of an *ssrA*-tagged GFP (section 4.3.2), the rate is instead slowed down in the absence of SspB. The opposite behavior of the repressors and the GFP, despite having the same *ssrA*-tag suggests that the effect is instead related to the structural differences of the substrate. Our second hypothesis is that, in the absence of the SspB adapter, *ssrA*-tagged susbtrates can be degraded by other proteases with lower affinity but faster degradation rates. Multiple different proteases have been reported to recognize and degrade *ssrA*-tagged substrates (ClpAP [104], Lon [147], FtsH [148]), even though the vast majority of *ssrA*-tagged substrates are degraded by ClpXP in vivo [102] (which makes sense because of the very high affinity of SspB). The results rule out ClpAP and Lon, because deletion of these proteases do not block the ASV effect. This leaves only FtsH, whose deletion is not viable unless combined with a suppressor mutation (*fabZ* [127]) and was therefore not included in this initial screen. Testing this double mutant to see if deletion of FtsH remove the ASV effect would be a crucial experiment to reject this hypothesis.

The other mutants showed minor or expected changes in the period. Deletion of Lon (here combined with a supressor mutation to prevent filamentation, *sulA* [122]) reduced only slightly the degradation rates, consistent with a minor role in degrading *ssrA*-tagged substrates. ClpA, ClpB, ClpS and HslV deletions showed no major changes in the period, while ClpP or ClpX deletions essentially eliminated degradation (period comparable to the repressilator without degradation tags, Figure 3.10). Note that deleting ClpP or ClpX funnels a large pool of the *ssrA* substrates (about 1 in 200 translated protein is naturally *ssrA*-tagged [100, 102]) to the other proteases, which could saturate them and make these particular mutants difficult to interpret.

YbaQ is a predicted transcription factor and was identified as a substrate of ClpXP [40], and we conjectured it could be a transcriptional regulator of ClpXP. However, deletion of that gene did not impact significantly degradation rates. Finally, YdcN was included being an unrelated protein as a control.

These results emphasize the importance of measuring proteolysis in vivo and the surprising effects of the complex proteolytic environment. This further motivated the development of a new assay to measure proteolysis rates in vivo in single cells.

## 4.3 Measuring instantaneous proteolysis rates in vivo

### 4.3.1 Our approach

A simple description of degradation dynamics formulates a differential equation for the concentration of the degradation-tagged substrate as:

$$\frac{dy(t)}{dt} = p(t) - \beta(t)y(t) \tag{4.1}$$

where $y(t)$ is the concentration of the substrate (i.e. molecules per volume), $p(t)$ the time-varying normalized production rate and $\beta(t) = \beta_{deg}(t) + \beta_{dil}(t)$ the time-dependent elimination rate per molecule (including protein degradation and dilution).By using concentration of molecules (or molecules per cell area assuming the radius of the bacterium is constant), we can approximate the loss of molecules at cell division by exponential dilution (as we and others have used previously [69]). The production rate is therefore normalized per area (more details in section B.3.2), which remove cell cycle dependence (as seen in chapters 2 and 3).

In order to measure $\beta_{deg}(t)$, we need to be able to measure the concentration of the substrate over time, its production rate and the dilution rate (i.e. growth rate over time). We therefore use fluorescent proteins targeted for degradation as substrate, an approach subject to limitations as discussed below. Most of this chapter will focus on *ssrA*-tagged substrates in order to compare to the main previous report of proteolysis in single cells [81] .

To measure the production rate, we use a translational fusion of CFP and YFP, linked by a se-

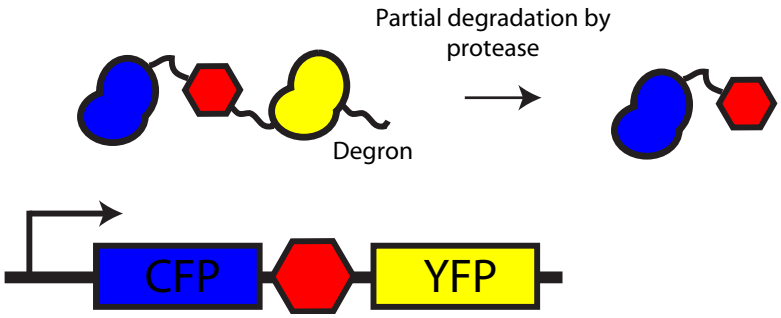



**Figure 4.5:** Schematic representation of our approach to measure instantaneous degradation rates in single cells. CFP and YFP are translationally fused and linked by a roadblock ($GA_{r30}$-Titin I27-$GA_{r30}$). The substrate is targeted for degradation by a C-terminal *ssrA* tag, and the slippery sequence causes partial degradation of the substrate.

quence that inhibits degradation of the other fluorescent protein (Figure 4.5). It has been previously shown [132] that a "slippery" sequence (alanine and glycine repeats[3]), followed by a "roadblock" (a protein domain with a structure that is difficult to degrade) cause the protease to stall in its processivity and fall off the substrate. The resulting substrate lacks a degradation and will no longer be recognized and degraded by the protease. We therefore use a $GA_{r30}$-Titin I27-$GA_{r30}$ domain as a linker (Materials and Methods, section B.3), which has been shown to cause partial degradation of 77% of *ssrA*-tagged substrate by ClpXP [132]. Note that we expect our rate of partial degradation to be much greater because fluorescent proteins themselves can act as roadblock, as we will show in section 4.4.3. Because of the double roadblock-slippery sequence, we expect about 95% of substrates to be partially processed[4]. We will refer to this setup as the slippery construct.

The number of CFP molecules produced will thus equal the number of YFP molecules produced, but only the latter are degraded. The production rate $p(t)$ can be calculated with the CFP (as in

[3]Interestingly, this sequence originate from the Epstein-Barr virus nuclear antigen-1 (EBNA1) to evade degradation by the proteasome [149–151]

[4]Of the 23% of the proteases that degrades the titin domain without falling, we expect only 23% to degrade the CFP. Therefore, only $\sim$5 % of the slippery construct will be completely degraded.

section B.2.3) after fluorescent units conversion. This conversion factor can be calculated by using a strain where the ClpXP protease is deleted, or by simply removing the tag on the substrate. Because we can calculate the instantaneous production rate, we can express the slippery construct using *any* time-varying rates. We will use both a constitutive expression or oscillatory inductions using the repressilator, and the type of induction will be specified in the respective sections.

The experiments were conducted in our microfluidic device (Figure 3.1), allowing us to follow hundreds of cell lineages under constant growth conditions as they grow and divide for hundreds of consecutive generations.

An example experimental time trace with calculated degradation is shown in Figure 4.6, where the substrate's lifetime ($\tau(t) = 1/\beta(t)$) is plotted alongside the CFP and YFP intensity. The construct was expressed at a constant but noisy level in order to observe degradation at different substrate concentrations. Multiple checks were performed to validate these degradation rates. The conversion factor between CFP and YFP units was obtained in two different ways which gave similar results: using the *clpXP* mutant (Figure A.2) and using the fact that when dilution completely dominates the half-life is equal to one generation. We also observed protease saturation at the same substrate concentration using an orthogonal method (Figures 4.7 and 4.9). Finally, the saturation curve obtained in the next section is an internal validation and will be compared to in vitro results.

### 4.3.2 Measuring the saturation curve

We started by expressing the construct at intermediate levels using a leaky repressed promoter (Figure 4.6). By using natural plasmid fluctuations to generate a range of substrate concentrations, we

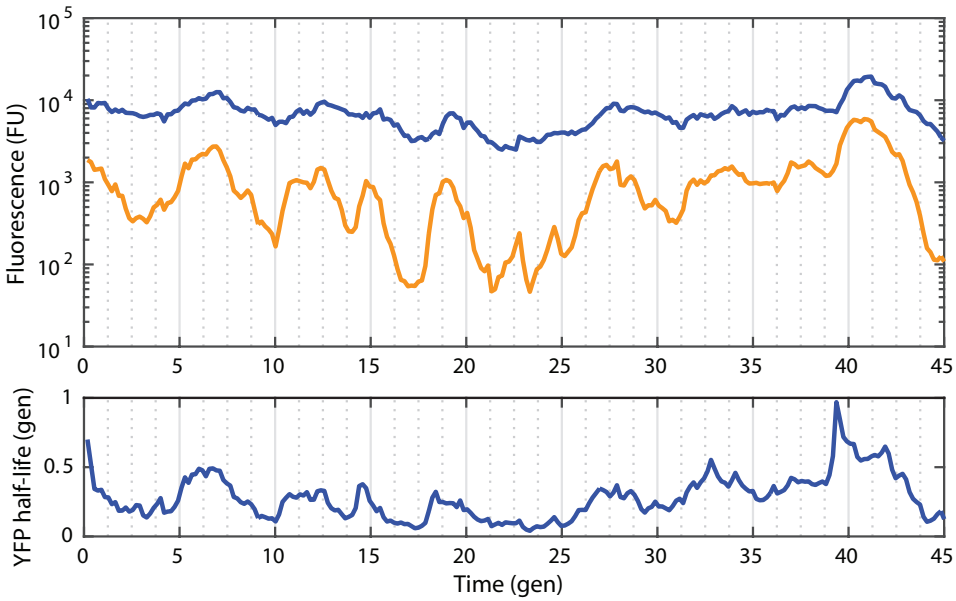



**Figure 4.6:** Example time trace showing the degradation rate of the slippery construct over time (SL86). The construct is repressed by the TetR repressor, which is expressed at low levels to generate fluctuations in gene expression. The fluorescence of the CFP (blue trace) and the YFP (yellow trace), as well as the substrate half-life ($\tau_{1/2}(t) = \log 2/\beta(t)$) fluctuate over time. Degradation rates are calculated with equation (4.1) (Materials and Methods, section B.3)

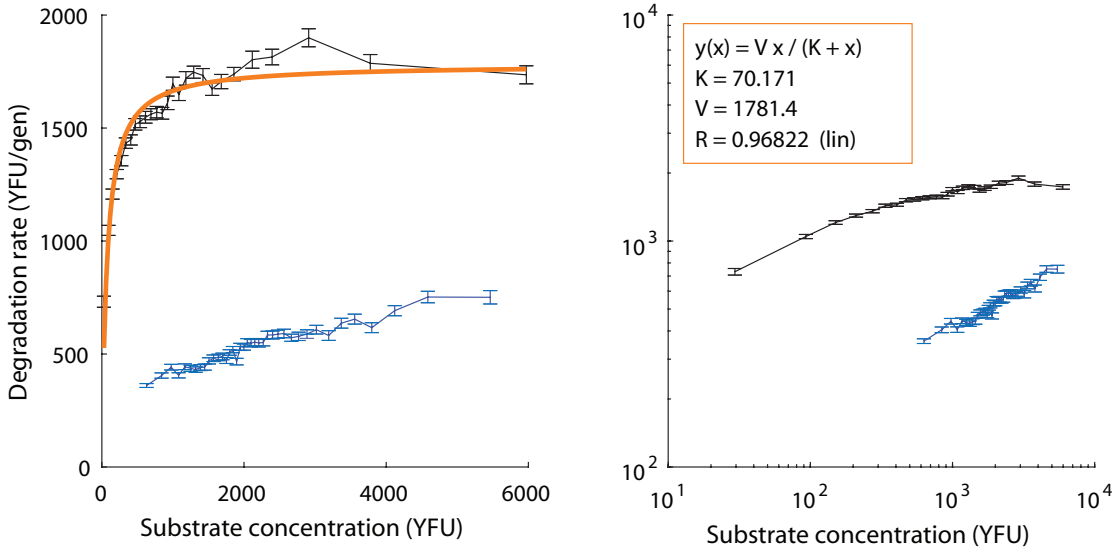



**Figure 4.7:** Saturation curve of the slippery construct in the WT and *sspB* mutant strain. The total degradation rate ($\beta_{deg}(t)y(t)$) is shown as a function of the substrate concentration ($y(t)$). The WT (SL86, black curve) fits well to Michaelis-Menten kinetics (orange curve). The *sspB* mutant (blue curve, LPT311) shows decreased degradation rates. Each point is a bin of 200 time points averaged in *x* and *y*, with the error bars indicating standard error on the mean.

could build a saturation curve (Figure 4.7), which showed the dependence of the total degradation rate ($\beta_{deg}(t)y(t)$) as a function the substrate concentration. The saturation curve fits quite well to the expectation from Michaelis-Menten kinetics:

$$\beta_{deg}(t,y)y(t) = \frac{y(t)V_{max}}{y(t)+K_M} \tag{4.2}$$

which was observed in vitro [110, 152] but not yet in vivo. The affinity is also remarkably strong, with a $K_M$ of ~70 FU, close to our detection limit. In the SspB deletion strain, we see that, as expected, the binding affinity goes down, but our range of concentrations is then not wide enough to conclude whether proteolysis in that strain also follows Michaelis-Menten kinetics.

Note that these results apparently contradict the observation in section 4.2 that *ssrA*-tagged substrates are degraded faster in the absence of the SspB adapter. However, the structure of the degraded substrates are different: in the former case, the substrates were the three repressors of the repressilator (CI, LacI and TetR), whereas here the substrate is a fluorescent protein. Most or all known fluorescent proteins have a structure inherently difficult to unfold and degrade, and indeed we only observed degradation of fluorescent proteins when targeted to the ClpXP protease via the C-terminus (section 4.4.3). This emphasizes the importance of the substrate's local structure for degradation, and we posit that here we are measuring mostly ClpXP-mediated degradation due to the stability of the substrate (similarly discussed in reference [138]). The ClpAP protease could possibly degrade this construct, although its degradation of *ssrA*-tagged substrate has generally been considered minor compared to the other proteases [102, 104, 138, 147]. In the case of the repressors targeted for degradation, we hypothesize that other proteases are participating in their degradation.

### Comparison to literature and validation

We measured degradation rates along the saturation curve varying from first-order to zero-order kinetics. These findings contrast with the only other single-cell measures of protein degradation in bacteria [81], which report exclusively zero-order kinetics. This discrepancy may be explained by complications of the pulse-decay method when assessing degradation of high-affinity substrates, which will be discussed in section 4.4.

Using the Michaelis-Menten fit, we can obtain the substrate half-life in the linear unsaturated

regime ($y \ll K_M$):

$$\tau_{1/2} = \ln 2 \frac{K_M}{V_{max}} = 0.76 \text{ min} \tag{4.3}$$

This is strikingly close to what has been reported by bulk in vivo measurements: natural *srA*-tagged substrates with a half-life of <0.5 min [102] and 0.5 min for CI-SsrA [110]. We expect the degradation rates in our case to be lower because 1) the substrate is a GFP, which has a structure that is particularly difficult to degrade (see section 4.4.3) and 2) the protease might spend some time on the slippery sequence before releasing the substrate.

One obvious next step is to convert the FU into actual protein numbers (such as in reference [153]). However, we can still use existing in vitro data as a rough validation of our saturation curve. Indeed, as the saturation curve contains two independent parameters ($K_M$ and $V_{max}$), we can use one to find a protein to FU conversion factor and validate the other. Using an in vitro $K_M = 75$ nM (at 30°C, [152]) and 1.7 nM = 1 molecule/cell [138], we obtain a $V_{max} = 1120$ YFP/gen, or 0.95 YFP/min/ClpXP (assuming 43 ClpXP complexes per cell [138]). This is again strikingly close to the in vitro reported $V_{max} = 1.2$ GFP/min/ClpXP (at 30°C [152]), and an order of magnitude closer than previous attempts at comparing in vivo to in vitro data [138]. In that previous report, the authors did not measured degradation rates directly, but rather compared the concentration of GFP with and without *srA* tag. As the authors dutifully noted, their rates include degradation of the substrate before folding, which is an order of magnitude faster [108]. Indeed, with such a strong binding affinity, the half-life of these molecules is close to the time GFP takes to fold (GFP takes 1 min to fold in vitro [154]). This comparison was included as a rough validation, because we expect

all of these parameters to change at different temperature (e.g. see Table 4.2) and in vivo. There is also considerable differences in the measurement of the parameters in different reports (e.g. see references [110, 113, 155, 156]). The data further provide a rough estimate of the actual number of substrates we expressing in the cell (1 yellow FU/YFU $\approx$ 0.6 molecules/cell).

## 4.4 Comparison to pulse-decay method and limitations

The main objective of this section is to use the traditional pulse-decay method (i.e. pulse the substrate and measure its decay) in order to compare it to our approach. The results will explain the discrepancy with previous reports while highlighting the advantages and limitations of our approach.

### 4.4.1 Limitations of pulse-decay method

In order to compare to previous measurements of single cell degradation rates (and to provide yet another validation), we replicated a typical pulse-decay experiment by placing the slippery construct (CFP-*slip*-YFP-SsrA, section 4.3.1) under the control of the repressilator as an autonomous pulse generator (Figure 4.8). The periodicity of the oscillations can be seen clearly in the CFP channel: every 14 generations, a strong induction is followed by complete repression indicated by exponential dilution. As expected, the YFP peaks decay much faster due to degradation.

Surprisingly, many peaks that can be seen in CFP channel do not produce a corresponding YFP peak. It appears that the small induction peaks are below the YFP detection limit. In figure 4.8, the second peak corresponds to ten times less substrates than the first peak (as measured on the CFP peak), but the number of substrate measured is at least 50 fold less (below the detection limit). This

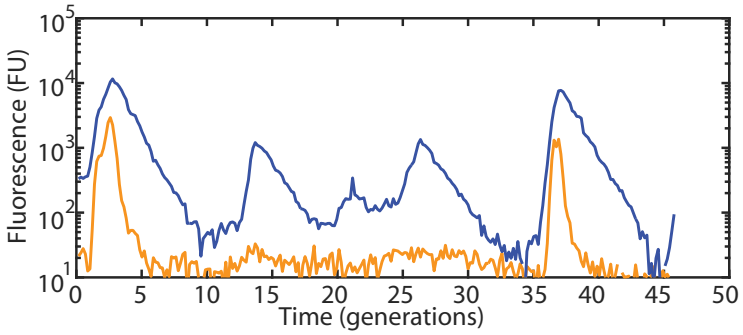



**Figure 4.8:** Representative time trace of the slippery construct (CFP-Slip-YFP-SsrA) induced by the repressilator (SL61). The oscillations, with a period of 14 generations, can clearly be seen in the CFP trace (blue). The CFP is simply diluted out exponentially, while the YFP (yellow trace) is degraded. Some peaks in the YFP are below the detection threshold, even though we should be able to detect them if the substrate half-life was constant.

non-linearity suggests that the proteases are saturated during the large induction peaks.

To better understand this effect, we used the distribution of peak heights generated by the repressilator to plot the relation between the "input" peak height (number of substrates produced, as estimated from the CFP peak height) and the "output" peak height (number of substrates observed, from the actual YFP peak height). The saturation effect can be seen in Figure 4.9, by analyzing three distinct zones. At very low peak heights, the number of substrates is below the detection limit, which we call "missing peaks". At intermediate values, the slope on the loglog plot is approximately two: that is, if we produce twice as many substrates, we observe four times as many, demonstrating that the degradation rate is changing as a function of the substrate concentration. Finally, we observe complete saturation of the proteases when the slope goes back to one. The saturation point occurs around a concentration of 2,000-3,000 YFU (yellow FU), which corresponds to where we observe saturation on the saturation curve (Figure 4.7).

These results highlight the difficulties of measuring degradation rates by using a pulse-decay

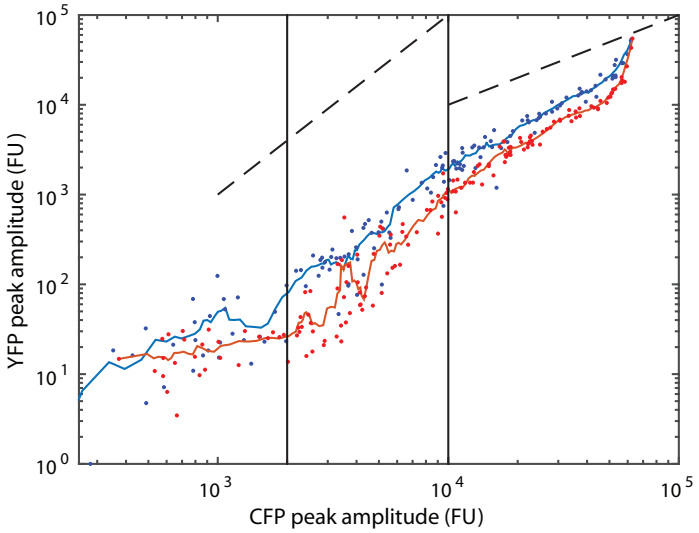



**Figure 4.9:** As the induction of the slippery construct (SL61) reaches the detection limit, the proteases are saturating. The relation between the ``input'' peak height (number of substrates produced, CFP peak height) and the ``output'' peak height (number of substrates observed, YFP peak height) is shown. Two different experiments with different exposure settings (blue and red, the red has higher CFP exposure) are shown, with each dot representing one peak, and the trace is a moving average of 8 points. Three distinct zones have been separated by vertical lines. On the left, the YFP peaks are below the detection limit. In the center, the YFP peaks can be detected, and their amplitude increase approximately as the square of the CFP peak amplitude, indicating decreasing degradation rates. On the right, the proteases are completely saturated, and the relation between CFP and YFP peak heights become linear. Dashed lines representing slopes of two and one are shown to guide the eye. Note that we observe camera saturation at $2^{16} = 65,536$ FU.

method for substrate with high affinity and high degradation rates. If the pulses are too low, the degradation rate is so high as to make it impossible to observe the pulses (partly due to the fluorescent protein maturation), and if the pulses are too high, the protease saturates. This perhaps explains why previous in vivo degradation results have reported only protease saturation [81], and showcases an advantage of the instantaneous measurements we report here.

### 4.4.2 Manipulating the degradation rate

Despite the limitations from the pulse-decay method, we used it to characterize degradation in different proteolytic environments using a variety of mutants due to the simplicity of the assay. To increase degradation, we inserted a second copy of the $P_{clpPX}$-*clpPX* operon close to the replication origin. To decrease degradation, we used a partially functional *clpP-mVenus* translational fusion previously characterized [76]. We also grew cells in a different temperature (32°C) and used a modified *ssrA* tag that was shown to improve delivery of the substrate by SspB to ClpXP (*ssrA*-NYNY [152]).

We started by pulsing different substrates (CFP-SsrA, CFP-SsrA-NY and the slippery construct) using the repressilator, and simply measured the frequency of observed peaks as a proxy for degradation rates (Table 4.2). The peak frequency was either calculated by using a stable FP reporter for the oscillations or approximated using the numbers of peaks per cell division observed. Lower frequency means more missing peaks, suggesting higher degradation rates.

The results are mostly as expected. The over- and under-expression appears to increase and decrease degradation, respectively. Removing SspB decreases the degradation rate, an effect which is exaggerated when the substrate carries the mutated *ssrA* tag with improved delivery to the protease.

**Table 4.2:** Approximate frequency of observed peaks for different substrates, tags and conditions. The peaks were driven by the repressilator, and the frequency of observed peaks is used as a proxy for degradation rates.

| | CFP-SsrA | CFP-SsrA-NYNY | CFP-slip-YFP-SsrA |
|---|---|---|---|
| WT | 0.43 | 0.29 | 0.68 |
| WT 32°C | 1 | 0.76 | |
| *sspB* | 1 | 0.96 | |
| *sspB* 32°C | 1 | 0.71 | |
| $P_{clpPX}$-*clpPX* over | 0.14 | | |
| *clpP-mVenus* | 0.89 | | |
| $P_{clpPX}$ decoy | 1 | | |

Degradation is also lower at lower temperature, as expected. Finally, the slippery construct appears to be degraded more slowly than only the CFP-SsrA, which makes sense if the protease spends some time on the slippery sequence before releasing the substrate.

Representative peaks are plotted in Figure 4.10. Again, the limitations of this method comes to light. For example, the decay of the peaks between the WT and the *sspB* mutants are remarkably similar, even though we have measured higher degradation for the WT (Figure 4.7 and Table 4.2). This suggests again that this method measures degradation rates closer to protease saturation, as the degradation rates is more similar between these strains at saturation. In fact, the *rise* of the peak is much sharper for the WT, indicating higher degradation at low abundances. Why isn't it possible to measure the decay at low abundances? It appears that a small fraction of the substrates cannot be degraded and are only diluted, which appears as a "kink" on the decay curve (discussed in section 4.4.3), possibly because of partial degradation where only the tag is removed. These non-degradable substrates therefore hide the "interesting" part of the decay curve.

Measuring degradation rates from this curve is thus quite challenging. The top part of the decay

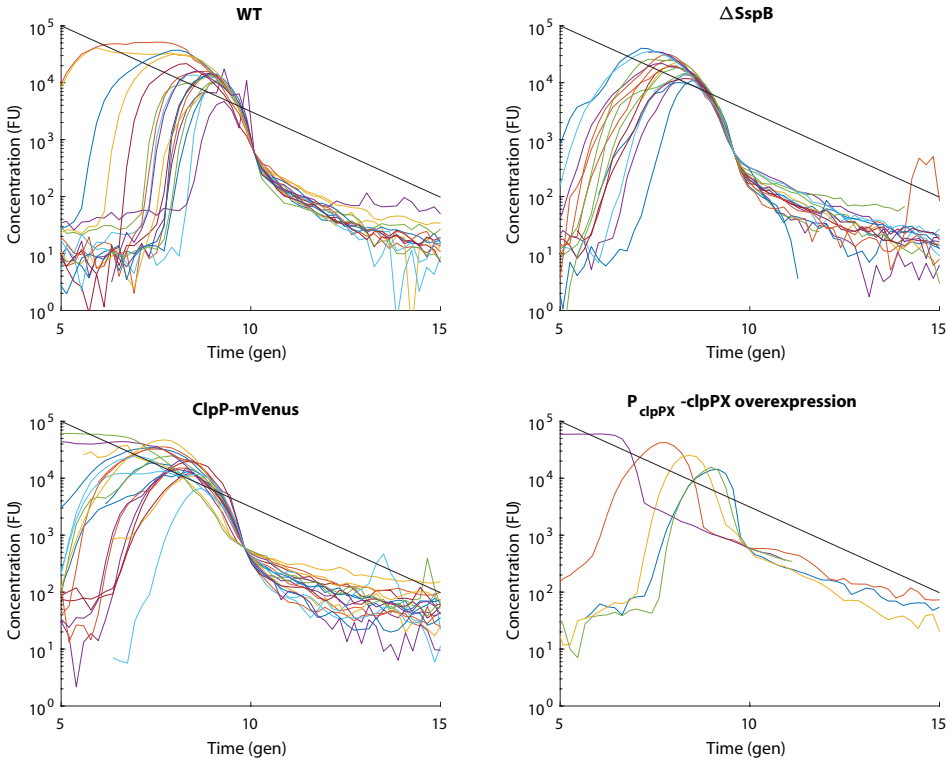



**Figure 4.10:** Comparison of pulse-decay measurements across different mutants. Peaks were induced autonomously using the repressilator, and the fluorescence decays over time as production is turned off. A solid line is shown indicating the dilution rate. Peaks are horizontally aligned using an arbitrary set value. The strains are, in the clockwise direction, WT mVenus-SsrA (LPT232), Δ*sspB* mVenus-SsrA(LPT241), $P_{clpPX}$-*clpPX* overexpression CFP-SsrA (LPT269) and *clpP-mVenus* CFP-SsrA (LPT266).

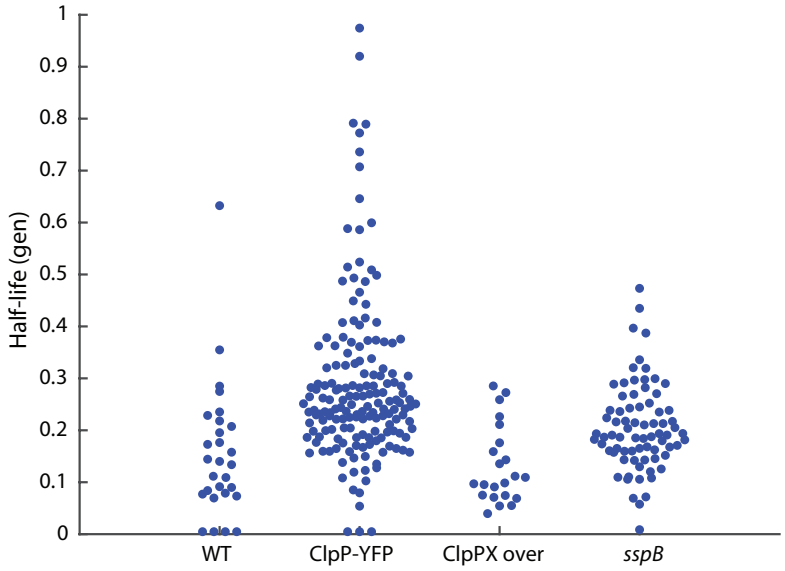



**Figure 4.11:** Half-lives of *ssrA*-tagged fluorescent proteins as measured per the pulse-decay method. The half-lives are longer in the partially functional ClpP-mVenus translational fusion (LPT266) and in the *sspB* mutant (LPT241) than in the WT (LPT232), but shorter in the $P_{clpPX}$-*clpPX* overexpression. Each point represents a fit to the exponential part of the decay curve and their *x* positions have been moved to avoid overlap.

curve is limited by saturation of the protease and the switch to full gene repression, while the bottom curve is hidden due to the kink. We can nevertheless fit the limited region of the decay curve to obtain the approximate half-life (Figure 4.11).

As expected, degradation is faster in the strain that over-expresses ClpP and slower in the ClpP under-expression or *sspB* mutants. The half-lives are substantially higher than those obtained from the saturation curve (0.03 gen, Figure 4.7). This is consistent with the fact that the half-lives of Figure 4.11 are measured outside of the first-order regime (closer to saturation), therefore providing an upper bound on the substrate's half-lives (i.e. the actual half-lives are lower).

Taken together, these results highlight the pitfalls of measuring degradation rates using the pulse-decay methods, especially for high affinity substrates with half-lives $\leq$ 1 min. To the best of our

knowledge, the higher degradation rates of *ssrA*-tagged substrates by ClpXP with SspB was not yet measured in vivo[5]. Our approach therefore provides the first confirmation of the in vitro measurement.

### 4.4.3 Fluorescent proteins can act as degradation roadblocks

#### A fraction of *ssrA*-tagged fluorescent proteins are not degraded

As mentioned in the previous section, using the pulse-decay method highlights that a fraction of the substrates targeted for degradation cannot be degraded and are simply diluted out (Figure 4.10). As the decay reaches around 1% of the peak height, it switches to exponential dilution. Because the lifetime of the degraded substrates is at least 10 times smaller than the stable population, we estimate that less than 0.1% of produced proteins cannot be degraded (i.e. 1/10 of the steady-state levels of the substrate). This fraction is within range of errors in transcription or translation that would cause the substrate to no longer be recognized by ClpXP[6].

We also suggest an additional possibility: fluorescent proteins can act as roadblock and cause partial degradation (i.e. clipping off the tag). In fact, this phenomenon can be seen in Dirk Landgraf's PhD thesis [160]. He was observing the targeting of an RpoS750-Venus fusion to ClpXP on gel (Figure 4.12). We believe the band he labeled as "degradation product" is actually the fluorescent protein (the gel is labeled with GFP antibodies). As we will see in section 4.4.3, we did not observe any degradation of the *fluorescence* of that construct, supporting that only the RpoS750 is degraded. Taken

[5] Without using synthetic degradation tags [157] or double mutants [110]

[6] $10^{-3}$-$10^{-4}$ for errors in translation and $10^{-4}$-$10^{-5}$ for errors in transcription [158, 159]. Mutation of one of the last three amino acids is sufficient to abolish recognition by ClpXP [146].

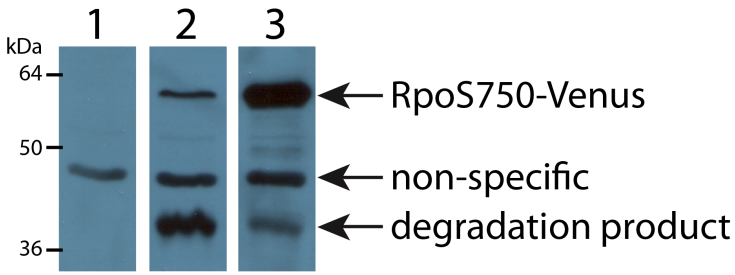



**Figure 4.12:** Western blot using $\alpha$GFP antibody showing the partial degradation of the RpoS750-Venus construct. The full construct is efficiently reduced to the Venus in the WT (lane 2), while deletion of RssB (essential for RpoS degradation) leads to accumulation of RpoS750-Venus (lane 3). A non-specific band is shown to cross-react in the WT MC4100 (lane 1). The expected molecular weights are 55.4 kDa for RpoS750-Venus and 27 kDa for Venus. Figure reproduced with permission from reference [160].

together, these results suggest that, at least in certain configurations, fluorescent proteins can act as roadblocks causing partial degradation and release by the proteases.

### Fluorescent proteins targeted for degradation by many natural tags are not degraded

We targeted fluorescent proteins for degradation by ClpXP, ClpAP and Lon using degrons that are normally encoded on the genome (Figures 4.13, A.3, A.4 and Table 4.3). Only the SulA tag produce any visible fluorescence decay beyond dilution with a $\tau_{deg,1/2} = 75$ min – comparable to degradation of GFP-SsrA by Lon in vitro [147] but much lower than 1 min, the in vivo half-life of SulA [161]. The other substrates showed fluorescence decay following exponential dilution due to growth.

We suggest that the FPs are partially degraded by the proteases (like in Figure 4.12) with these degrons because their structure makes them resilient to degradation. This is perhaps not so surprising as degradation of GFP has only been reported in vivo using the *ssrA* tag. Many of these results were

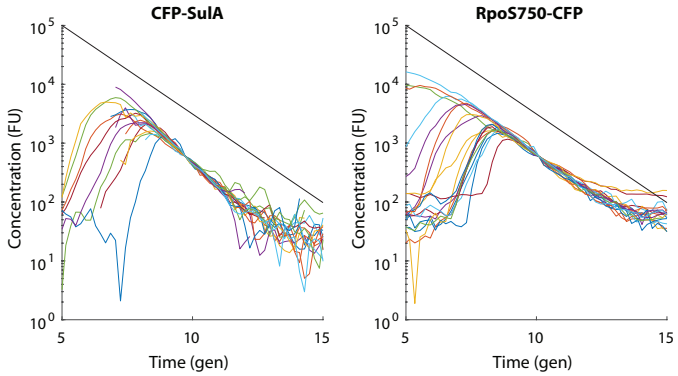



**Figure 4.13:** Pulse-decay measurement of fluorescent proteins targeted for degradation with genomic tags. CFP-SulA is slowly degraded (SL49, $\tau_{deg,1/2}$ = 75 min) while RpoS750-CFP (LPT263) follows exponential dilution due to cell growth. The solid lines indicate the decay rate for exponential growth.

in fact expected: ClpXP can degrade FP from their C-terminus [146, 154] but apparently not their N-terminus [129][7]. The N-end rule degron we used require pre-processing by an enzyme to expose the tag [96, 118], which might depend on recognition of the whole protein rather than simply the degradation tag. The RepA and MuA tags have been reported degraded in vitro or not in this exact configuration. It is possible that the exact position of the degron with respect to the FP structure has an impact on the rate of partial degradation. Finally, the Lon protease has very weak unfoldase activity and was reported to degrade FP with their normal structures very slowly [147, 162].

This result highlights the limitations of using FP for measuring degradation. We will discuss alternative methods that could overcome this issue in section 4.7, and will focus the rest of the results on degradation of *ssrA*-tagged proteins.

[7]This result was reported as "Personal communication from J.A. Kenniston"

**Table 4.3:** Summary of FP-degrons pairs assessed for degradation by loss of fluorescence.

| | Protease | Observed Degradation | References |
|---|---|---|---|
| $Dps_{1-20}$-CFP | ClpXP, ClpAP | No | In vivo, Dps $\tau_{1/2}$ = 10 min, [40]<br>In vitro, ClpXP $Dps_{2-12}$-Arc, $\tau_{1/2}$ = 5 min [40]<br>In vitro, ClpAPS $Dps_{6-16}$-GFP, $\tau_{1/2}$ = 15 min [96] |
| $IscS_{1-11}$-CFP, $IscS_{1-11}$-mNeonGreen | ClpXP | No | In vitro, $IscS_{1-11}$-Arc, $\tau_{1/2}$ = 15 min [40]. |
| CFP-$SulA_{20C}$ | Lon | Yes,$\tau_{1/2}$ = 75 min | In vitro, GFP-titinI27-$sulA_{20C}$, $\tau_{1/2}$ = 45 min [129] |
| RpoS750-CFP, RpoS750-mNeonGreen | ClpXP | No | In vivo, RpoS750-LacZ, $\tau_{1/2}$ = 3 min [163] |
| CFP-RpoS750 | ClpXP | No | |
| CFP-$RepA_{1-15}$ | ClpAP | No | In vitro, GFP-$RepA_{1-15}$, $\tau_{1/2}$ = 3 min [164] |
| CFP-$MuA_{2-11}$ | ClpXP | No | In vitro, GFP-MuA, $\tau_{1/2}$ = 25 min [164] |

## 4.5 Interference with the cell

In this section, we will investigate whether our degradation measurements interfere with natural cellular processes. We will use transcriptional reporters for various genes, and then further verify if the *reporters* are interfering with the cells.

### 4.5.1 High concentration of *ssrA*-tagged substrates induce production of ClpXP

We monitored levels of the ClpXP protease using a transcriptional reporter (including all reported transcription start sites of the *clpPX* operon) while pulsing CFP-SsrA using the repressilator (Figure 4.14). At high substrate concentration, the levels of the proteases increased by about 50% compared to without substrate, and there was no increase when the substrate was not tagged for degradation.

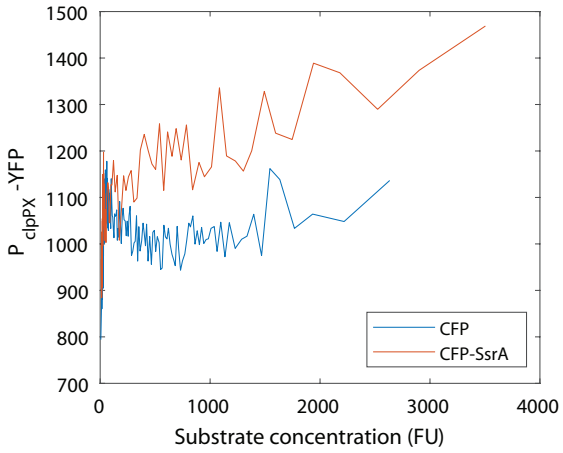



**Figure 4.14:** Transcriptional reporter for *clpPX* increases with higher level of *ssrA*-tagged substrates, but not with untagged proteins. Concentration of the translational reporter as a function of the substrate concentration. The substrates (CFP, SL148, blue trace, and CFP-SsrA, SL149, red trace) were autonomously induced using the repressilator. Each point is a bin of 50 time points averaged in *x* and *y*.

This effect was also observed in the translational fusion *clpP-mVenus* integrated at the native locus (Figure A.5). This compensation reveals a negative feedback mechanism where the substrate concentration affects its own degradation rate.

### Toxicity of the slippery substrate

We also assessed whether this feedback was present while inducing the slippery construct using the repressilator. We mutated the YFP of the construct to make it non-fluorescent in order to measure $P_{clpPX}$-*yfp* expression (i.e. CFP-*slip*-darkYFP). High substrate concentration also increased protease expression (Figure 4.15), but there are two surprising observations.

First, the substrate accumulates in much higher concentration than just the CFP (almost three-fold). This is not expected as the CFP is not degraded, is expressed from the same promoter and should have the same translation since the region close to the RBS is identical. Longer mRNA have

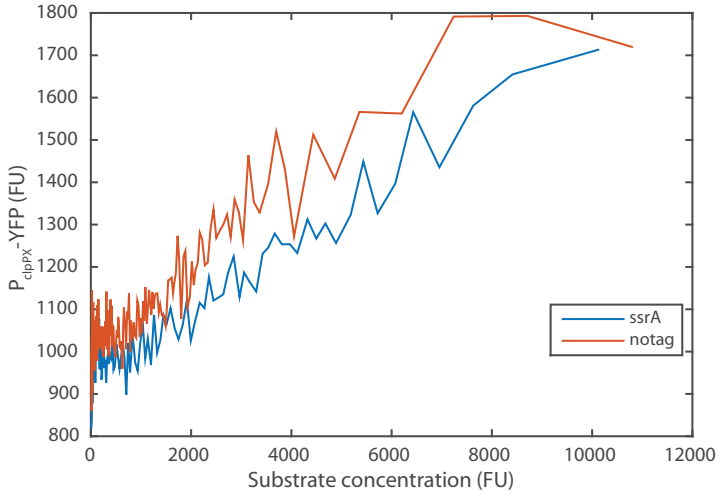



**Figure 4.15:** The slippery construct induces expression of $P_{clpPX}$. Concentration of the transcriptional reporter as a function of the substrate concentration. The increase in the levels of the transcriptional reporter are slightly higher when the slippery construct in untagged. The substrates (CFP-slip-darkYFP, SL153, red trace, and CFP-slip-darkYFP-SsrA, SL123, blue trace) were autonomously induced using the repressilator. Each point is a bin of 50 time points averaged in *x* and *y*.

been shown to be effectively stabilized against RNAses, as translation can begin while transcription is still in progress [165], which could contribute to an increase of ∼50%. However, we suggest that the higher levels reflect toxicity of the substrate, as we observed a drastic reduction (up to twofold) in growth rate at high substrate concentration (Figure 4.16). If these slowdowns in growth are not balanced by reduction in expression of the substrate, they would cause its accumulation.

Secondly, the slippery construct without degradation tags induced the same, if not greater, elevation in protease expression as its *ssrA*-tagged counterpart. This suggests a second facet to this response, that the slippery construct itself can cause production of proteases. This will be discussed further in section 4.6.

In order to validate that the increase in protease levels resulted in higher degradation rates, we

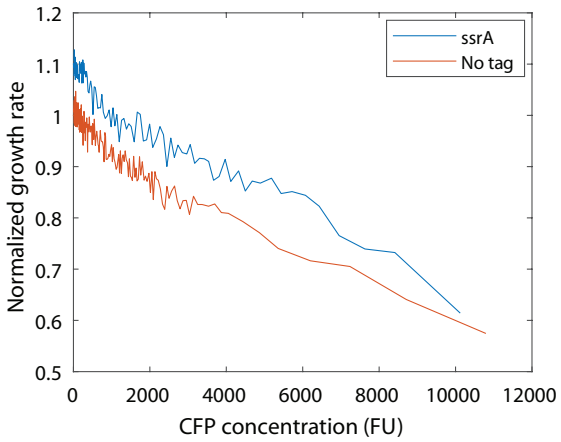



**Figure 4.16:** Toxicity of the slippery construct. The normalized growth rate slows down as a function of the substrate concentration. The substrates (CFP-slip-darkYFP, SL153, red trace, and CFP-slip-darkYFP-SsrA, SL123, blue trace) were autonomously induced using the repressilator. Each point is a bin of 50 time points averaged in *x* and *y*.

constructed a time-dependent saturation curve (Figure 4.17). By separating degradation rates temporally, we saw that the rates were much higher after the substrate concentration had crossed a set threshold than before (i.e. hysteresis). This supports the previous results: high concentration of substrates causes the production of more proteases and higher degradation rates.

In order to better understand this response and toxicity, we monitored downstream expression of stress sigma factors. We observed that RpoH-dependent (heat shock sigma factor) expression increased by about twofold (Figure 4.18). The transcription of the RpoH gene also increased (Figure A.7), although this particular reporter exhibited high levels of toxicity (discussed in section 4.5.2). RpoS-dependent expression was only marginally increased and $P_{RpoE}$ was unchanged (Figure A.8).

We evaluated the role of SspB, ClpXP and RpoS in the toxicity and the production of proteases by deletion mutations (Figure 4.19). The mutants did not inhibit the response (in fact they slightly increased it), indicating that these proteins are not necessary in the potential feedback loop. How-

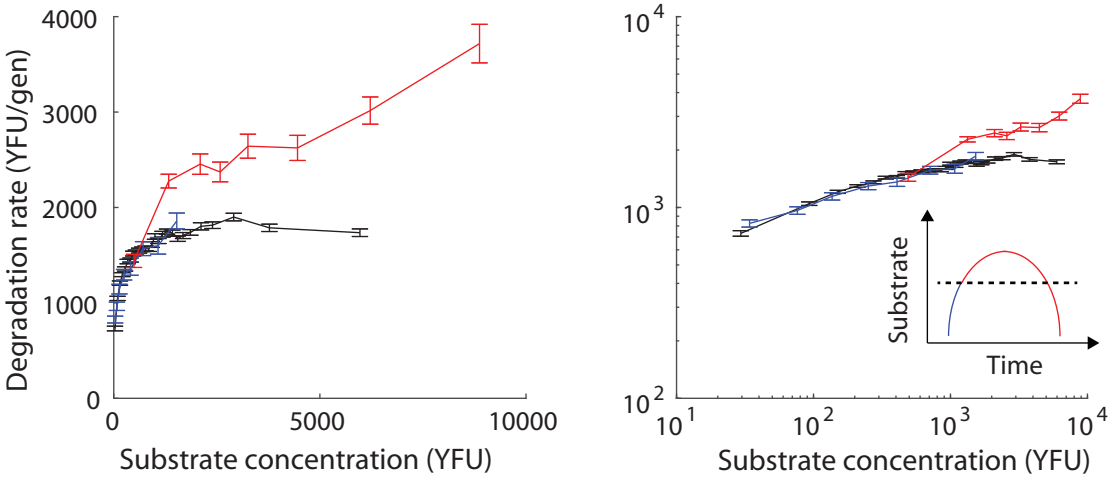



**Figure 4.17:** Saturation curve of the slippery construct in the WT is dependent on the history of substrate concentration. The total degradation rate ($\beta_{deg}(t)y(t)$) is shown as a function of the substrate concentration ($y(t)$). The substrate was autonomously induced using the repressilator (SL61, blue and red curves) or expressed at constant levels (SL83, black curve). The induction peaks have been separated using a threshold of 2,000 YFU (see inset): time points before the threshold was reached are binned in blue, while time points after the threshold are binned in red. The blue points overlap with the constant induction curve (black), while the red curve shows much higher degradation rates, suggesting higher levels of the proteases. Each point is a bin of 50 or 200 time points averaged in *x* and *y*, for the blue/red and black curves, respectively, with the error bars indicating standard error on the mean.

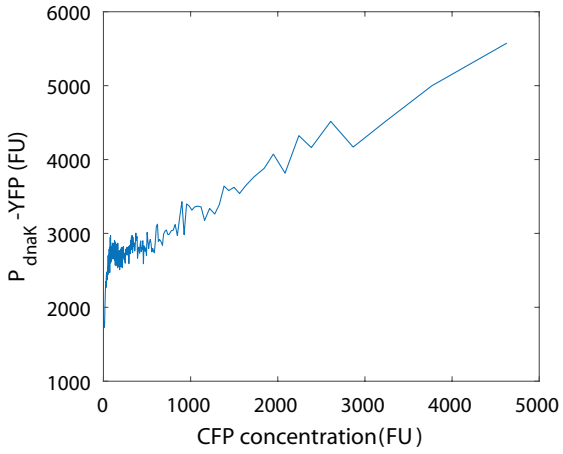



**Figure 4.18:** An RpoH dependent gene is expressed in the presence of the slippery construct. The transcriptional reporter $P_{dnaK}$ expression is shown as a function of the slippery construct concentration (CFP-slip-darkYFP-SsrA, SL156). The substrate was autonomously induced using the repressilator. Each point is a bin of 50 time points averaged in *x* and *y*.

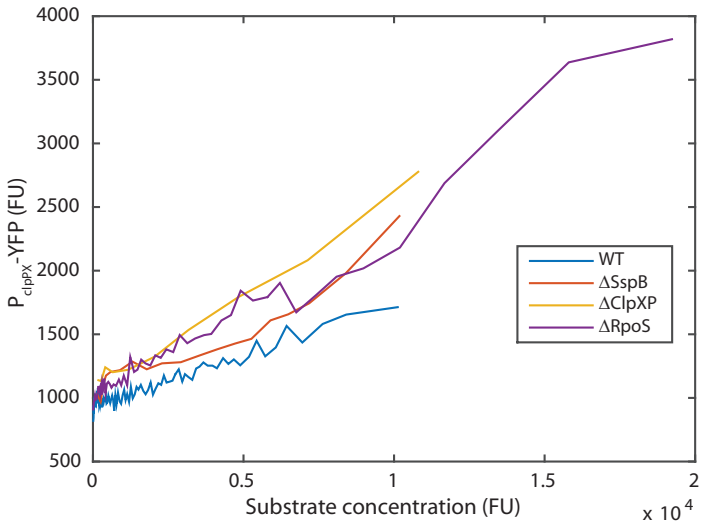



**Figure 4.19:** Mutation of SspB, ClpXP or RpoS do not inhibit the response to the slippery construct. Concentration of the transcriptional reporter ($P_{clpPX}$-YFP) as a function of the substrate concentration. The increase in the levels of the reporter is slightly higher in the mutants (*sspB*, red trace, SL159, *clpXP*, yellow trace, SL160, *rpoS*, purple trace, SL158) than in the wild-type (blue trace, SL123). The substrate (CFP-slip-darkYFP-SsrA) was autonomously induced using the repressilator. Each point is a bin of 50 time points averaged in *x* and *y*. Note that these strains exhibit high levels of cell death.

ever, all these knock-out mutants showed extreme sensitivity to substrate over-expression (about 90% of the cells died after each substrate induction), which is surprising since the mutants are very healthy without the substrate over-expression. This suggests that these three proteins play an important role against the toxicity of the slippery substrate.

### ATP levels are affected by high concentration of *ssrA*-tagged substrate

Because ATP plays an important role in degradation, we measured changes in ATP concentration during repressilator-induced peaks of *ssrA* substrates by using the QUEEN ATP sensor [166]. This sensor is composed of a circularly permutated GFP fused to an ATP-binding domain. The binding

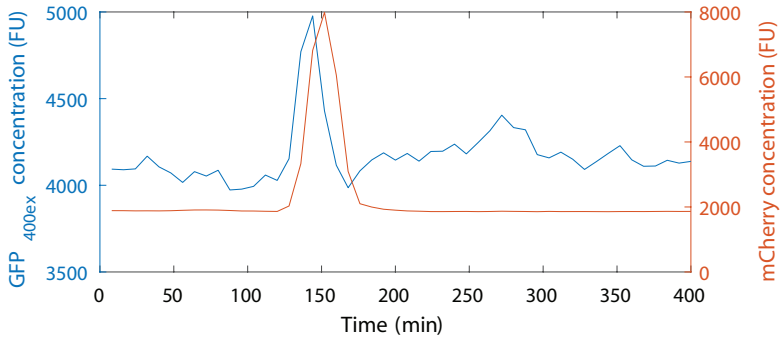



**Figure 4.20:** ATP concentration increases with induction of *ssrA*-tagged substrates. Representative time trace of mCherry-SsrA (red) induction using the repressilator, with the QUEEN sensor (blue trace) reporting ATP levels (SL46). The concentration was obtained using the average of a small region comprising the cell in ImageJ.

of ATP to that substrate therefore influences fluorescence and the concentration of ATP can be measured by the ratio of fluorescence at two different excitation wavelengths. Because the sensor is a GFP, and we are pulsing mCherry-SsrA, we did not have a segmentation marker and could not perform the automated segmentation routine. However, we could still obtain preliminary results by manually analyzing the data in ImageJ. Representative time traces are shown in Figures 4.20 and A.9. Virtually all observed peaks in mCherry were correlated with peaks in the fluorescence ratio, which correspond to an increase in ATP. The height of the mCherry peaks also seemed to correlate with the increase in ATP. Using a mutant ATP sensor insensitive to changes in ATP concentration confirmed that the increase in fluorescence was not due to spectral bleed-through.

These results should be considered as preliminary, as the lack of image segmentation prevents quantitative analysis. We would also need to calibrate the sensor to convert the 400ex/494ex ratio into ATP concentration. However, it suggests that high quantities of *ssrA*-tagged proteins cause an increase in the concentration of ATP in the cells.

### 4.5.2 Transcriptional reporter interferes with gene expression

We noticed an increase in cell death in strains carrying the $P_{clpPX}$-*yfp* transcriptional reporter (located on the low copy, 5-10 per cell, pSC101 plasmid), which is surprising given that these reporters are commonly viewed as "readouts" that do not interfere with the cell. To assess whether the $P_{clpPX}$-*yfp* transcriptional reporter was interfering with the cell, we measured the frequency of repressilator-driven peaks of CFP-SsrA with and without the transcriptional reporter (Table 4.2). Surprisingly, we found that the transcriptional reporter decreased the degradation rates (peak frequency went from 0.43 to 1), even slightly more than the translational fusion (0.89), suggesting that transcriptional reporter could lower the levels of the ClpXP protease.

Moreover, the slowdown in growth due to the slippery construct was also dependent on the particular transcriptional reporter. This resulted in differences in the maximum substrate concentration reached (e.g. Figure 4.15 vs. Figure 4.18). In fact, removing the $P_{clpPX}$-*yfp* almost completely eliminated the toxicity; showing that the toxicity is mainly observed when both the slippery construct and the transcriptional reporter are present.

## 4.6 Discussion

We have developed and validated the first method for measuring instantaneous proteolysis in single cells. Using this method, we measured the first saturation curve for protein degradation in vivo. This showed that, as expected, the SspB-ClpXP complex has a very high affinity to *ssrA*-tagged substrates. The saturation curve showed Michaelis-Menten kinetics, while deletion of the SspB adapter

greatly reduced the affinity. We would need to obtain a greater range of concentration to conclude about the kinetics of the *sspB* mutant.

We validated our approach with multiple different methods. The necessary fluorescent units conversion factor was obtained in different ways which gave similar results. We compared the saturation curve to established in vivo results by calculating the substrate half-life, and to in vitro results using both the binding affinity and the degradation rate at saturation. The congruency of these findings supports the accuracy of our method for measuring in vivo degradation rates.

Previous reports have shown protein degradation in single cells to be zero-order (i.e. saturated [81]), without any report of the SspB adapter influencing the in vivo kinetics. By using similar pulse-decay methodology, we showed that these results are due to the challenges of this particular experimental method. The substrates have an extremely low $K_M$ and high $V_{max}$, such that small pulses are not detected and big pulses saturate the proteases. Indeed, the degradation rates of ClpXP close to saturation are similar with or without the adapter.

Using an indirect measurement of proteolysis – with the period of the repressilator – we saw that deletion of SspB could *increase* degradation in some context. We suggested that this difference is due to the structure of the substrates targeted for degradation: GFP was degraded more slowly without SspB while the repressors (CI, LacI and TetR) were degraded more quickly. This highlights the importance of the substrate structure for degradation: we could only measure GFP degradation by ClpXP through its C-terminus, while more natural *ssrA*-tagged substrates are known to be degraded by Lon, FtsH and ClpAP. This has in fact been observed in the literature, although not discussed. Degradation of GFP-*ssrA* was found to be faster with SspB [138], but three different combinations

of mutants show faster degradation of natural *ssrA*-tagged substrates without SspB [102]. Note that the single *sspB* knock-out has curiously been omitted from many studies [102, 110, 138].

This suggest an additional role for the adapter protein in the queuing dynamics: directing substrate with unknown structure (any protein can be *ssrA*-tagged in the cell), to a specific protease which can degrade very resilient structures quickly (ClpXP). This prevents other proteases from becoming saturated with substrates that they cannot degrade, while keeping a layer of redundancy if ClpXP is saturated. In fact, SspB and ClpS have been shown to directly inhibit the ClpAP degradation of *ssrA*-tagged proteins [146, 155].

### Proteolytic response

Although we did not exploit the temporal dynamics of the measurement extensively in this study, it did help us uncover one type of degradation response: a potential feedback loop for maintaining homeostatic proteolysis. We observed two facets to this response, which may or may not be related: an increase in the levels of the ClpXP protease due to presence of *ssrA*-tagged substrates and a heat shock response in the presence of the slippery construct.

The first part does suggest a feedback control on degradation rates. The simplest model would be that a substrate for degradation up-regulates ClpXP. Saturation of the protease would result in accumulation of the regulator, which in turns would up-regulate ClpXP. This response did not seem to depend on SspB, which makes sense as this response happens when ClpXP is saturated, where SspB makes little difference.

The presence of the slippery construct (CFP-YFP fusion linked by alanine-glycine repeats and a

titin domain) also induces production of ClpXP, although it was independent of whether the construct was *ssrA*-tagged. This response is likely mediated by the heat shock sigma factor, as RpoH-dependent expression increased more than two-fold during the response. The *clpPX* operon is known to be positively regulated by RpoH, through the ClpPp3 promoter [167]. We suggest two different mechanisms for up-regulation of RpoH. The Lon protease has been shown to be activated by the presence of unfolded proteins and to degrade DnaA, the initiator of DNA replication, in *Caulobacter crescentus* [168]. The slippery sequence could be recognized by Lon as an unfolded substrate, and lowering the levels of DnaA would increase RpoH transcription because DnaA represses its production [169]. Additionally, complete elimination of DnaA could cause DNA replication arrest, suggesting a potential toxicity mechanism.

Secondly, RpoH could be stabilized via the binding of DnaK to the slippery construct. DnaK is responsible for FtsH-mediated degradation of RpoH [97, 98] and binds to unstructured hydrophobic regions of proteins [97, 126]. Therefore, the presence of the slippery construct could stabilize RpoH if DnaK binds to the slippery sequence. This mechanism of RpoH activation is thought to be the main activation of the heat shock response, by effectively sensing misfolded protein with the chaperone DnaK.

RpoH has a half-life of 1 min in normal conditions, which is consistent with it being present in 120 copies [170] and produced around 3424 molecules/gen [171]. A small decrease in the degradation rate of RpoH could increase markedly the number of RpoH present and explain the response observed. This is consistent with the fact that the response was independent of ClpXP, RpoS and SspB. The response was slightly stronger in the construct that lacked the *ssrA* tag, which supports

that this response is due to the presence of the slippery sequence, since the degradation tag might enable removal of part of that sequence.

The slippery construct could also cause a response similar to the presence of large amounts of *ssrA*-tagged proteins. Protease such as Lon and FtsH are known to recognize unfolded proteins and target them for degradation. Saturation of this protease could then cause accumulation of a regulator of proteolysis (e.g. RpoH or a yet unknown regulator). Another possible scenario, although unlikely, is that the slippery construct is recognized as a foreign protein and targeted for degradation by an unknown mechanism. A recent mechanism similar to the ubiquitin targeting system in eukaryotes has recently been found in *B. subtilis*, where phosphorylated arginine was discovered as a marker for degradation [172]. The authors also reported that a foreign protein, casein, was arginine phosphorylated and degraded. The crucial experiment will be to remove the titin domain and/or the slippery sequence from our construct to see how it influences the response.

The slippery construct also showed toxicity that was observable as drastic slowdown in growth. This was not due to phototoxicity. We suggest two hypotheses for this toxicity. First, the increase in degradation might cause an imbalance in the degradation of toxin-anti-toxin modules (or DnaA, as previously discussed). Many of these modules are degraded by the Lon protease (Rel [139], MazEF [140], CspD [141] and HipBA [142]), which is up-regulated during the heat shock. For example, faster degradation of an anti-toxin could cause the cells to enter persister-like states [142], while slower degradation of the CspD toxin could inhibit DNA replication [141]. The second possibility we advance is that, if this substrate is recognized for degradation by a protease that cannot unfold it, it could saturate the protease. This could cause toxicity through the just described mechanism or

via saturation of the essential protease FtsH.

The toxicity was greatly amplified by two factors: 1) deletion of ClpXP, SspB or RpoS and 2) the presence of 10 copies of a decoy $P_{clpPX}$ promoter. The importance of ClpXP and SspB in the protection against this effect suggest that this toxicity is related to proteolysis. RpoS is slightly more surprising, because its downstream expression remained relatively low throughout the response (although that was without the additional toxicity from the decoy $P_{clpPX}$ promoters). The mutant is also healthy under normal conditions, although it is impaired in exit from stationary phase. We hypothesize that this could prevent the cells from exiting the slow growth toxicity from the toxins and cause cell death. However, RpoS has an important role in the protection of many other stress, such as osmotic shock, oxidative stress, etc. [173], so its protective mechanism might act differently.

It was also quite surprising that the presence of a few copies of a transcriptional reporter decreased protein degradation. It suggests that a transcription factor present in low numbers is titrated away from the native $P_{clpPX}$ locus and reducing expression from the native promoter. In a different study, we have observed a fourfold reduction in transcription by titration of a transcription factor present in 250 molecules per cell (with the same pSC101 plasmid origin [174]). This suggests that it could be enough to titrate RpoH, normally present in 120 copies [170], and reduce expression of ClpXP. The results do not exclude the possibility of another unknown transcriptional regulator of ClpXP being titrated by the decoy promoter.

Multiple other evidences point to RpoH being the mediator of the response. First, the increase in DnaK expression was twice as high as the ClpPX increase, similarly to what has been reported by over-expressing RpoH [167]. Secondly, increase in ATP levels have also been associated with the heat

**Table 4.4:** Summary of alternative methods tested and technical issues.

| | Constructs | Issue |
|---|---|---|
| cp-GFP | cp6-sfGFP, cp6-mVenus | No signal |
| | cp8-sfGFP-SulA, Dps-cp8-sfGFP | Not degraded |
| splitFP | $GFP_{1-10}$, $GFP_{11}$, $CFP_{1-10}$, $CFP_{11}$ | No signal |
| HaloTag | Dps-Halotag7 | No in vivo dye |
| Indirect readout | $P_{spoIIE}$-*cfp*, Spo0A, Spo0A-SsrA, IscS-Spo0A, Spo0A-SulA | Threshold too high for use in fast growing cells |

shock response in *E. coli* [175]. Thirdly, we also observed high toxicity from the slippery construct in combination with a $P_{rpoH}$ transcriptional reporter – presumably lowering RpoH expression again via titration – providing further support for an important role of RpoH in this response.

## 4.7 Alternative methods for measuring proteolysis in single cells

In the previous sections, we highlighted the advantages and disadvantages of our approach compared to other methods for measuring proteolysis. The main drawbacks came from the use of difficult to degrade fluorescent proteins as reporters. In this section, we explore alternatives methods to measure degradation in single cells. These techniques did not prove successful in their current form, and the results are summarized in Table 4.4.

### 4.7.1 Circular permutation of the GFP

The structure of GFP is a beta barrel consisting of eleven beta strands. One idea is to use circular permutation of the fluorescent protein so that the N and C termini open at a different location. Circular permutations expose a different region to unfold to the protease, and it has been shown to

facilitate degradation [162] since the local structure (where the tag is located) is very important for degradation [107].

We tested cp6-sfGFP, cp6-mVenus and cp8-sfGFP as substrate for degradation. The cp6-FPs were not bright enough to be used in our assay (even without degradation tag), while the cp8 were bright but not degraded from either termini with the tags tried (N for ClpXP and C for Lon).

### 4.7.2 SplitGFP

An alternative method would be to use a split GFP that has recently been used for protein localization [176]. The GFP is split into two separate sub-units: the first ten beta sheets and the eleventh beta sheet. The protein becomes fluorescent when the two subunits reconstitute. We could therefore attach a degradation tag on either end of the eleventh beta strand and it would likely be much easier to unfold and degrade than the whole protein. We tested a split GFP and CFP, and it both cases we could not get any signal, even without attaching a degradation tag on the eleventh beta strand, which could be due to slow maturation of the chromophore.

### 4.7.3 HaloTag

The HaloTag is a modified bacterial enzyme that can covalently bind to a molecule of interest, such as a dye [177, 178]. It could therefore be a more natural substrate for degradation and potentially much easier to degrade. The challenging part is to find dyes that can penetrate the live cell membrane. A new dye (JF549 [179]) has recently been used to label penicillin-binding protein (PBP) in the cell membrane of *E. coli* and *B. subtilis* [180, 181]. We tested this dye to label ClpP in live *E. coli*

cells and found that it did not enter the bacteria.

### 4.7.4 Indirect readout via a transcription factor

Finally, another alternative that is less intrusive is to use an indirect read-out instead of direct labeling; for example by targeting a transcription factor for degradation. Spo0A, the *B. subtilis* sporulation master regulator, is known to regulate positively the *spoIIE* promoter [182, 183], and was previously used in *E. coli*. We could therefore tag Spo0A, and use the *spoIIE* promoter to induce a fluorescent protein as a read-out. We could make a calibration curve (e.g. Spo0A concentration vs induction level) by using an untagged Spo0A, along with a transcriptional reporter for Spo0A or a translational fusion of Spo0A with a fluorescent protein. The disadvantage of this method is that we rely on an indirect read-out: they might be delays in the read-out, the sensitivity will be limited on some region of the concentrations, etc.

We tested this approach (Figures A.10-A.13), and found that the SpoIIE promoter was only responsive in slow growing cells. In these cells, we saw that the peaks were smaller when Spo0A was targeted for degradation, as expected if it were degraded. However, in fast growing cells, even the untagged Spo0A could not produce a downstream response. This is probably because the *spoIIE* promoter is activated at high threshold concentration of Spo0A [183] that are not reached in fast growing cells. While there are promoters that are activated at low threshold, most of them are known to be indirectly activated by a different transcription factor. The challenge of this method is to make a reporter that is active across a wide range of concentration, which in theory could be achieved by combining promoters activated at different thresholds.

## 4.8 Conclusion and future directions

We developed and validated the first method to measure the temporal dynamics of protein degradation rates in single *E. coli* cells. By using this tool, we measured the first in vivo saturation kinetics, which validated the increase in affinity of the ClpXP protease to *srA*-tagged substrate with the SspB adapter protein in vivo. We also measured degradation indirectly using the repressilator, showing the importance of the substrate's structure for degradation and the role that plays the different protease queues in the cells. This emphasizes the importance of understanding the single cell, in vivo kinetics of degradation.

This method also helped us discover a potential feedback loop on degradation enabling homeostatic degradation rates, and highlighted the toxicity of our slippery construct. There is however many experiments needed to explore and validate this response.

Taken together, these experimental results have opened up more questions than they have answered, and the future work can be divided on two fronts.

### Measurement of degradation rates and in vivo kinetics

Given the results on the toxicity of the slippery construct and its interference with degradation, the obvious improvement to the methodology would be to use an operon structure instead of the translational fusion. By having the untagged reporter on the same transcript as the tagged reporter, their production would be very similar (because translation noise is much smaller than transcription noise) but not identical. In other words, we would sacrifice a bit of precision for accuracy. If

translation noise appears to be significant, we could explore using translational coupling similar to ribosomal operon.

We could obtain a wider range of concentration in the saturation curve by using an inducible promoter (e.g. $P_{LlacO}$ and IPTG). We could then test the kinetics model for the *sspB* mutant. By testing all the different protease mutants (i.e. of section 4.2), we could measure how the interaction between the queues impact the saturation curve. For example, we could test if the curve obtained for the *sspB* mutant was a combination of Michaelis-Menten kinetics between different proteases, and exactly what pathways are possible for degradation of GFP-SsrA. By measuring the actual conversion factor between fluorescent units and number of proteins, we could compare these curves to the ones measured in vitro, to see how different the kinetics are in the complex cellular environment.

Measuring the saturation curve could also be extended to different degrons. As ClpAP can degrade GFP from its N-terminus, we could explore the N-end rule recognition via the ClpS adapter, provided that we expose properly the recognition sequence (either by fusing the whole protein or via SUMO-mediated cleavage).

We could explore the competition between substrates that are targeted to the same protease, to the same protease but with a different adapters or to different proteases. This could be an interesting way to circumvent limitations of fluorescent proteins: by using a "dark" substrate as a competitor, we could measure how it affects degradation of *ssrA*-tagged GFP. By changing and measuring the constant expression of the competitor, we could see how that affect the saturation curve for GFP-SsrA. This could help us better understand the role of the different peptidases, unfoldases and adapters queues in managing a fluctuating number of different substrates in the cell.

Using the temporal dynamics would enable us to probe how proteolysis changes across growth conditions. We could use a setup recently developed by our colleagues to flow a growing culture in the microfluidic device to measure how degradation rates change along the growth curve. Proteolysis is regulated across different conditions and plays an important role in the activation of stress responses, so understanding how the saturation curve changes across conditions is important.

A major challenge in the field is the identification of adapters or regulators that are not essential for degradation. Our method provides the best resolution on the measurement of degradation rates in vivo. We could therefore use this platform to screen for fine-tuned regulation of proteolysis using a newly developed method to screen in the microfluidic device. By measuring the saturation curve of each mutant, we could identify changes in both $V_{max}$ and $K_M$ caused by deletion or overexpression of particular proteins.

Finally, the development of alternative methods to measure proteolysis would open up more possibilities of degrons-proteases pairs by using more natural substrates. The most promising method is the indirect readout, although it remains challenging to engineer a substrate that can provide sensitivity in a wide range of concentration. An interesting possibility is to use the repressilator as an indirect readout for degradation. Although it is limited to measuring average rates over a short time period and would require more validation, we could use our multi-reporter repressilator to measure the decay of one of the repressor. Simultaneously measuring the substrate peak height and the end of the decay (induction of next component) would allow us to measure the decay time for a particular decay path. We could then test different hypotheses on the saturation curve, and even use this readout in a screen for factors that influence proteolysis of that particular degradation tag.

These experiments could also be performed in *B. subtilis* using the HaloTag as a substrate, because of the availability of the in vivo dye. *B. subtilis* has a similar apparatus of degradation machinery, with ClpC, ClpE and ClpX associating with ClpP, ClpYQ, LonA, LonB and FtsH. These proteases play an important role in the competence development, the response to various stresses and the protein quality control network. Interestingly, ClpCP recognizes phosphorylated arginine as a marker for degradation in a striking similarity to the ubiquitin system in Eukaryotes. All the questions we raised in *E. coli* would apply to this organism as well.

### Proteolysis response

The immediate next step for understanding the proteolytic response would be to disentangle the two facets observed. Are the reponses to *ssrA*-tagged substrate and the slippery sequences part of the same response? To test this hypothesis, we could measure RpoH dependent expression during the presence of *ssrA*-tagged substrate. To verify that RpoH is the factor responsible for the increase in $P_{clpPX}$ expression, we could mutate the known RpoH promoter on our transcriptional reporter to see whether that abolishes the increase. This would also test simultaneously whether the interference from the transcriptional reporter was due to titration of RpoH.

We would then try to pinpoint the exact mechanism of the response. How is RpoH activated? We could test the hypothesis of DnaK titration by removing the slippery sequences and the titin domain. By expressing these sequences by themselves, we could assess whether they are targeted for degradation and whether they are sufficient to induce the response. What types of degradation tags provoke the response (i.e. how specific is the response)? By testing different degrons, we could

see which proteases are involved in the response and if saturation of a particular protease induce the response. Finally, by using our microfluidic screen, we could use a deletion library to find other potential unknown factors implicated in the response.

# A

# Supplementary figures

## A.1 Chapter 2: Fluctuation timescales in micro-organisms

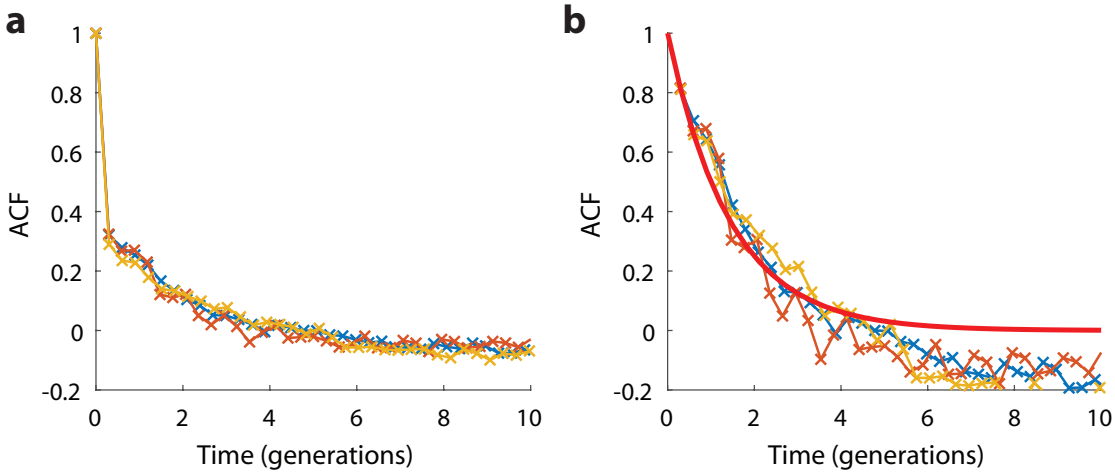



**Figure A.1:** Autocorrelation functions show the absence of slow fluctuations for reporters with low signal. a) The autocorrelation functions for the translational reporters with low signal show a rapid loss of correlation after one time point, which suggests that a large fraction of the variance is due to uncorrelated white noise (blue, FadB-Venus, LPT77, red, KatE-Venus, LPT78, orange, NrdG-Venus, LPT81). b) By correcting for the white noise, the reporters also decay with a timescale of one generation. The decay curve have been normalized so that the first time point equals $2^{-t/\tau_{div}}$. An exponential with a half-life of one generation is indicated in red.

## A.2 Chapter 4: Protein degradation kinetics in single bacterial cells

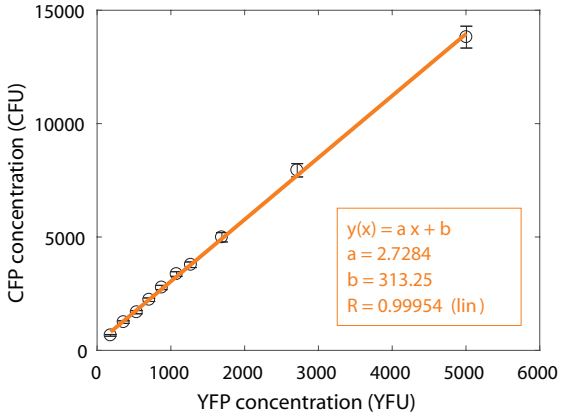



**Figure A.2:** A conversion factor between yellow and cyan fluorescent units (YFU, CFU) can be obtained by expressing the slippery construct in a *clpXP* mutant strain (SL70). The CFP concentration is shown as a function of the YFP concentration, where each point is a bin of 50 time points averaged in *x* and *y* and the error bars indicate standard error on the mean. The linear fit gives: 1 YFU = 2.73 CFU, and the constant represent the difference in background autofluorescence between the two channels. Note that the imaging conditions differ slightly in that particular experiment, and we generally used a 1 YFU = 4/5 * 2.73 CFU = 2.18 CFU.

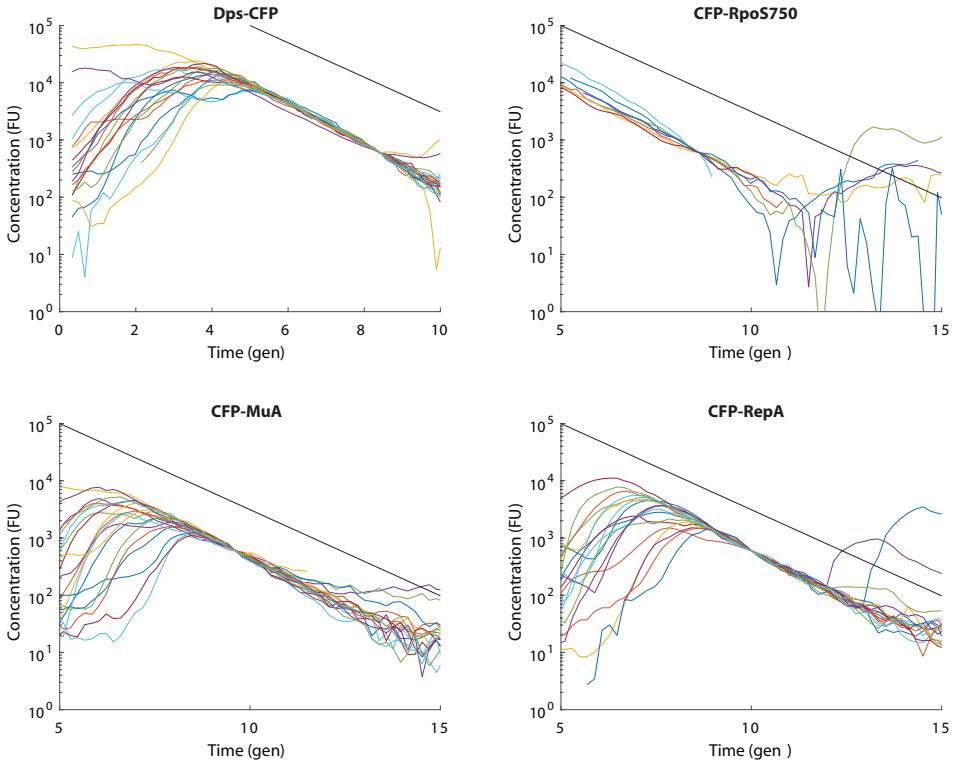



**Figure A.3:** Pulse-decay measurement of fluorescent proteins targeted for degradation with genomic tags show exponential dilution due to cell growth (no degradation). The strains are from top left to bottom right, SL57 Dps-CFP, LPT309 CFP-RpoS750, LPT310 CFP-MuA, LPT312 CFP-RepA. Solid lines indicate exponential dilution due to cell growth.

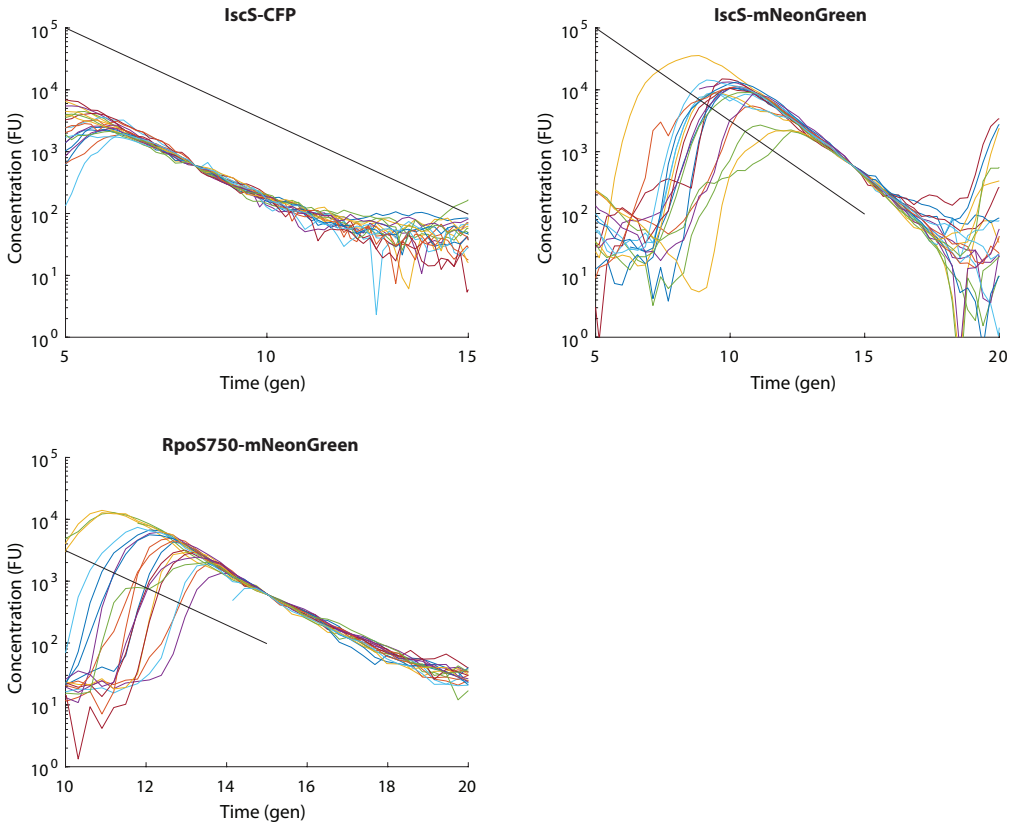



**Figure A.4:** Pulse-decay measurement of fluorescent proteins targeted for degradation with genomic tags show exponential dilution due to cell growth (no degradation). The strains are from top left to bottom right, LPT248 IscS-CFP, SL154 IscS-mNeonGreen, SL155 RpoS750-mNeonGreen. Solid lines indicate exponential dilution due to cell growth.

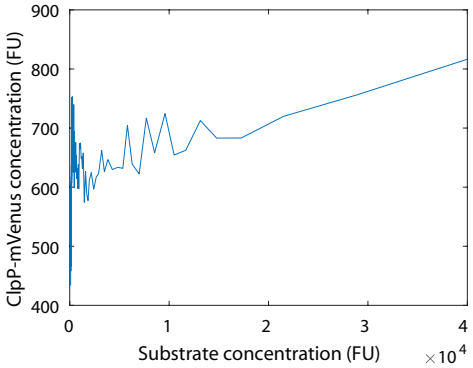



**Figure A.5:** The levels of the translational fusion ClpP-mVenus increase as a function of the substrate concentration (LPT266). The substrate (CFP-SsrA) is autonomously induced using the repressilator. Each point is a bin of 50 time points averaged in *x* and *y*

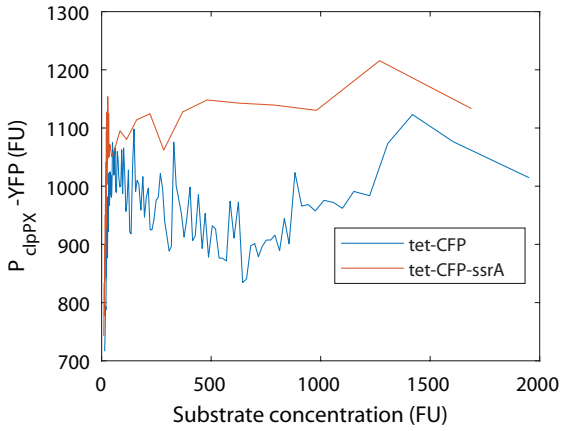



**Figure A.6:** Transcriptional reporter levels ($P_{clpPX}$-mVenus) as a function of the substrate concentration in the *sspB* mutant. The levels increase slightly for the *ssrA*-tagged substrate (SL163), but not for the CFP (SL162). The substrates (CFP-SsrA and CFP) are autonomously induced using the repressilator. Note that the maximum concentration are lower than in Figure 4.14 most likely because the experiment was much shorter. Each point is a bin of 50 time points averaged in *x* and *y*

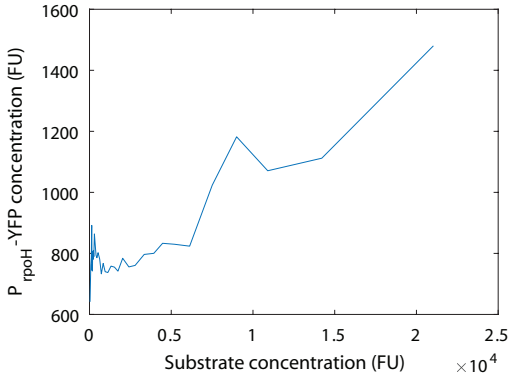



**Figure A.7:** Transcriptional reporter levels ($P_{rpoH}$-YFP) increase as a function of the slippery substrate concentration (SL125). The substrate ( CFP-slip-darkYFP-SsrA) is autonomously induced using the repressilator. Note that there is high levels of cell death. Each point is a bin of 50 time points averaged in *x* and *y*.

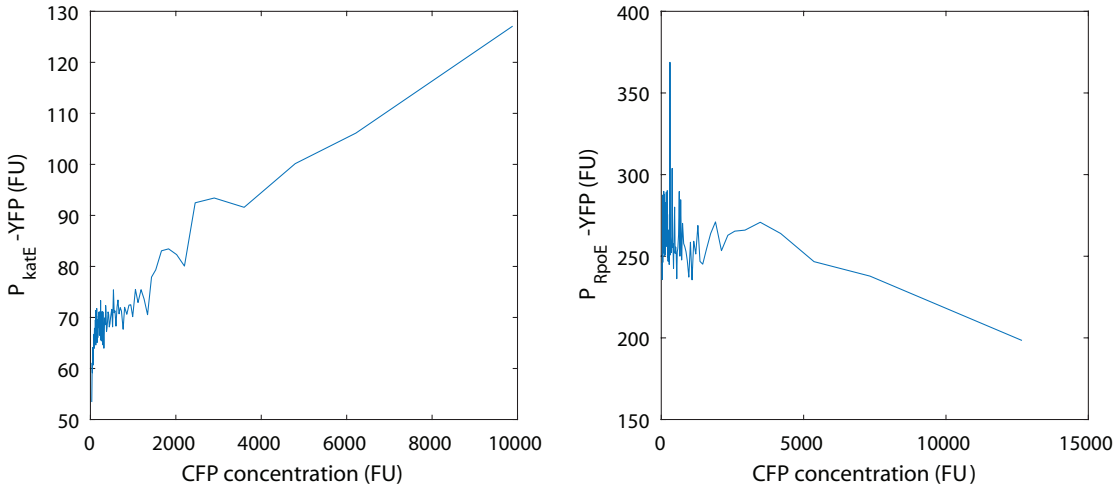



**Figure A.8:** Transcriptional reporter levels as a function of the slippery substrate concentration. RpoS-dependent expression ($P_{katE}$-YFP, SL157) increases slightly at high levels of the substrate, but this might be due to cell close to death. Levels entering stationnary are ≈350 FU. $P_{rpoE}$-YFP levels do not change as a function of substrate concentration. The substrate ( CFP-slip-darkYFP-SsrA) is autonomously induced using the repressilator. Each point is a bin of 50 time points averaged in *x* and *y*.

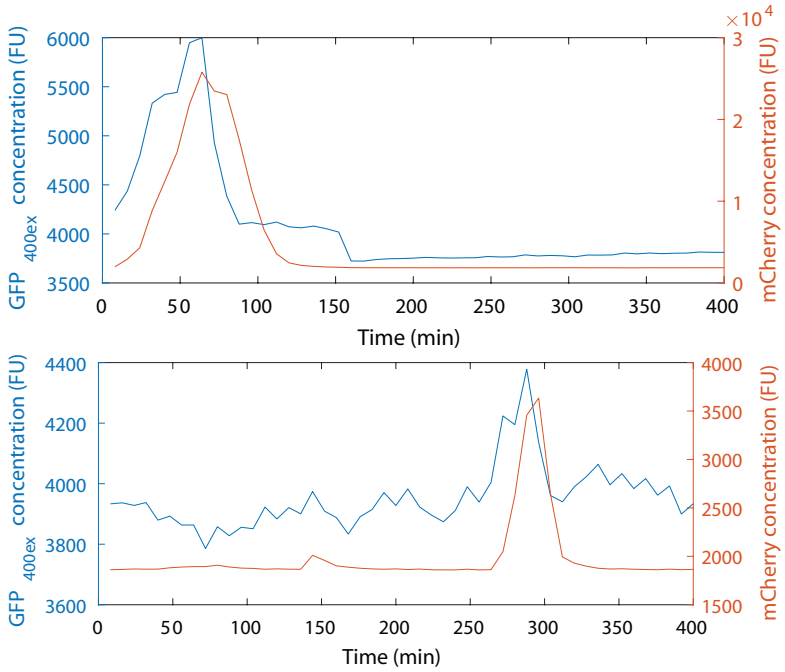



**Figure A.9:** ATP concentration increases with induction of *ssr*A-tagged substrates. Representative time traces of mCherry-SsrA (red) induction using the repressilator, with the QUEEN sensor (blue trace) reporting ATP levels (SL46). The concentration was obtained using the average of a small region comprising the cell in ImageJ.

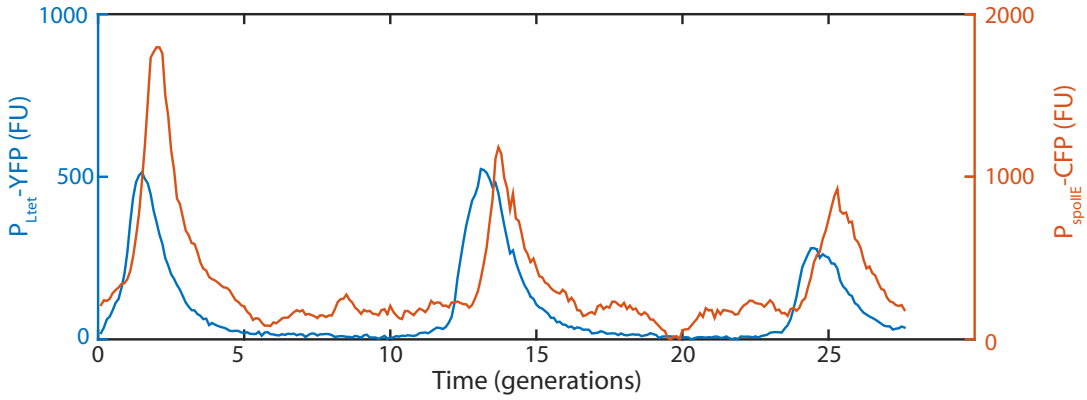



**Figure A.10:** Time trace of indirect degradation reporter. The oscillations of the repressilator, shown using the repressilator ($P_{Ltet}$-mVenus, blue trace), are used to induce a substrate ($P_{Ltet}$-Spo0A, SL89). The levels of the substrate are inferred by a Spo0A-dependent promoter ($P_{SpoIIE}$-CFP, red trace). Automated segmentation was not possible due to the slow cell growth in the long channels ($\tau_{div} \approx 45$ min). The concentration was obtained using the average of a small region comprising the cell in ImageJ.

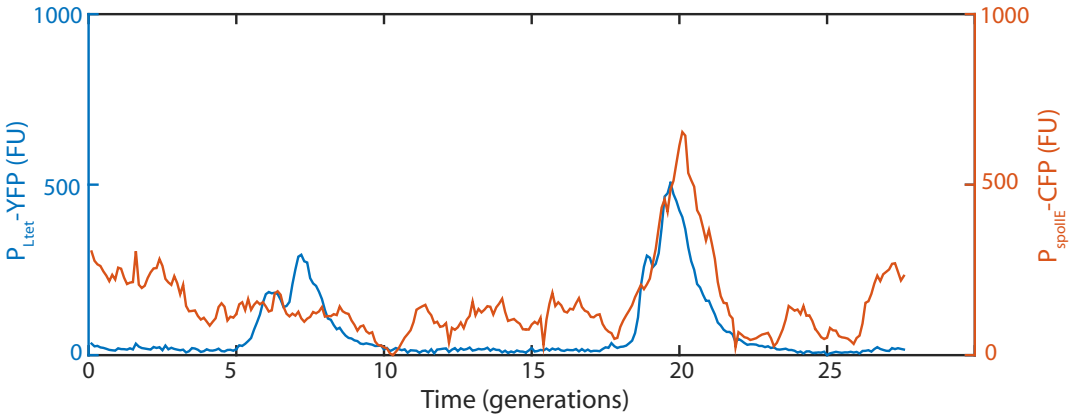



**Figure A.11:** Time trace of indirect degradation reporter. The oscillations of the repressilator, shown using the repressilator ($P_{Ltet}$-mVenus, blue trace), are used to induce a substrate ($P_{Ltet}$-Spo0A-SulA, SL90). The levels of the substrate are inferred by a Spo0A-dependent promoter ($P_{SpoIIE}$-CFP, red trace). The levels of the substrates appear lower than when it was untagged. Automated segmentation was not possible due to the slow cell growth in the long channels ($\tau_{div} \approx 45$ min). The concentration was obtained using the average of a small region comprising the cell in ImageJ.

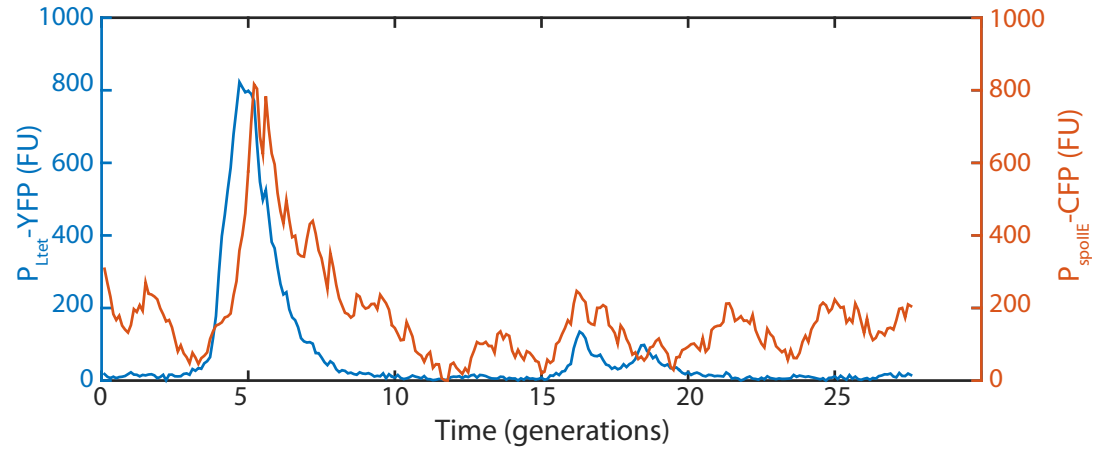



**Figure A.12:** Time trace of indirect degradation reporter. The oscillations of the repressilator, shown using the repressilator ($P_{Ltet}$-mVenus, blue trace), are used to induce a substrate ($P_{Ltet}$-IscS-Spo0A, SL91). The levels of the substrate are inferred by a Spo0A-dependent promoter ($P_{SpoIIE}$-CFP, red trace). The levels of the substrates appear lower than when it was untagged. Automated segmentation was not possible due to the slow cell growth in the long channels ($\tau_{div} \approx 45$ min). The concentration was obtained using the average of a small region comprising the cell in ImageJ.

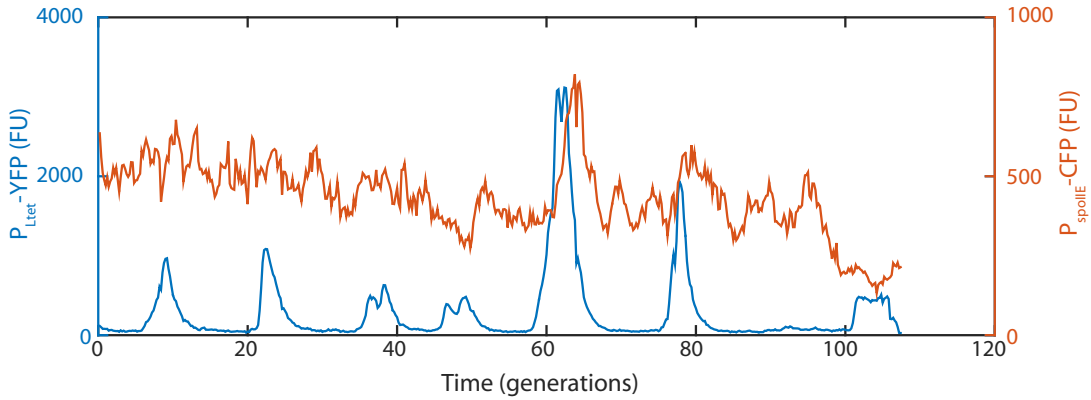



**Figure A.13:** Time trace of indirect degradation reporter. The oscillations of the repressilator, shown using the repressilator ($P_{Ltet}$-mVenus, blue trace), are used to induce a substrate ($P_{Ltet}$-Spo0A, SL89). The levels of the substrate are inferred by a Spo0A-dependent promoter ($P_{SpoIIE}$-CFP, red trace). The downstream reporter is not activated in the fast-growing cell ($\tau_{div} \approx 25$ min).

# B

# Materials and methods

## B.1 Chapter 2: Fluctuation timescales in micro-organisms

### B.1.1 Imaging protocol

For details on the chip preparation, cell preparation and microscopy image acquisition, see section B.2.2. The protocol to prepare the wafer of the microfluidic device is presented in section B.4.

The medium used for imaging contained contained: 1 x M9 salts, 0.2% (w/v) glucose, 0.1 mM

$CaCl_2$, 1mM $MgSO_4$, 20 $\mu$g/mL Uracil, 0.2g/L casamino acids (BD Biosciences) and 0.85 g/L pluronic F108 (Sigma Aldrich, included as a passivating agent).

### B.1.2 Data processing

For details on the segmentation and data analysis, see section B.2.3.

#### Autocorrelation estimate

The autocorrelation functions were calculated according to equations (B.3) and (B.4) in section B.2.3.

When provided, the large lag errors estimate [41] was calculated using:

$$\mathrm{SE}(r_k) = \sqrt{\frac{1}{N}\left(1 + 2\sum_{j=1}^{q\leq k} r_j^2\right)} \tag{B.1}$$

where $\mathrm{SE}(r_k)$ is the standard error estimated at time lag $k$, $N$ is the number of points used to estimate that particular lag, $q$ is the cutoff at which point the theoretical autocorrelation is considered 0 and $r_k$ is the theoretical autocorrelation function. We used the estimated autocorrelation as the best estimate of the theoretical autocorrelation.

### B.1.3 Stochastic simulations

The stochastic simulations were performed using the Gillespie algorithm [83] implemented in Matlab (The Mathworks).

### B.1.4 Plasmids and strains

The strains used in this chapter are listed in table B.1. The wild-type strain used was MG1655.

pLPT30 was constructed using isothermal assembly (ITA) [184]. Primers were ordered from Integrated DNA technologies (IDT) and PCR were performed with Accuprime Pfx (Life technologies) or Phusion polymerase (New England BioLabs) according to manufacturers' protocols. The plasmid was verified by DNA sequencing (Genewiz). The promoter was PCRed from the transcriptional reporter library [39]. The promoter was inserted into the chromosome using Tn7 integration [185] to generate LPT40.

LPT75 to LPT81 were generated by P1 phage transduction from strains SX1225 (CGSC 12780), SX1301 (CGSC 12780), SX1442 (CGSC12997), SX1512 (CGSC 13067), SX1526 (CGSC 13081), SX1589 (CGSC 13144) and SX1968 (CGSC 13523).

The construction of the NDL strains has been described in reference [36].

**Table B.1:** List of strains and plasmids used in chapter 2.

| Plasmid | | | |
|---|---|---|---|
| Strain | Parent | Operation | Genotype |
| pLPT30 | pPM136 | | miniR1 par+ $P_{rpsL}$-mCFP |
| Strains | | | |
| Strain | Parent | Operation | Genotype |
| MG1655 | | | 'Wild type' *E. coli* K12 strain, CGSC 7740 |
| LPT40 | NDL93 | Tn7 integration | $\Delta motA$ attTn7::$P_{rpsL}$-mVenus, pLPT30 |
| LPT74 | | P1 transduction | $glmS$::$P_{RNA1}$-mCherry/mKate2 $\Delta motA$ $seqA$::$seqA$-Venus-camR |
| LPT75 | | P1 transduction | $glmS$::$P_{RNA1}$-mCherry/mKate2 $\Delta motA$ $atpD$::$atpD$-Venus-camR |
| LPT76 | | P1 transduction | $glmS$::$P_{RNA1}$-mCherry/mKate2 $\Delta motA$ $clpP$::$clpP$-Venus-camR |
| LPT77 | | P1 transduction | $glmS$::$P_{RNA1}$-mCherry/mKate2 $\Delta motA$ $fadB$::$fadB$-Venus-camR |
| LPT78 | | P1 transduction | $glmS$::$P_{RNA1}$-mCherry/mKate2 $\Delta motA$ $katE$::$katE$-Venus-camR |
| LPT79 | | P1 transduction | $glmS$::$P_{RNA1}$-mCherry/mKate2 $\Delta motA$ $sulA$::$sulA$-Venus-camR |
| LPT81 | | P1 transduction | $glmS$::$P_{RNA1}$-mCherry/mKate2 $\Delta motA$ $nrdG$::$nrdG$-Venus-camR |
| NDL93 | MG1655 | | $\Delta motA$ |
| NDL162 | NDL93 | | $glmS$::$P_{RNA1}$-mCherry/mKate2 $\Delta motA$ |
| NDL205 | | | $glmS$::$P_{RNA1}$-mCherry/mKate2 $\Delta motA$ $lon$::$lon$-Venus-camR |
| NDL207 | | | $glmS$::$P_{RNA1}$-mCherry/mKate2 $\Delta motA$ $ligA$::$ligA$-Venus-camR |
| NDL208 | | | $glmS$::$P_{RNA1}$-mCherry/mKate2 $\Delta motA$ $gapA$::$gapA$-Venus-camR |

## B.2 Chapter 3: Synchronous long-term oscillations in a synthetic gene circuit

### B.2.1 Data and materials availability

Segmented and assembled single-cell traces files are accessible online. The masks files used for the microfluidic device master fabrication are available on request. The plasmids have been deposited to the Addgene plasmid repository, and the plasmids maps are available online.

### B.2.2 Imaging protocol

#### Chip preparation

Dimethyl siloxane monomer (Sylgard 184) was mixed in a 10:1 ratio with curing agent, defoamed, poured onto the silicon wafer, degassed for 1 hour and cured at 65°C for 1 hour. Individual chips were then cut and the inlets and outlets were punched with a biopsy puncher. Bonding to KOH-cleaned cover slips was ensured using oxygen plasma treatment (30 sec at 50 W and $O_2$ pressure at 170 mTorr) on the day the experiments start. The chips were then incubated at 95°C for at least 30 min to reinforce the bonding.

#### Cell preparation

*E. coli* strains were grown overnight in LB with appropriate antibiotics and diluted 1:100 ∼2-3 hours before the beginning of the experiments in imaging media, consisting of M9 salts, 10% (v/v) LB, 0.2% (w/v) glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1.5 $\mu$M thiamine hydrochloride and 0.85 g/L

Pluronic F-108 (Sigma Aldrich, included as a passivating agent). The cells were loaded into the device at $OD_{600}$ 0.2-0.4, and centrifuged on a custom-machined holder that could fit into a standard table-top centrifuge at 5000 *g* for 10 min to insert them into the side-channels. The feeding channels were connected to syringes filled with imaging media using Tygon tubing (VWR), and media was pumped using syringe pumps (New Era Pump System) initially at a high rate of 100 $\mu$L/min for 1 hour, to clear the inlets and outlets. The media was then pumped at 5-10 $\mu$L/min for the duration of the experiment and cells were allowed to adapt to the device for multiple hours before imaging was started.

### Microscopy and image acquisition

Images were acquired using a Nikon Ti inverted microscope equipped with a temperature-controlled incubator, an Orca R2 CCD camera (Hamamatsu), a 60X Plan Apo oil objective (NA 1.4, Nikon), an automated xy-stage (Ludl) and light engine LED excitation source (Lumencor). All experiments were performed at 37°C. Microscope control was done with MATLAB (Mathworks) scripts interfacing with $\mu$Manager [186]. Typical exposure was low (50-100 ms) in order to reduce photobleaching, and the reporter channels were acquired using 2x2 binning (CCD chip dimension of 1344 x 1024 pixels, effective pixel size of 129nm x 129nm). 16 bits TIFF images were taken every 5-8 minutes, and focal drift was controlled via the Nikon PerfectFocus system, as well as a custom routine based on *z*-stack images of a sacrificial position. The following filter sets were used for acquisition: GFP (Semrock GFP-3035B), RFP (Semrock mCherry-A), YFP (Semrock YFP-2427A) and CFP (Semrock CFP-2432A).

#### Conditioned medium

The conditioned medium was obtained by growing the strain used in the experiment until $OD_{600}$ = 2.0, and then the culture was rapidly sterilized with a 0.2$\mu$m filter and kept at 4°C until the experiment.

### IPTG

When indicated, the imaging medium was supplemented with isopropyl $\beta$-D-1-thiogalactopyranoside (IPTG) for the duration indicated in the figure.

## B.2.3 Data processing

#### Segmentation

Image segmentation and single-cell trace assembly were performed similarly to a previously described procedure [38]. Briefly, the segmentation was done using images from a bright, constituvely expressed ($P_{RNAI}$ promoter on the chromosome or on the plasmid) CFP or RFP. The rough channel boundaries were estimated, in order to reject out-of-channel cells, with a simple threshold followed by erosion, opening and dilation of the mask. The contrast of the fluorescent image was enhanced using a 'unsharp mask'. Then, the edges of the cells were detected using the Laplacian of Gaussian method. Cells joined by their poles (as indicated by objects with definite constrictions) were separated, and spurious non-cell objects were rejected using their size, orientation and shape. Finally, the boundaries were refined using opening, thickening and active contours. The parameters used for

these functions were optimized specifically for the combination of the strain, growth conditions and microscope setup. We will share the code used on request, but the specific parameters will need to be re-optimized depending on the exact setup.

We chose to follow only the cells at the top (closed end) of the channel (the 'mother' cell), as it made compiling the single-cell traces much easier as these cells stay in place for the duration of the experiment. Due to physical limitations of the setup, the segmentation mask was slightly mis-registered with respect to the 'data channel' (i.e. GFP or YFP). Each object was then registered to the proper channel before the data were extracted.

For the triple reporter strains (that did not contain a specific segmentation fluorophore), we combined the three reporters channels according to their signal-to-noise ratio to obtain an effective segmentation channel. In some cases, we used an alternative procedure where the whole channels were segmented, and the top 50 pixels were used as a 'cell' which represented one to two cells. Because the oscillations for these analyses were very slow and the cells in the channels are very close in phase (e.g. see Extended Data Figure 3.9g), the two methods gave very similar results to the normal procedure, but in general the 'channel' analysis worked more reliably. This analysis was only used in Figure 3.2b.

The background fluorescence was corrected by subtracting the median value of the fluorescent images (the cells represent a very small fraction of the image). We then estimated the concentration of fluorophore using the average of the background-substracted intensities inside the segmentation mask. Nearly identical results were obtained using the 'peak' intensity (median of the top 10 % of the pixel in the segmentation mask), but for simplicity we only report the results obtained with the

mean.

### Single-cell traces construction

The temporal information of the cell data (intensities, area, etc.) was then compiled into single-cell traces by matching the centroid of the cells from frame $n$ to frame $n+1$. As there was little drift, this procedure was very reliable, and we prevented spurious matches by setting an upper limit on the centroid distance. We identified cell divisions by sudden decreases in cell area; if the cell area dropped to less than 60% of its previous value, a division was called.

### Production rate estimation

To estimate the production rate of fluorophore, we used the derivative of the concentration, as it was more robust to errors in segmentation. Let $T$ be the total intensity, $A$ the area of the cell and $C$ the concentration. Since $T(t) = A(t)C(t)$,

$$\begin{aligned} \frac{dT}{dt} &= C\frac{dA}{dt} + A\frac{dC}{dt} \\ \frac{1}{A}\frac{dT}{dt} &= C\frac{1}{A}\frac{dA}{dt} + \frac{dC}{dt} \\ P(t) &= C(t)g(t) + \frac{dC}{dt} \end{aligned} \tag{B.2}$$

where $P(t) = \frac{1}{A}\frac{dT}{dt}$ is the normalized production rate and $g(t)$ is the growth rate (e.g. $g = \ln(2)/\tau_{div}$ for exponential growth, $A(t) = A_0 2^{t/\tau_{div}}$, $\tau_{div}$ being the doubling time). Equation (B.2) was used for estimating the production rate in the paper, with $g(t)$ estimated for each cell

cycle with the initial and final area and $dC/dt$ using Tikhonov regularization to enforce smoothness [187]. The normalization factor was kept as small as possible, but similar results were obtained for factors one order of magnitude smaller or larger. In practice, photobleaching or degradation of the fluorescent protein can affect the estimation of the production rate. One can account for these effects by using an effective half-life instead of $\tau_{div}$ (e.g. $\tau_{eff}^{-1} = \tau_{div}^{-1} + \tau_{photo}^{-1} + \tau_{deg}^{-1}$, where $\tau_{photo}$ and $\tau_{deg}$ are the photobleaching and degradation half-lives, respectively). These effects were negligible for GFP (even with the *asv* degradation tags), so we chose to report the production rate without correction.

### Autocorrelation function and power spectrum estimation

The autocorrelation functions were estimated by averaging the correlation functions of the individual cells, as it was more robust to outliers, and using the unbiased estimator. Similar functions were obtained by taking directly the autocorrelation of the population, but needed manual curation of the data to remove dead cells or filaments.

$$A(\tau) = \left\langle \frac{\langle (x_i(t) - \langle x_i(t) \rangle_t)(x_i(t+\tau) - \langle x_i(t) \rangle_t) \rangle_t}{\langle x_i(t)^2 \rangle_t} \right\rangle_i \tag{B.3}$$

where $x_i(t)$ is the production rate or the concentration (indicated in the figure caption) of the $i^{\text{th}}$ cell at time $t$. Averaging of the correlations functions of the cells was done taking into account the finite length of the time series (each cell has a different number of samples for a specific time lag). If $A_i$ is the autocorrelation of cell $i$,

$$\hat{A}[j] = \frac{\sum_i (L_i - j)\hat{A}_i[j]}{\sum_i (L_i - j)} \tag{B.4}$$

with $A(j\Delta t) = \hat{A}[j]$, $\Delta t$ the time between images, $j$ the discrete delay index and $L_i$ the number of points in time trace $i$. The brackets are used to emphasize discrete sampling. The autocorrelations were cropped to a constantly decreasing envelope to keep only time lags with good estimates. This resulted in correlation functions very similar to the ones obtained by using the biased estimator, albeit with a slightly larger envelope.

The power spectrum was then estimated by taking the discrete Fourier transform (DFT) of the windowed autocorrelation function [188, 189]:

$$P[k] = \mathrm{DFT}_N(a[m]) \tag{B.5}$$

where $\mathrm{DFT}_N$ is the $N$ point DFT and $a[m]$ is the windowed symmetric autocorrelation:

$$a[m] = \begin{cases} \hat{A}\left[|M - m|\right] w[m] & \text{for } 0 \leq m \leq 2M \\ 0 & \text{for } 2M < m \leq N \end{cases} \tag{B.6}$$

and $w[m]$ is a window function. Then,

$$P[k] = P(\omega)|_{\omega = 2\pi k/N} \tag{B.7}$$

$$P(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\theta) W(\omega - \theta) d\theta \tag{B.8}$$

where $X(\omega)$ is the power spectrum of the signal and $W(\omega)$ the Fourier transform of the window function. We are therefore sampling the power spectrum of the signal convolved with $W(\omega)$. This is a consistent estimator of the power spectrum (it converges to the actual power spectrum as the amount of data goes to infinity) [190]. We used a triangular window function to avoid negative spectral leakage, and the length of the window function ($2M$) was chosen to maximize the resolution without introducing too much noise (50-225 frames, depending on the period of the oscillations). The approximate resolution loss was indicated by a red line of width $2\pi/M$ ($1/(M\Delta t)$) in the figures.

### Period histograms and phase drift estimation

Peak-to-peak distances were evaluated by finding maxima using the *findpeaks* MATLAB function. The traces were first smoothed using a 3 or 5 points moving average and peaks were rejected if they were closer than 3 or 5 frames to avoid double counting, or smaller than the average of the trace. The peaks were then manually curated; this was especially useful for the noisy oscillators. Note that the average period was slighly shorter than the first maximum of the autocorrelation, most likely because longer periods have higher intensities and thus more weights in the correlation (but not in the period histogram).

The period histograms were made by using the peak-to-peak distance. The squared error on the $n^{\text{th}}$ period grew linearly with $n$, as expected for this type of oscillator undergoing a random walk in phase. We therefore used the coefficient of variation (CV, standard deviation divided by the mean) of the period as an indicator of phase drift; the normalization makes comparison between oscillators

of different frequencies straightforward.

Most of the strains had a phase drift of 30-35% per period; except for the repressilator without degradation but with the titration sponge, where it was only 14%. Since the variance increased linearly, we can express the variance for $n$ periods ($\sigma_n^2$) as a function of the variance for one ($\sigma_1^2$):

$$\begin{aligned} \sigma_n^2 &= n\sigma_1^2 \\ \sigma_n &= \sigma_1\sqrt{n} \\ &= \langle \text{period} \rangle \times \text{CV}\sqrt{n} \end{aligned}$$

Hence, it would take ∼13 periods (∼179 generations) to obtain a standard deviation of half a period.

Another measure of the phase drift is the *average* time to reach half a period of phase drift, or the average first passage time. This could be calculated by drawing randomly directly from the period histogram until the first time the phase drift is reached, because subsequent periods were exceptionally well approximated as independent (Figure 3.1f). This creates a distribution of first passage times, and after $10^5$ iterations, we converge on an average first passage time of $\sim$ 18 periods ($\sim$ 240 generations), again for the repressilator without degradation with titration (LPT64).

### B.2.4 Macroscopic image acquisition

#### IPTG synchronization and flask experiment

In order to synchronize the phase of the oscillators in the population, we diluted the strains in imaging medium supplemented with appropriate antibiotics and 1 mM IPTG so that they would be early exponential ($OD_{600} \sim 0.2$) 8 hours later ($\sim 1 : 10^6$) at 37°C. For the unsynchronized control, we did the same procedure but did not include IPTG. After, we diluted the cultures to $OD_{600}$ 0.05 every 50 min, while taking the fluorescent images of the (undiluted) flasks. The $OD_{600}$ of the imaged culture varied slightly, but the effect was negligible, as can be seen on the unsynchronized control (Figure 3.3b).

#### Photo acquisition

Photos were acquired using a digital camera setup equipped with emission filters and LEDs fluorescent excitation [191]. A custom written software controls a Canon T3i digital single lens reflex (DSLR) camera with a Canon EF-S 60mm USM lens, placed in front of a Starlight express filter wheel, and appropriate LEDs for excitation. A long exposition time of 10 s was used for the flask – enabling the use of small $OD_{600}$ – while the exposition time of the plates was 0.1-2 s.

#### Microscopy

Images for figure 3c were acquired using a Olympus MVX10 Macroview microscope equipped with a Zeiss AxioCam MRc camera. Flurophores were excited using a Lumen200 fluorescence illumina-

tion system (Prior Scientific) and we used the following Olympus filter sets: CFP (U-M40001XL), YFP (U-M49003XL) and mCherry (U-M49008XL).

### B.2.5 Strains

#### Plasmids

All plasmids used in this study are listed in Table B.2.

The constructs were made using either isothermal assembly (ITA) [184] or 'round-the-horn' site-directed mutagenesis (http://openwetware.org/wiki/%27Round-the-horn_site-directed_mutagenesis). Primers and gBlocks were ordered from Integrated DNA technologies (IDT) and PCR were performed with Accuprime Pfx (Life technologies) or Phusion polymerase (New England BioLabs) according to manufacturers' protocols. All plasmids were verified by DNA sequencing (Genewiz). In the text, *laa* is used as a shorthand for the native *ssrA*-(LAA) (AANDENYALAA) and *asv* for the synthetic *ssrA*-(ASV) (AANDENYAASV) degradation tag.

pZE21-GFPasv refers to the original reporter [45], while pZS1-lTlrLLtCL to the original repressilator plasmid. Both were kind gifts of M. Elowitz. The $P_{LlacO-1}$ and $P_{LtetO-1}$ are the hybrid promoters orginally used by switching the binding sites of bacteriophage $\lambda$ promoter $P_L$ with lacO1 and tetO2 binding sequences, respectively [192]. They are abbreviated $P_{Llac}$ and $P_{Ltet}$ in the text for simplicity. $P_R$ refers to the native bacteriophage $\lambda$ promoter.

pLPT20 was constructed by assembling *mVenus* (kind gift of the P. Cluzel, Ref. [193] with A206K), $P_{RNA1}$ and *cfp* (kind gift of the P. Cluzel) into a temporary reporter plasmid (pZE21-GFPasv deriva-

tive), PCR-amplified and inserted into pZS1-lTlrLLtCL plasmid digested at the unique AatII site.

pLPT41, pLPT42, pLPT43 and pLPT57 were constructed by 'round-the-horn' PCR from pZE21-GFPasv removing different part of the construct. For pLPT57, we merely removed the *asv* degradation tag. The *pep* in pLPT43 refers to keeping a peptide consisting of the N-terminal 40 amino acids fragment of *gfpmut3* (MRKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYG) followed by the *asv* degradation tag. In pLPT41, we kept the promoter, ribosome binding site (RBS) and terminator intact while removing the *gfp*. For pLPT42, we also removed the promoter.

pLPT47 was assembled from a synthesized gBlock and fragment of pLPT1. The $P_{const}$ refers to a synthetic, intermediate strength and constitutive promoter from the iGEM Registry of standard biological parts (http://parts.igem.org/, BBa_J23107, tttacggctagctcagccctaggtattatgctagc). We used a Y66L mutation on the GFP to make it non-fluorescent [194] and therefore compatible with the YFP reporter of the integrated repressilator.

pLPT107 was constructed by inserting synthesized gBlocks as well as PCR-amplified fragments into pLPT20 backbone. $P_R$-*mKate2* was added and the $P_{RNA1}$ promoter was changed to a $P_{Llac}$. The *mKate2* was a gift from D. Landgraf and consists of the *mCherry* N-terminal 11 amino acids followed by the *mKate2* sequence (used to improve translational efficiency).

pLPT119 was assembled from PCR fragments of pLPT20, simply removing the degradation tags on the repressors.

### Bacterial strains

All *E. coli* strains used in this study are listed in Table B.3. The wild-type strain was MC4100.

DHL708 was a kind gift of D. Landgraf [76] and was built by deleting the *clpPX* operon with lambda-Red mediated homologous recombination. The FRT-flanked Kan cassette was then flipped out using the FLP recombinase (pCP20).

NDL319 was constructed by inserting the $P_{RNA1}$-*mKate2* hybrid cassette close to (while preserving) the neutral Tn7 insertion locus [185].

**Table B.2:** Plasmid list

| Plasmid | Parent | Ori | Antibiotic | Genotype |
|---|---|---|---|---|
| pDHL474 | Gift of D. Landgraf | ColE1 | Kan | $P_{LtetO1}$-*mCherry-asv* |
| pDHL490 | Gift of D. Landgraf | ColE1 | Kan | $P_{LtetO1}$-*mCherry-laa* |
| pZE21-GFPasv | Gift of M. Elowitz | ColE1 | Kan | Reporter, $P_{LtetO1}$-*gfpmut3-asv* |
| pZS1-lTlrLLtCL | Gift of M. Elowitz | pSC101 | Amp | Repressilator, $P_{LtetO1}$-*cI-laa*, $P_R$-*lacI-laa*, $P_{LlacO1}$-*tetR-laa* |
| pLPT20 | pZS1-lTlrLLtCL | pSC101 | Amp | Integrated repressilator, $P_{LtetO1}$-*cI-laa*, $P_R$-*lacI-laa*, $P_{LlacO1}$-*tetR-laa*, $P_{LtetO1}$-*venus*, $P_{RNA1}$-*cfp* |
| pLPT41 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$ |
| pLPT42 | pZE21-GFPasv | ColE1 | Kan | empty |
| pLPT43 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-*pep-asv* |
| pLPT47 | pZE21-GFPasv | ColE1 | Kan | $P_{const}$-*darkgfp-asv* |
| pLPT57 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-*gfpmut3* |
| pLPT107 | pLPT20 | pSC101 | Amp | Triple reporter repressilator, $P_{LtetO1}$-*cI-laa*, $P_R$-*lacI-laa*, $P_{LlacO1}$-*tetR-laa*, $P_{LtetO1}$-*venus*, $P_{LlacO1}$-*cfp*, $P_R$-*mKate2* |
| pLPT119 | pLPT20 | pSC101 | Amp | Integrated repressilator no degradation tags, $P_{LtetO1}$-*cI*, $P_R$-*lacI*, $P_{LlacO1}$-*tetR*, $P_{LtetO1}$-*venus*, $P_{RNA1}$-*cfp* |
| pLPT144 | LPT42 | ColE1 | Kan | $P_{LlacO1}$, $P_R$ |
| pLPT145 | LPT41 | ColE1 | Kan | $P_{LtetO1}$, $P_{LlacO1}$, $P_R$ |
| pLPT146 | LPT42 | ColE1 | Kan | $P_{LlacO1}$ |
| pLPT147 | LPT42 | ColE1 | Kan | $P_R$ |
| pLPT148 | LPT41 | ColE1 | Kan | $P_{LlacO1}$, $P_{LtetO1}$ |
| pLPT149 | LPT41 | ColE1 | Kan | $P_{LtetO1}$, $P_R$ |
| pLPT151 | pDHL490 | ColE1 | Kan | $P_{LtetO1}$-*mCherry-laa*, $P_{const}$-*darkgfp-asv* |

**Table B.3:** Strain list and summary of experimental results

| Strain | Parent | Genotype | Plasmids | Period (gen) | Correlation after one period | Generation time (min) |
|---|---|---|---|---|---|---|
| DHL708 | MC4100 | $\Delta clpXP$ | | | | |
| LPT25 | MC4100 | | pLPT20 | 5.6 | 0.18 | 26 |
| LPT44 | LPT25 | | pLPT20, pLPT41 | 5 | 0.2 | 29 |
| LPT45 | LPT25 | | pLPT20, pLPT42 | 5.5 | 0.14 | 27 |
| LPT46 | LPT25 | | pLPT20, pLPT43 | 4.2 | 0.16 | 27 |
| LPT53 | LPT25 | | pLPT20, pLPT47 | 2.3 | 0.09 | 27 |
| LPT54 | LPT25 | | pLPT20, pDHL474 | 2.6 | 0.07 | 27 |
| LPT55 | LPT25 | | pLPT20, pDHL490 | 5.7 | 0.2 | 28 |
| LPT60 | MC4100 | | pZS1-lTlrLLtCL, pLPT57 | 5 | 0.18 | 26 |
| LPT61 | DHL708 | $\Delta clpXP$ | pLPT20 | 10.4 | 0.09 | 25 |
| LPT64 | LPT61 | $\Delta clpXP$ | pLPT20, pDHL474 | 13.9 | 0.5 | 27 |
| LPT113 | DHL708 | $\Delta clpXP$ | pLPT107 | 14 | 0.2 | 26 |
| LPT117 | LPT113 | $\Delta clpXP$ | pLPT107, pLPT43 | 14.8 | 0.5 | 28 |
| LPT118 | LPT113 | $\Delta clpXP$ | pLPT107, pDHL474 | 14.9 | 0.4 | 29 |
| LPT120 | MC4100 | | pLPT119 | 10 | 0.07 | 25 |
| LPT124 | LPT120 | | pLPT119 pLPT41 | 12.9 | 0.3 | 25 |
| LPT125 | LPT120 | | pLPT119 pLPT43 | | | |
| LPT127 | DHL708 | $\Delta clpXP$ | pLPT107, pLPT41 | 14.4 | 0.5 | 28 |
| LPT143 | LPT61 | $\Delta clpXP$ | pLPT20, pLPT41 | | | |
| LPT150 | MC4100 | att::Tn7 $P_{clpPX}$-*clpP-msfgfp* | | | | |
| LPT152 | LPT61 | $\Delta clpXP$ | pLPT20, pLPT145 | 15.6 | 0.48 | 26 |

**Table B.3:** Strain list and summary of experimental results (continued)

| Strain | Parent | Genotype | Plasmids | Period (gen) | Correlation after one period | Generation time (min) |
|---|---|---|---|---|---|---|
| LPT153 | LPT61 | $\Delta clpXP$ | pLPT20, pLPT146 | | | |
| LPT154 | LPT61 | $\Delta clpXP$ | pLPT20, pLPT147 | | | |
| LPT155 | LPT61 | $\Delta clpXP$ | pLPT20, pLPT148 | | | |
| LPT156 | LPT61 | $\Delta clpXP$ | pLPT20, pLPT149 | 13.3 | 0.47 | 25 |
| LPT157 | LPT61 | $\Delta clpXP$ | pLPT20, pLPT144 | | | |
| LPT158 | MC4100 | | pDHL490 | | | |
| LPT159 | LPT150 | att::Tn7 $P_{clpPX}$-*clpP-mgfpmut3* | pLPT20 | 5.4 | 0.17 | 25 |
| LPT165 | LPT159 | att::Tn7 $P_{clpPX}$-*clpP-mgfpmut3* | pLPT20, pDHL474 | 2.8 | 0.1 | 26 |
| LPT172 | MC4100 | | pLPT151 | | | |
| MC4100 | | | | | | |
| NDL319 | MC4100 | att::Tn7 $P_{RNA1}$-*mKate2* | | | | |
| NDL332 | NDL319 | att::Tn7 $P_{RNA1}$-*mKate2* | pZE21-GFPasv, pZS1-lTlrLLtCL | 2.4 | 0.15 | 25 |

## B.3 Chapter 4: Protein degradation kinetics in single bacterial cells

### B.3.1 Imaging protocol

For details on the chip preparation, cell preparation and microscopy image acquisition, see section B.2.2. The protocol to prepare the wafer of the microfluidic device is presented in section B.4.

### B.3.2 Data processing

For details on the segmentation and data analysis, see section B.2.3.

#### Estimation of production and degradation rates

We consider cells growing and dividing exponentially, and use exponential dilution as an approximation for binomial partitioning of molecules at cell division. We express all parameters in concentration (i.e. molecules per unit volume, or per area if the bacterium radius is constant). We can therefore write:

$$\begin{aligned} x &\xrightarrow{p(t)} x+1 \\ x &\xrightarrow{\beta(x,t)x} x-1 \end{aligned} \tag{B.9}$$

where is $x$ is the concentration of molecules, $p(t)$ the time-dependent production rate per area, and $\beta(x,t)$ the total elimination rate per molecule. The total elimination rate can be decomposed as:

$\beta(x,t) = \beta_{deg}(x,t) + \beta_{dil}(t) + \beta_{bleach}$. We can then write down:

$$\frac{d\langle x(t)\rangle}{dt} = p(t) - \beta(x,t)x(t) \tag{B.10}$$

To calculate the production rate of the slippery construct, we measured the production rate of the CFP, with $\beta(x,t) = \beta_{dil}(t) = g(t)$, where $g(t)$ is the growth rate of the cell. The growth rate was calculated for each cell division, using $g(t) = \ln(A_f/A_i)/\Delta t$, where $A_f$ and $A_i$ are the final and initial cell area, and $\Delta t$ the time interval for that particular cell division. This gives the same equation as (B.2) in section B.2.3.

To calculate the degradation rates, we used the same equation, but with $p(t)$ calculated using the CFP and converted in the proper fluorescence units (e.g. using Figure A.2), and $\beta(x,t) = \beta_{deg}(x,t) + \beta_{dil}(t) + \beta_{bleach}$. The photobleaching rate was estimated with a half-life of 10 cell generations, using the slippery construct in the *clpXP* mutant strain.

### B.3.3 Plasmids and strains

The plasmids and strains used in this chapter are listed in tables B.4 and B.5. The wild-type strain was MC4100.

The constructs were made using either isothermal assembly (ITA) [184] or 'round-the-horn' site-directed mutagenesis (http://openwetware.org/wiki/%27Round-the-horn_site-directed_mutagenesis). Primers and gBlocks were ordered from Integrated DNA technologies (IDT) and PCR were performed with Accuprime Pfx (Life technologies) or Phusion poly-

merase (New England BioLabs) according to manufacturers' protocols. All plasmids were verified by DNA sequencing (Genewiz). In the text, *laa* is used as a shorthand for the native *ssrA*-(LAA) (AANDENYALAA) and *asv* for the synthetic *ssrA*-(ASV) (AANDENYAASV) degradation tag. The *ssrA*-(NY) refers to the *ssrA* tag extension [152] (AANDENYNYALAA).

pZE21-GFPasv refers to the original reporter [45], while pZS1-lTlrLLtCL to the original repressilator plasmid. Both were kind gifts of M. Elowitz. The $P_{LlacO-1}$ and $P_{LtetO-1}$ are the hybrid promoters orginally used by switching the binding sites of bacteriophage $\lambda$ promoter $P_L$ with lacO1 and tetO2 binding sequences, respectively [192]. They are abbreviated $P_{Llac}$ and $P_{Ltet}$ in the text for simplicity. $P_R$ refers to the native bacteriophage $\lambda$ promoter.

The deletion strains were constructed by P1 phage transduction from the Keio collection [195] (obtained from CGSC). The FRT-flanked Kan cassette was then flipped out using the FLP recombinase (pCP20).

NDL319 was constructed by inserting the $P_{RNA1}$-*mKate2* hybrid cassette close to (while preserving) the neutral Tn7 insertion locus (noted as *glmS*).

The chromosomal insertions were obtained using Tn7 integration [185].

BBa_J23114 is a constitutive promoter from the iGEM Registry of standard biological parts (`http://parts.igem.org/`. The natural promoter used were obtained from the complete intergenic region and includes all reporter transcription start sites.

The dark fluorescent proteins were obtained by using a Y66L mutation them non-fluorescent [194].

The *slip* linker represents $GA_{r30}$-TitinI27-$GA_{r30}$, where $GA_{r30}$ is (AGAGGGAGAG GAG-

GAGGGAG AGGAGAGGAG) and based on reference [132].

**Table B.4:** List of strains used in chapter 4

| Strain | Parent | Genotype | Plasmids |
|---|---|---|---|
| DHL708 | MC4100 | $\Delta clpXP$ | |
| DHL776 | MC4100 | $\Delta clpX$ | |
| DHL971 | MC4100 | clpP-mVenus FRT-KanR-FRT | |
| NDL319 | MC4100 | $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid | |
| TB781 | MC4100 | $clpP$::Cat | |
| LPT25 | MC4100 | MC4100 | pLPT20 |
| LPT53 | LPT25 | MC4100 | pLPT20, pLPT47 |
| LPT133 | MC4100 | $\Delta ybaQ$ | pLPT20, pLPT47 |
| LPT134 | MC4100 | $\Delta ydcN$ | pLPT20, pLPT47 |
| LPT139 | MC4100 | $\Delta ybaQ$ | pLPT20 |
| LPT140 | MC4100 | $\Delta ydcN$ | pLPT20 |
| LPT185 | NDL319 | $clpS$::FRT-Kan-FRT, $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid | pLPT20 |
| LPT186 | NDL319 | $clpB$::FRT-Kan-FRT, $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid | pLPT20 |
| LPT187 | NDL319 | $sspB$::FRT-Kan-FRT, $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid | pLPT20 |
| LPT188 | NDL319 | $hslV$::FRT-Kan-FRT, $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid | pLPT20 |
| LPT189 | MC4100 | $clpA$::FRT-Kan-FRT | pLPT20 |
| LPT190 | TB781 | $clpP :: Cat$ | pLPT20 |
| LPT191 | DHL776 | $\Delta clpX$ | pLPT20 |
| LPT198 | LPT185 | $clpS$::FRT-Kan-FRT,$glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid | pLPT20, pLPT95 |
| LPT199 | LPT186 | $clpB$::FRT-Kan-FRT, $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid | pLPT20, pLPT95 |
| LPT200 | LPT187 | $sspB$::FRT-Kan-FRT, $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid | pLPT20, pLPT95 |
| LPT201 | LPT188 | $hslV$::FRT-Kan-FRT, $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid | pLPT20, pLPT95 |
| LPT202 | LPT189 | $clpA$::FRT-Kan-FRT | pLPT20, pLPT95 |
| LPT203 | LPT190 | $clpP :: Cat$ | pLPT20, pLPT47 |
| LPT204 | LPT191 | $\Delta clpX$ | pLPT20, pLPT95 |
| LPT228 | NDL319 | attTn7:: $P_{Ltet01}$-mVenus, $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid | |
| LPT232 | LPT228 | attTn7:: $P_{Ltet01}$-mVenus, $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid | pLPT196, pLPT226 |
| LPT241 | NDL319 | $\Delta sspB$, $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid | pLPT196, pLPT227 |
| LPT248 | LPT228 | attTn7:: $P_{Ltet01}$-mVenus, $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid | pLPT196, pSL55 |
| LPT263 | LPT228 | attTn7:: $P_{Ltet01}$-mVenus, $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid | pLPT196, pLPT251 |

**Table B.4:** (continued) List of strains used in chapter 4

| | | | |
|---|---|---|---|
| LPT266 | NDL319 | *clpP*::*clpP*-mVenus, *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid | pLPT196, pLPT226 |
| LPT269 | NDL319 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid, attTn7:: $P_{clpPX}$-*clpPX* | pLPT196, pLPT226 |
| LPT309 | NDL319 | attTn7::$P_{Ltet01}$-mVenus, *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid | pLPT305 |
| LPT310 | NDL319 | attTn7::$P_{Ltet01}$-mVenus, *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid | pLPT307 |
| LPT311 | SL83 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid, lacIq, tetR, Δ*sspB* | pSL61 |
| LPT312 | LPT228 | attTn7::$P_{Ltet01}$-mVenus, *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid | pLPT196, pLPT306 |
| LPT323 | NDL319 | Δ*lon*, *sulA*::FRT-Kan-FRT | pLPT20 |
| LPT324 | NDL319 | Δ*lon*, *sulA*::FRT-Kan-FRT | pLPT20, pLPT95 |
| SL46 | NDL319 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid, BBa_J23114-QUEEN-7u | pLPT196, pDHL490 |
| SL49 | LPT228 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid, attTn7::$P_{Ltet01}$-mVenus | pLPT196, pSL49 |
| SL52 | LPT228 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid, attTn7::$P_{Ltet01}$-mVenus | pLPT196, pSL52 |
| SL53 | | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid lacIq tetR Δ*sspB* | pLPT196, pSL52 |
| SL57 | LPT228 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid, attTn7::$P_{Ltet01}$-mVenus | pLPT196, pSL57 |
| SL61 | NDL319 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid | pLPT196, pSL61 |
| SL70 | | Δ*clpXP* | pLPT196, pSL61 |
| SL83 | | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid, lacIq, tetR | |
| SL86 | SL83 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid, lacIq, tetR | pSL61 |
| SL89 | LPT228 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid, attTn7::$P_{Ltet01}$-mVenus | pSL89, pLPT290 |
| SL90 | LPT228 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid, attTn7::$P_{Ltet01}$-mVenus | pSL89, pLPT284 |
| SL91 | LPT228 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid, attTn7::$P_{Ltet01}$-mVenus | pSL89, pLPT285 |
| SL123 | NDL319 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid | pSL123, pSL124 |
| SL124 | NDL319 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid | pSL123, pSL125 |
| SL125 | NDL319 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid | pSL123, pSL126 |
| SL148 | NDL319 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid | pSL148, pSL124 |
| SL149 | NDL319 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid | pLPT226, pSL124 |
| SL153 | NDL319 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid | pSL153, pSL124 |
| SL154 | NDL319 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid | LPT196, pSL154 |
| SL155 | NDL319 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid | LPT196, pSL155 |
| SL156 | NDL319 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid | pSL153, pSL156 |
| SL157 | NDL319 | *glmS*:: $P_{RNAi}$-mCherry/mKate2 hybrid | pSL153, pSL157 |

**Table B.4:** (continued) List of strains used in chapter 4

| | | | |
|---|---|---|---|
| SL158 | SL151 | $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid, $\Delta rpoS$ | pSL153, pSL124 |
| SL159 | SL127 | $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid, $\Delta sspB$ | pSL153, pSL124 |
| SL160 | SL129 | $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid, $\Delta clpXP$ | pSL153, pSL124 |
| SL162 | SL127 | $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid $\Delta sspB$ | pSL148, pSL124 |
| SL163 | SL127 | $glmS$:: $P_{RNA1}$-mCherry/mKate2 hybrid $\Delta sspB$ | pLPT226, pSL124 |

**Table B.5:** List of plasmids used in chapter 4.

| Plasmid | Parent | Ori | Antibiotic Resistance | Genotype |
|---|---|---|---|---|
| pZE21-GFPasv | Gift of Elowitz | ColE1 | Kan | Reporter,$P_{Ltet01}$-GFPmut3-$asv$ |
| pZS1-lTlrLLtCL | Gift of Elowitz | pSC101 | Amp | $P_{LtetO1}$-$cI$-$ssrA$, $P_R$-$lacI$-$ssrA$, $P_{LlacO1}$-$tetR$-$ssrA$ |
| pDHL474 | Gift of Dirk Landgraf | ColE1 | Kan | $P_{LtetO1}$1-mCherry-$asv$ |
| pDHL490 | Gift of Dirk Landgraf | ColE1 | Kan | $P_{LtetO1}$1-mCherry-$ssrA$ |
| pLPT20 | pZS1-lTlrLLtCL | pSC101 | Amp | $P_{LtetO1}$-$cI$-$ssrA$, $P_R$-$lacI$-$ssrA$, $P_{LlacO1}$-$tetR$-$ssrA$,$P_{Ltet01}$-Venus, $P_{RNA1}$-CFP |
| pLPT47 | pZE21-GFPasv | ColE1 | Kan | Pconst-darkGFP-$asv$ |
| pLPT95 | pLPT47 | ColE1 | Cam | Pconst-darkGFP-$asv$ |
| pLPT196 | pLPT119 | pSC101 | Amp | $P_{LtetO1}$-$cI$, $P_R$-$lacI$, $P_{LlacO1}$-$tetR$ |
| pLPT226 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-mCFP-$ssrA$ |
| pLPT227 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-mVenus-$ssrA$ |
| pLPT251 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-$rpoS$750-mCFP |
| pLPT284 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-$IscS_{1-11}$-$spo0A$ |
| pLPT285 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-$spo0A$-$SulA_{20C}$ |
| pLPT290 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-$spo0A$ |
| pLPT305 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-mCFP-$rpoS$750 |
| pLPT306 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-mCFP-$RepA_{1-15}$ |
| pLPT307 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-mCFP-$MuA_{2-11}$ |
| pSL49 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-mVenus-$SulA_{20C}$ |
| pSL52 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-mVenus-$ssrA$(NY) |
| pSL55 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-$IscS_{1-11}$-mCFP |
| pSL57 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-$dps$-mVenus |
| pSL61 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-CFP-$slip$-mVenus-$ssrA$ |
| pSL89 | pLPT196 | pSC101 | Amp | $P_{LtetO1}$-$cI$, $P_R$-$lacI$, $P_{LlacO1}$-$tetR$, $P_{spoIIE}$-CFP |
| pSL123 | pSL61 | ColE1 | Kan | $P_{LtetO1}$-CFP-$slip$-darkmVenus-$ssrA$ |
| pSL124 | pLPT196 | pSC101 | Amp | $P_{LtetO1}$-$cI$, $P_R$-$lacI$, $P_{LlacO1}$-$tetR$, $P_{clpXP}$-mVenus |
| pSL125 | pLPT196 | pSC101 | Amp | $P_{LtetO1}$-$cI$, $P_R$-$lacI$, $P_{LlacO1}$-$tetR$, $P_{rpoE}$-mVenus |
| pSL126 | pLPT196 | pSC101 | Amp | $P_{LtetO1}$-$cI$, $P_R$-$lacI$, $P_{LlacO1}$-$tetR$, $P_{rpoH}$-mVenus |
| pSL148 | pLPT226 | ColE1 | Kan | $P_{LtetO1}$-mCFP |
| pSL153 | pSL123 | ColE1 | Kan | $P_{LtetO1}$-CFP-$slip$-darkmVenus |
| pSL154 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-$IscS_{1-11}$-mNeonGreen |
| pSL155 | pZE21-GFPasv | ColE1 | Kan | $P_{LtetO1}$-$rpos$750-mNeonGreen |
| pSL156 | pLPT196 | pSC101 | Amp | $P_{LtetO1}$-$cI$, $P_R$-$lacI$, $P_{LlacO1}$-$tetR$, $P_{dnaK}$-mVenus |
| pSL157 | pLPT196 | pSC101 | Amp | $P_{LtetO1}$-$cI$, $P_R$-$lacI$, $P_{LlacO1}$-$tetR$, $P_{katE}$-mVenus |

## B.4 Microfluidic master fabrication

Fabrication of the *E. coli* mother machine was carried out using standard UV photolithography in a clean room environment. The device was designed using AutoCad, and quartz-chrome photomasks were ordered from Toppan Inc. and from the Center for Nanoscale Systems at Harvard University. We modified the fabrication procedure from the method described in the original report on the mother machine. The fabrication requires three independent layers: an Su8 'base' coat, cell channels, and feeding channels. The process parameters below should be regarded as useful starting points, and all feature dimensions must be empirically verified to ensure that they meet experimental requirements. The tolerances for this device are stringent; a difference of $\pm 0.5$ $\mu$m in the height or width of a cell channel will compromise function. Note: for all spin coater steps described below, the following shorthand notation is used: speed (rpm)/acceleration (rpm/sec)/time (sec).

### B.4.1 First Layer: Su8 Base Coat

The first layer of the master contains no features. It consists of a thin, uniform layer of completely cured Su8. This layer is intended to serve as an adhesive to improve the retention of the cell channels deposited in the second layer. Without this base, the cell channels are easily removed from the wafer surface during development.

1. Place a new 3″ Si wafer (we used 380 $\mu$m TEST grade wafers from University Wafer) in a dish of fresh acetone. Sonicate at high power for 5 minutes.
2. Sequentially rinse the wafer with streams of methyl alcohol (MeOH), isopropyl alcohol (IPA) and $H_2O$ ($\sim$ 10 seconds per solvent).

3. Place wafer on 2″ spin chuck and spin seconds at 500 rpm.
4. While spinning, sequentially rinse the wafer with streams of MeOH, IPA and $H_2O$.
5. Spin wafer 1 minute at 3,000 rpm to dry.
6. Dehydrate wafer for 15 minutes on a hot plate set to 150-200°C.
7. Set spin program to: Step 1: 500/100/10, Step 2: 3000/300/60.
8. Place the dehydrated wafer onto the spin coater chuck and dispense a small (cover ∼2/3 of the wafer surface) amount of Su8 2000.5 photoresist (Microchem) with a pipette. Run the spin program. This should result in a coat of ∼0.5 $\mu$m.
9. Soft bake wafer (in order) for 1 minute at 65°C, 1 minute at 95°C, 1 minute at 65°C.
10. Expose wafer for 5 seconds with no mask at 25 mW/cm$^2$ (I-line).
11. Post exposure bake the wafer (in order) for 1 minute at 65°C, 1 minute at 95°C, 1 minute at 65°C.
12. Do not develop or hard bake the wafer. Instead, proceed directly to the second layer protocol.

### B.4.2 Second Layer: Cell Channels

This set of steps lays down the channels that house the cells in the final device. The tolerances for this layer are very stringent; the exposure dose and contact between mask and wafer must be optimized. We recommend trying a range of exposure parameters to ensure that a useful device is obtained. We also stress the importance of the very long post exposure bake time in the process below. In our hands, this greatly improves the retention of the cell channels during development.

1. Set spin program to: Step1: 500/100/10, Step 2: 2000/300/60.

2. Place the wafer onto the spin coater chuck and dispense a small (cover 2/3 of the wafer surface) amount of Su8 2001 photoresist with a pipette. Run the spin program. This should result in a coat of $\sim$1.5 $\mu$m.
3. Soft bake wafer (in order) for 1 minute at 65°C, 3 minutes at 95°C, 1 minute at 65°C.
4. Expose wafer for 0.75 seconds (25 mW/cm$^2$, I-line) through cell channel mask in vacuum contact mode.
5. Bake wafer for 1 minute at 65°C, 20 minutes at 95°C, 1 minute at 65°C.
6. Develop wafer for 30 seconds with very gentle agitation in Su8 developer.
7. Rinse wafer for 10 seconds with IPA.
8. Check completeness of development process. If undeveloped Su8 remains on the wafer (other than the desired cell channels) repeat developer treatment for 10 seconds.
9. Hard bake wafer for 10 minutes at 150°C.
10. Verify channel height using a profilometer. The expected height is 1.5 $\mu$m. If the channel dimensions lie outside of your expected tolerance bounds, the process must be repeated with modified spin coating parameters.

### B.4.3 Third Layer: Feeding Channels

The final layer of the device forms the medium flow channels. The dimensions of these features are not critical: we have used feeding channels of widely varying dimension to similar effect. The alignment is sensitive to large errors, however. The alignment between feeding channels and cell channels must be accurate (down to a couple of microns) in order to ensure that the cell channels are of the desired final length.

1. Set spin program to: Step 1: 500/100/10, Step 2: 5000/300/60.

2. Place the wafer onto the spin coater chuck and dispense a small (cover $\sim 2/3$ of the wafer surface) amount of Su8 2025 photoresist with a pipette being careful not to introduce bubbles. Run the spin program. This should result in a coat of $\sim 15$ $\mu$m.

3. Soft bake the wafer (in order) for 1 minute at 65°C, 4 minutes at 95°C, 1 minute at 65°C.

4. With an Su8-developer-soaked swab, clean the newly-deposited photoresist off of the alignment marks to make them visible for the alignment process.

5. Soft bake the wafer (in order) for 1 minute at 65°C, 4 minutes at 95°C, 1 minute at 65°C.

6. Align feeding channel mask to the alignment marks on the wafer. Apply vacuum contact and check alignment again. If the vacuum application skewed the alignment, repeat the alignment process.

7. Expose wafer for 10 seconds (25 mW/cm$^2$, I-line) through aligned feeding channel mask.

8. Bake wafer for 1 minute at 65°C, 4 minutes at 95°C, 1 minute at 65°C.

9. Develop wafer for 1.5 minutes in Su8 Developer with mild agitation.

10. Rinse wafer for 10 seconds in IPA. Check to ensure that the development is finished. If undesired photoresist remains, develop again for 20 seconds.

11. Hard bake wafer for 15 minutes at 150°C.

12. Verify channel height using a profilometer. The expected height is 15 $\mu$m. If the channel dimensions lie outside of your expected tolerance bounds, the process must be repeated with modified spin coating parameters.

# References


[1] Paulsson, J. Models of stochastic gene expression. *Phys Life Rev* **2**, 157–175 (2005).

[2] Maheshri, N. & O'Shea, E. K. Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu Rev Biophys Biomol Struct* **36**, 413–34 (2007).

[3] Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226 (2008).

[4] Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* **467**, 167–73 (2010).

[5] Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as a phenotypic switch. *Science* **305**, 1622–1625 (2004).

[6] Kint, C. I., Verstraeten, N., Fauvart, M. & Michiels, J. New-found fundamentals of bacterial persistence. *Trends Microbiol* **20**, 577–585 (2012).

[7] Yamanaka, Y., Lanner, F. & Rossant, J. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* **137**, 715–724 (2010).

[8] Sharma, S. V. *et al.* A Chromatin-Mediated Reversible Drug-Tolerant State in Cancer Cell Subpopulations. *Cell* **141**, 69–80 (2010).

[9] Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004).

[10] Lestas, I., Paulsson, J., Ross, N. & Vinnicombe, G. Noise in gene regulatory networks. *Automatic Control, IEEE* 189–200 (2008).

[11] Lestas, I., Vinnicombe, G. & Paulsson, J. Fundamental limits on the suppression of molecular fluctuations. *Nature* **467**, 174–178 (2010).

[12] Hilfinger, A. & Paulsson, J. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc Natl Acad Sci U S A* **108**, 12167–12172 (2011).

[13] Hilfinger, A., Chen, M. & Paulsson, J. Using Temporal Correlations and Full Distributions to Separate Intrinsic and Extrinsic Fluctuations in Biological Systems. *Phys Rev Lett* **109**, 248104 (2012).

[14] Hilfinger, A., Norman, T. M. & Paulsson, J. Exploiting Natural Fluctuations to Identify Kinetic Mechanisms in Sparsely Characterized Systems. *Cell Systems* **2**, 251–259 (2016).

[15] Hilfinger, A., Norman, T. M., Vinnicombe, G. & Paulsson, J. Constraints on Fluctuations in Sparsely Characterized Biological Systems. *Phys Rev Lett* **116**, 1–5 (2016).

[16] Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005).

[17] Yu, J., Xiao, J., Ren, X., Lao, K. & Xie, X. S. Probing gene expression in live cells, one protein molecule at a time. *Science* **311**, 1600–1603 (2006).

[18] Cai, L., Friedman, N. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).

[19] Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nat Genet* **31**, 69–73 (2002).

[20] Zenklusen, D., Larson, D. R. & Singer, R. H. Single-RNA counting reveals alternative modes of gene expression in yeast. TL - 15. *Nature structural & molecular biology* **15 VN - r**, 1263–1271 (2008).

[21] Blake, W. J., KAErn, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).

[22] Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4**, e309 (2006).

[23] Paré, A. *et al.* Visualization of Individual Scr mRNAs during Drosophila Embryogenesis Yields Evidence for Transcriptional Bursting. *Curr Biol* **19**, 2037–2042 (2009).

[24] Gottschling, D. E., Aparicio, O. M., Billington, B. L. & Zakian, V. A. Position effect at S. cerevisiae telomeres: Reversible repression of Pol II transcription. *Cell* **63**, 751–762 (1990).

[25] Hernday, A., Krabbe, M., Braaten, B. & Low, D. Self-perpetuating epigenetic pili switches in bacteria. *Proc Natl Acad Sci U S A* **99 Suppl 4**, 16470–6 (2002).

[26] Rando, O. J. & Verstrepen, K. J. Timescales of Genetic and Epigenetic Inheritance. *Cell* **128**, 655–668 (2007).

[27] Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M. & Elowitz, M. B. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* **440**, 545–550 (2006).

[28] Acar, M., Becskei, A. & van Oudenaarden, A. Enhancement of cellular memory by reducing stochastic transitions. *Nature* **435**, 228–231 (2005).

[29] Elf, J., Paulsson, J., Berg, O. G. & Ehrenberg, M. Near-critical phenomena in intracellular metabolite pools. *Biophys J* **84**, 154–170 (2003).

[30] Taniguchi, Y. *et al.* Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).

[31] Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005).

[32] Austin, D. W. *et al.* Gene network shaping of inherent noise spectra. *Nature* **439**, 608–611 (2006).

[33] Dunlop, M. J., Cox, R. S., Levine, J. H., Murray, R. M. & Elowitz, M. B. Regulatory activity revealed by dynamic correlations in gene expression noise. *Nat Genet* **40**, 1493–1498 (2008).

[34] Tanouchi, Y. *et al.* A noisy linear map underlies oscillations in cell size and gene expression in bacteria. *Nature* **523**, 357–60 (2015).

[35] Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444**, 643–646 (2006).

[36] Lord, N. *Fluctuation Timescales in Bacterial Gene Expression*. Ph.D. thesis, Harvard University (2013).

[37] Wang, P. *et al.* Robust growth of Escherichia coli. *Curr Biol* **20**, 1099–1103 (2010).

[38] Norman, T. M., Lord, N. D., Paulsson, J. & Losick, R. Memory and modularity in cell-fate decision making. *Nature* **503**, 481–6 (2013).

[39] Zaslaver, A., Bren, A., Ronen, M. & Itzkovitz, S. A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nat Methods* **3**, 623–628 (2006).

[40] Flynn, J. M., Neher, S. B., Kim, Y. I., Sauer, R. T. & Baker, T. a. Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol Cell* **11**, 671–83 (2003).

[41] Box, G., Jenkins, G. M. & Reinsel, G. C. *Time Series Analysis: Forecasting and Control* (Prentice Hall, Englewood Cliffs, NJ, 1994), 3rd edn.

[42] Kitagawa, G. *Introduction to time series modeling* (Chapman and Hall/CRC, 2010).

[43] Larrabee, K. L., Phillips, J. O., Williams, G. J. & Larrabee, A. R. The relative rates of protein synthesis and degradation in a growing culture of Escherichia coli. *J Biol Chem* **255**, 4125–4130 (1980).

[44] Jung, Y. *Emergence and transmission of noise created at cell division*. Ph.D. thesis, Harvard University (2017).

[45] Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).

[46] Gardner, T., Cantor, C. & Collins, J. Construction of a genetic toggle switch in Escherichia coli. *Nature* 339–342 (2000).

[47] Huh, D. & Paulsson, J. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat Genet* **43**, 95–100 (2011).

[48] Huh, D. & Paulsson, J. Random partitioning of molecules at cell division. *Proc Natl Acad Sci U S A* **108**, 15004–15009 (2011).

[49] Potvin-trottier, L., Lord, N. D., Vinnicombe, G. & Paulsson, J. Synchronous long-term oscillations in a synthetic gene circuit. *Nature* **538**, 1–4 (2016).

[50] Nakajima, M. *et al.* Reconstitution of Circadian Oscillation of Cyanobacterial KaiC Phosphorylation in Vitro. *Science* **308**, 414–415 (2005).

[51] Teng, S.-W., Mukherji, S., Moffitt, J. R., de Buyl, S. & O'Shea, E. K. Robust circadian oscillations in growing cyanobacteria require transcriptional feedback. *Science* **340**, 737–40 (2013).

[52] Mihalcescu, I., Hsing, W. & Leibler, S. Resilient circadian oscillator revealed in individual cyanobacteria. *Nature* **430**, 81–85 (2004).

[53] Chabot, J. R., Pedraza, J. M., Luitel, P. & van Oudenaarden, A. Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock. *Nature* **450**, 1249–1252 (2007).

[54] Friedland, A. E. *et al.* Synthetic gene networks that count. *Science* **324**, 1199–202 (2009).

[55] Daniel, R., Rubens, J. R., Sarpeshkar, R. & Lu, T. K. Synthetic analog computation in living cells. *Nature* (2013).

[56] Fung, E. *et al.* A synthetic gene-metabolic oscillator. *Nature* **435**, 118–122 (2005).

[57] Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–519 (2008).

[58] Tigges, M., Marquez-Lago, T. T., Stelling, J. & Fussenegger, M. A tunable synthetic mammalian oscillator. *Nature* **457**, 309–312 (2009).

[59] Danino, T., Mondragón-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463**, 326–30 (2010).

[60] Mondragón-Palomino, O., Danino, T., Selimkhanov, J., Tsimring, L. & Hasty, J. Entrainment of a population of synthetic genetic oscillators. *Science* **333**, 1315–9 (2011).

[61] Prindle, A. *et al.* Rapid and tunable post-translational coupling of genetic circuits. *Nature* (2014).

[62] Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P. & Endy, D. Amplifying Genetic Logic Gates. *Science* (2013).

[63] Tabor, J. J. *et al.* A synthetic genetic edge detection program. *Cell* **137**, 1272–81 (2009).

[64] Paulsson, J., Berg, O. G. & Ehrenberg, M. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. *Proc Natl Acad Sci USA* **97**, 7148–7153 (2000).

[65] Vilar, J. M. G., Kueh, H. Y., Barkai, N. & Leibler, S. Mechanisms of noise-resistance in genetic oscillators. *Proc Natl Acad Sci USA* **99**, 5988–92 (2002).

[66] McKane, a. J. & Newman, T. J. Predator-Prey Cycles from Resonant Amplification of Demographic Stochasticity. *Phys Rev Lett* **94**, 218102 (2005).

[67] Keiler, K., Waller, P. & Sauer, R. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271**, 990–993 (1996).

[68] Andersen, J. B. *et al.* New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol* **64**, 2240–2246 (1998).

[69] Paulsson, J. & Ehrenberg, M. Noise in a minimal regulatory network: plasmid copy number control. *Q Rev Biophys* **34**, 1–59 (2001).

[70] Moriya, T., Yamamura, M. & Kiga, D. Effects of downstream genes on synthetic genetic circuits. *BMC Syst Biol* **8**, S4 (2014).

[71] Berg, O. G., Paulsson, J. & Ehrenberg, M. Fluctuations in repressor control: thermodynamic constraints on stochastic focusing. *Biophys J* **79**, 2944–2953 (2000).

[72] Niederholtmeyer, H. *et al.* Rapid cell-free forward engineering of novel genetic ring oscillators. *eLife* **4**, 1–18 (2015).

[73] Van der Pol, B. On relaxation-oscillations. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **2**, 978–992 (1926).

[74] Brewster, R. *et al.* The Transcription Factor Titration Effect Dictates Level of Gene Expression. *Cell* **156**, 1312–1323 (2014).

[75] Verdú, S. Poisson communication theory. *Invited talk, March* 1–4 (1999).

[76] Landgraf, D., Okumus, B., Chien, P., Baker, T. A. & Paulsson, J. Segregation of molecules at cell division reveals native protein localization. *Nat Methods* **9**, 480–482 (2012).

[77] Gora, K. G. *et al.* Regulated proteolysis of a transcription factor complex is critical to cell cycle progression in Caulobacter crescentus. *Mol Microbiol* **87**, 1277–1289 (2013).

[78] Goldbeter, a. & Koshland, D. E. An amplified sensitivity arising from covalent modification in biological systems. *Proc Natl Acad Sci U S A* **78**, 6840–6844 (1981).

[79] Berg, O. G., Paulsson, J. & Ehrenberg, M. Fluctuations and quality of control in biological cells: zero-order ultrasensitivity reinvestigated. *Biophys J* **79**, 1228–1236 (2000).

[80] Cherry, J. L. & Adler, F. R. How to make a biological switch. *J Theor Biol* **203**, 117–133 (2000).

[81] Wong, W. W., Tsai, T. Y. & Liao, J. C. Single-cell zeroth-order protein degradation enhances the robustness of synthetic oscillator. *Mol Syst Biol* **3**, 130 (2007).

[82] Buchler, N. E. & Louis, M. Molecular titration and ultrasensitivity in regulatory networks. *J Mol Biol* **384**, 1106–19 (2008).

[83] Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **93555**, 2340–2361 (1977).

[84] Herman, C., Thévenet, D., D'Ari, R. & Bouloc, P. Degradation of sigma 32, the heat shock regulator in Escherichia coli, is governed by HflB. *Proc Natl Acad Sci U S A* **92**, 3516–20 (1995).

[85] Flynn, J. M., Levchenko, I., Sauer, R. T. & Baker, T. a. Modulating substrate choice: the SspB adaptor delivers a regulator of the extracytoplasmic-stress response to the AAA+ protease ClpXP for degradation. *Genes Dev* **18**, 2292–301 (2004).

[86] Muffler, A., Fischer, D., Altuvia, S., Storz, G. & Hengge-Aronis, R. The response regulator RssB controls stability of the sigma(S) subunit of RNA polymerase in Escherichia coli. *The EMBO journal* **15**, 1333–9 (1996).

[87] Yu, A. Y. H. & Houry, W. A. ClpP: A distinctive family of cylindrical energy-dependent serine proteases. *FEBS Lett* **581**, 3749–3757 (2007).

[88] Brötz-Oesterhelt, H. *et al.* Dysregulation of bacterial proteolytic machinery by a new class of antibiotics. *Nat Med* **11**, 1082–1087 (2005).

[89] Kirstein, J. *et al.* The antibiotic ADEP reprogrammes CIP, switching it from a regulated to an uncontrolled protease. *EMBO Molecular Medicine* **1**, 37–49 (2009).

[90] Li, D. H. S. *et al.* Acyldepsipeptide antibiotics induce the formation of a structured axial channel in ClpP: A model for the ClpX/ClpA-bound state of ClpP. *Chem Biol* **17**, 959–969 (2010).

[91] Lee, B.-G. *et al.* Structures of ClpP in complex with acyldepsipeptide antibiotics reveal its activation mechanism. *Nature structural & molecular biology* **17**, 471–478 (2010).

[92] Leung, E. *et al.* Activators of Cylindrical Proteases as Antimicrobials: Identification and Development of Small Molecule Activators of ClpP Protease. *Chem Biol* **18**, 1167–1178 (2011).

[93] Dougan, D. A. (ed.) *Regulated Proteolysis in Microorganisms*, vol. 66 of *Subcellular Biochemistry* (Springer Netherlands, Dordrecht, 2013).

[94] Baker, T. a. & Sauer, R. T. ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim Biophys Acta* **1823**, 15–28 (2012).

[95] Maglica, Ž., Kolygo, K. & Weber-Ban, E. Optimal Efficiency of ClpAP and ClpXP Chaperone-Proteases Is Achieved by Architectural Symmetry. *Structure* **17**, 508–516 (2009).

[96] Ninnis, R. L., Spall, S. K., Talbo, G. H., Truscott, K. N. & Dougan, D. A. Modification of PATase by L/F-transferase generates a ClpS-dependent N-end rule substrate in Escherichia coli. *The EMBO journal* **28**, 1732–44 (2009).

[97] Guisbert, E., Herman, C., Lu, C. Z. & Gross, C. A. A chaperone network controls the heat shock response in E. coli. *Genes Dev* **18**, 2812–2821 (2004).

[98] Rodriguez, F. *et al.* Molecular Basis for Regulation of the Heat Shock Transcription Factor sigma32 by the DnaK and DnaJ Chaperones. *Mol Cell* **32**, 347–358 (2008).

[99] Keiler, K. C. Mechanisms of ribosome rescue in bacteria. *Nat Rev Microbiol* **13**, 285–97 (2015).

[100] Moore, S. D. & Sauer, R. T. Ribosome rescue: tmRNA tagging activity and capacity in Escherichia coli. *Mol Microbiol* **58**, 456–466 (2005).

[101] Moore, S. D. & Sauer, R. T. The tmRNA System for Translational Surveillance and Ribosome Rescue. *Annu Rev Biochem* **76**, 101–124 (2007).

[102] Lies, M. & Maurizi, M. R. Turnover of endogenous SsrA-tagged proteins mediated by ATP-dependent proteases in Escherichia coli. *J Biol Chem* **283**, 22918–29 (2008).

[103] Komine, Y., Kitabatake, M., Yokogawa, T., Nishikawa, K. & Inokuchi, H. A tRNA-like structure is present in 10Sa RNA, a small stable RNA from Escherichia coli. *Proc Natl Acad Sci U S A* **91**, 9223–7 (1994).

[104] Gottesman, S. & Roche, E. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* 1338–1347 (1998).

[105] Weber-Ban, E. U., Reid, B. G., Miranker, A. D. & Horwich, A. L. Global unfolding of a substrate protein by the Hsp100 chaperone ClpA. *Nature* **401**, 90–93 (1999).

[106] Kenniston, J. A., Baker, T. A., Fernandez, J. M. & Sauer, R. T. Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell* **114**, 511–520 (2003).

[107] Kenniston, J. A., Burton, R. E., Siddiqui, S. M., Baker, T. A. & Sauer, R. T. Effects of local protein stability and the geometric position of the substrate degradation tag on the efficiency of ClpXP denaturation and degradation. *J Struct Biol* **146**, 130–140 (2004).

[108] Kenniston, J. a., Baker, T. a. & Sauer, R. T. Partitioning between unfolding and release of native domains during ClpXP degradation determines substrate selectivity and partial processing. *Proc Natl Acad Sci U S A* **102**, 1390–1395 (2005).

[109] Neher, S. B. *et al.* Proteomic profiling of ClpXP substrates after DNA damage reveals extensive instability within SOS regulon. *Mol Cell* **22**, 193–204 (2006).

[110] Levchenko, I., Meredith, S., Sauer, R. T. & Baker, T. A. A Specificity-Enhancing Factor for the ClpXP Degradation Machine. *Science* **289**, 2354–2356 (2000).

[111] Pratt, L. A. & Silhavy, T. J. The response regulator SprE controls the stability of RpoS. *Proc Natl Acad Sci U S A* **93**, 2488–2492 (1996).

[112] Zhou, Y., Gottesman, S., Hoskins, J. R., Maurizi, M. R. & Wickner, S. The RssB response regulator directly targets ??S for degradation by ClpXP. *Genes Dev* **15**, 627–637 (2001).

[113] Ortega, J., Lee, H. S., Maurizi, M. R. & Steven, A. C. ClpA and ClpX ATPases bind simultaneously to opposite ends of ClpP peptidase to form active hybrid complexes. *J Struct Biol* **146**, 217–226 (2004).

[114] Gottesman, S. & Clark, W P, M. M. R. The ATP-dependent Clp Protease of Escherichia coli. *Jbc* **265**, 7886–7893 (1990).

[115] Maglica, Ž., Striebel, F. & Weber-Ban, E. An Intrinsic Degradation Tag on the ClpA C-Terminus Regulates the Balance of ClpAP Complexes with Different Substrate Specificity. *J Mol Biol* **384**, 503–511 (2008).

[116] Dougan, D. A., Truscott, K. N. & Zeth, K. The bacterial N-end rule pathway: Expect the unexpected. *Mol Microbiol* **76**, 545–558 (2010).

[117] Wang, K. H., Oakes, E. S. C., Sauer, R. T. & Baker, T. A. Tuning the strength of a bacterial N-end rule degradation signal. *The Journal of biological chemistry* **283**, 24600–7 (2008).

[118] Schmidt, R., Zahn, R., Bukau, B. & Mogk, A. ClpS is the recognition component for Escherichia coli substrates of the N-end rule degradation pathway. *Mol Microbiol* **72**, 506–17 (2009).

[119] Tomoyasu, T., Mogk, A., Langen, H., Goloubinoff, P. & Bukau, B. Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the Escherichia coli cytosol. *Mol Microbiol* **40**, 397–413 (2001).

[120] Tsilibaris, V., Maenhaut-Michel, G. & Van Melderen, L. Biological roles of the Lon ATP-dependent protease. *Res Microbiol* **157**, 701–13 (2006).

[121] Gur, E. & Sauer, R. T. Recognition of misfolded proteins by Lon, a AAA+ protease. *Genes Dev.* **22**, 2267–2277 (2008).

[122] Gur, E. & Sauer, R. T. Degrons in protein substrates program the speed and operating efficiency of the AAA+ Lon proteolytic machine. *Proc Natl Acad Sci U S A* **106**, 18503–8 (2009).

[123] Tomoyasu, T. *et al.* Escherichia coli FtsH is a membrane-bound, ATP-dependent protease which degrades the heat-shock transcription factor sigma 32. *The EMBO journal* **14**, 2551–2560 (1995).

[124] Kanemori, M., Nishihara, K., Yanagi, H. & Yura, T. Synergetic roles of HslVU and other ATP-dependent protease in controlling in vivo turnover of s32 and abnormal proteins in Escerichia coli. *J Bacteriol* **179**, 7219–7225 (1997).

[125] Herman, C., Prakash, S., Lu, C. Z., Matouschek, A. & Gross, C. A. Lack of a robust unfoldase activity confers a unique level of substrate specificity to the universal AAA protease FtsH. *Mol Cell* **11**, 659–669 (2003).

[126] Straus, D. B., Walter, W. a. & Gross, C. a. The heat shock response of E. coli is regulated by changes in the concentration of $\sigma$32. *Nature* **329**, 348–351 (1987).

[127] Ogura, T. *et al.* Balanced biosynthesis of major membrane components through regulated degradation of the committed enzyme of lipid A biosynthesis by the AAA protease FtsH (HflB) in Escherichia coli. *Mol Microbiol* **31**, 833–844 (1999).

[128] Westphal, K., Langklotz, S., Thomanek, N. & Narberhaus, F. A trapping approach reveals novel substrates and physiological functions of the essential protease Ftsh in Escherichia coli. *J Biol Chem* **287**, 42962–42971 (2012).

[129] Gur, E., Vishkautzan, M. & Sauer, R. T. Protein unfolding and degradation by the AAA+ Lon protease. *Protein science : a publication of the Protein Society* **21**, 268–78 (2012).

[130] Lee, C., Schwartz, M. P., Prakash, S., Iwakura, M. & Matouschek, A. ATP-Dependent Proteases Degrade Their Substrates by Processively Unraveling Them from the Degradation Signal. *Mol Cell* **7**, 627–637 (2001).

[131] Vass, R. H. & Chien, P. Critical clamp loader processing by an essential AAA+ protease in Caulobacter crescentus. *Proc Natl Acad Sci U S A* **110**, 18138–43 (2013).

[132] Too, P. H. M., Erales, J., Simen, J. D., Marjanovic, A. & Coffino, P. Slippery substrates impair function of a bacterial protease ATPase by unbalancing translocation versus exit. *J Biol Chem* (2013).

[133] Coffino, P., Too, P. H.-m. & Erales, J. Letters : Slippery Substrates Impair ATP-dependent Protease Function by Slowing Unfolding. *J Biol Chem* **289**, 6–7 (2014).

[134] Nassif, N. D., Cambray, S. E. & Kraut, D. A. Slipping up: Partial substrate degradation by ATP-dependent proteases. *IUBMB Life* **66**, 309–317 (2014).

[135] Kraut, D. A. Slippery substrates impair ATP-dependent protease function by slowing unfolding. *J Biol Chem* **288**, 34729–34735 (2013).

[136] Okuno, T., Yamanaka, K. & Ogura, T. An AAA protease FtsH can initiate proteolysis from internal sites of a model substrate, apo-flavodoxin. *Genes Cells* **11**, 261–268 (2006).

[137] Koodathingal, P. *et al.* ATP-dependent proteases differ substantially in their ability to unfold globular proteins. *The Journal of biological chemistry* **284**, 18674–84 (2009).

[138] Farrell, C. M., Grossman, A. D. & Sauer, R. T. Cytoplasmic degradation of ssrA-tagged proteins. *Mol Microbiol* **57**, 1750–1761 (2005).

[139] Christensen, S. K., Mikkelsen, M., Pedersen, K. & Gerdes, K. RelE, a global inhibitor of translation, is activated during nutritional stress. *Proc Natl Acad Sci U S A* **98**, 14328–14333 (2001).

[140] Christensen, S. K., Pedersen, K., Hansen, F. G. & Gerdes, K. Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA. *J Mol Biol* **332**, 809–819 (2003).

[141] Langklotz, S. & Narberhaus, F. The Escherichia coli replication inhibitor CspD is subject to growth-regulated degradation by the Lon protease. *Mol Microbiol* **80**, 1313–1325 (2011).

[142] Hansen, S. *et al.* Regulation of the Escherichia coli HipBA toxin-antitoxin system by proteolysis. *PLoS One* **7** (2012).

[143] Ventura, M., Zhang, Z. & Cronin, M. The ClgR Protein Regulates Transcription of the clpP Operon in Bifidobacterium breve UCC 2003. *J Bacteriol* (2005).

[144] Bellier, A., Gominet, M. & Mazodier, P. Post-translational control of the Streptomyces lividans ClgR regulon by ClpP. *Microbiology* **152**, 1021–7 (2006).

[145] Personne, Y., Brown, A. C., Schuessler, D. L. & Parish, T. Mycobacterium tuberculosis ClpP proteases are co-transcribed but exhibit different substrate specificities. *PLoS One* **8**, e60228 (2013).

[146] Flynn, J. M. *et al.* Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. *Proc Natl Acad Sci U S A* **98**, 10584–9 (2001).

[147] Choy, J. S., Aung, L. L. & Karzai, A. W. Lon protease degrades transfer-messenger RNA-tagged proteins. *J Bacteriol* **189**, 6564–6571 (2007).

[148] Herman, C., Thévenet, D., Bouloc, P., Walker, G. C. & D'Ari, R. Degradation of carboxy-terminal-tagged cytoplasmic proteins by the Escherichia coli protease HflB (FtsH). *Genes Dev* **12**, 1348–1355 (1998).

[149] Levitskaya, J. *et al.* Inhibition of antigen processing by the internal repeat region of the Epstein-Barr virus nuclear antigen-1. (1995).

[150] Zhang, M. & Coffino, P. Repeat Sequence of Epstein-Barr Virus-encoded Nuclear Antigen 1 Protein Interrupts Proteasome Substrate Processing. *J Biol Chem* **279**, 8635–8641 (2004).

[151] Hoyt, M. a. *et al.* Glycine-alanine repeats impair proper substrate unfolding by the proteasome. *The EMBO journal* **25**, 1720–1729 (2006).

[152] Hersch, G. L., Baker, T. a. & Sauer, R. T. SspB delivery of substrates for ClpXP proteolysis probed by the design of improved degradation tags. *Proc Natl Acad Sci U S A* **101**, 12136–41 (2004).

[153] Bakshi, S., Siryaporn, A., Goulian, M. & Weisshaar, J. C. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Mol Microbiol* **85**, 21–38 (2012).

[154] Kim, Y. I., Burton, R. E., Burton, B. M., Sauer, R. T. & Baker, T. a. Dynamics of substrate denaturation and translocation by the ClpXP degradation machine. *Mol Cell* **5**, 639–648 (2000).

[155] Dougan, D. A., Reid, B. G., Horwich, A. L. & Bukau, B. ClpS, a substrate modulator of the ClpAP machine. *Mol Cell* **9**, 673–683 (2002).

[156] Nager, A. R., Baker, T. A. & Sauer, R. T. Stepwise unfolding of a beta barrel protein by the AAA+ ClpXP protease. *J Mol Biol* (2011).

[157] McGinness, K. E., Baker, T. a. & Sauer, R. T. Engineering controllable protein degradation. *Mol Cell* **22**, 701–7 (2006).

[158] Laughrea, M. Mistranslation in twelve Escherichia coli ribosomal proteins. *Eur. J. Biochem.* **64**, 59–64 (1987).

[159] Zaher, H. S. & Green, R. Fidelity at the Molecular Level: Lessons from Protein Synthesis. *Cell* **136**, 746–762 (2009).

[160] Landgraf, D. *Quantifying localizations and dynamics in single bacterial cells*. Ph.D. thesis, Harvard University (2012).

[161] Mizusawa, S. & Gottesman, S. Protein degradation in Escherichia coli: the lon gene controls the stability of sulA protein. *Proc Natl Acad Sci U S A* **80**, 358–362 (1983).

[162] Wohlever, M. L., Nager, A. R., Baker, T. A. & Sauer, R. T. Engineering fluorescent protein substrates for the AAA+ Lon protease. *Protein engineering, design & selection : PEDS* **26**, 299–305 (2013).

[163] Muffler, A., Traulsen, D. D., Lange, R. & Hengge-Aronis, R. Posttranscriptional osmotic regulation of the sS subunit of RNA polymerase in Escherichia coli. *J Bacteriol* **178**, 1607–1613 (1996).

[164] Hoskins, J. R., Yanagihara, K., Mizuuchi, K. & Wickner, S. ClpAP and ClpXP degrade proteins with tags located in the interior of the primary sequence. *Proc Natl Acad Sci U S A* **99**, 11037–42 (2002).

[165] Lim, H. N., Lee, Y. & Hussein, R. Fundamental relationship between operon organization and gene expression. *Proc Natl Acad Sci U S A* **108**, 10626–10631 (2011).

[166] Yaginuma, H. *et al.* Diversity in ATP concentrations in a single bacterial cell population revealed by quantitative single-cell imaging. *Sci Rep* **4**, 6522 (2014).

[167] Nonaka, G., Blankschien, M., Herman, C., Gross, C. a. & Rhodius, V. a. Regulon and promoter analysis of the E. coli heat-shock factor, sigma32 , reveals a multifaceted cellular response to heat stress. *Genes Dev.* **20**, 1776–1789 (2006).

[168] Jonas, K., Liu, J., Chien, P. & Laub, M. T. Proteotoxic stress induces a cell-cycle arrest by stimulating lon to degrade the replication initiator DnaA. *Cell* **154**, 623–636 (2013).

[169] Wang, Q. P. & Kaguni, J. M. DnaA protein regulates transcriptions of the rpoH gene of Escherichia coli. *J Biol Chem* **264**, 7338–7344 (1989).

[170] Grigorova, I. L., Phleger, N. J., Mutalik, V. K. & Gross, C. a. Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA. *Proc Natl Acad Sci U S A* **103**, 5332–5337 (2006).

[171] Li, G.-W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. *Cell* **157**, 624–635 (2014).

[172] Trentini, D. B., Suskiewicz, M. J., Deszcz, L. & Mechtler, K. Arginine phosphorylation marks proteins for degradation by the ClpCP protease. *Nature* **539**, 1–41 (2016).

[173] Battesti, A., Majdalani, N. & Gottesman, S. The RpoS-mediated general stress response in Escherichia coli. *Annu Rev Microbiol* **65**, 189–213 (2011).

[174] Uphoff, S. *et al.* Stochastic activation of a DNA damage response causes cell-to-cell mutation rate variation. *Science* **351**, 1094–1098 (2016).

[175] Soini, J. *et al.* Transient increase of ATP as a response to temperature up-shift in Escherichia coli. *Microb Cell Fact* **4**, 9 (2005).

[176] Kamiyama, D. *et al.* Versatile protein tagging in cells with split fluorescent protein. *Nat Commun* **7**, 11046 (2016).

[177] Los, G. V. *et al.* HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* **3**, 373–382 (2008).

[178] Ohana, R. F. *et al.* HaloTag7: A genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification. *Protein Expression Purif* **68**, 110–120 (2009).

[179] Grimm, J. B. *et al.* A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat Methods* **12**, 244–250 (2015).

[180] Bisson Filho, A. W. *et al.* Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division. *Science* (2017).

[181] Cho, H. *et al.* Bacterial cell wall biogenesis is mediated by SEDS and PBP polymerase families functioning semi-autonomously. *Nat. Microbiol* **1**, 16172 (2016).

[182] Ireton, K., Rudner, D. Z., Siranosian, K. J. & Grossman, A. D. Integration of multiple developmental signals in Bacillus subtilis through the Spo0A transcription factor. *Genes Dev* **7**, 283–294 (1993).

[183] Fujita, M., González-Pastor, J. E. & Losick, R. High- and low-threshold genes in the Spo0A regulon of Bacillus subtilis. *J Bacteriol* **187**, 1357–1368 (2005).

[184] Gibson, D., Young, L. & Chuang, R. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 12–16 (2009).

[185] McKenzie, G. J. & Craig, N. L. Fast, easy and efficient: site-specific insertion of transgenes into enterobacterial chromosomes using Tn7 without need for selection of the insertion event. *BMC Microbiol* **6**, 39 (2006).

[186] Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer control of microscopes using μManager. *Current protocols in molecular biology* **Chapter 14**, Unit14.20 (2010).

[187] Tikhonov, A. N. & Arsenin, V. Y. Solutions of Ill-Posed Problems. *Math Comput* **32**, 1320–2 (1978).

[188] Blackman, R. B. & Tukey, J. W. *The measurement of power spectra.* (Dover Publications, New York, NY, 1958).

[189] Oppenheim, A. V. & Schafer, R. W. *Discrete-time signal processing* (Prentice Hall, Upper Saddle River, NJ, 1989), 3 edn.

[190] Jenkins, G. M. & Watts, D. G. *Spectral Analysis* (Holden-Day, 1968).

[191] Chait, R., Shrestha, S., Shah, A. K., Michel, J.-B. & Kishony, R. A differential drug screen for compounds that select against antibiotic resistance. *PLoS One* **5**, e15179 (2010).

[192] Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **25**, 1203–10 (1997).

[193] Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat Biotechnol* **20**, 87–90 (2002).

[194] Rosenow, M. a., Huffman, H. a., Phail, M. E. & Wachter, R. M. The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation. *Biochemistry* **43**, 4464–72 (2004).

[195] Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006.0008 (2006).