



**HAL**  
open science

# Microbial growth control in changing environments : Theoretical and experimental study of resource allocation in *Escherichia coli*

Nils Giordano

► **To cite this version:**

Nils Giordano. Microbial growth control in changing environments : Theoretical and experimental study of resource allocation in *Escherichia coli*. Quantitative Methods [q-bio.QM]. Université Grenoble Alpes, 2017. English. NNT : 2017GREAV021 . tel-01685626

**HAL Id: tel-01685626**

**<https://theses.hal.science/tel-01685626v1>**

Submitted on 16 Jan 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## THÈSE

Pour obtenir le grade de

### DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Biologie cellulaire**

Arrêté ministériel du 25 mai 2016

Préparée au sein du **Laboratoire Interdisciplinaire de Physique**  
et de l'école doctorale **Chimie et Sciences du Vivant**

## Microbial growth control in changing environments

Theoretical and experimental study of resource allocation in  
*Escherichia coli*

Présentée par

**Nils GIORDANO**

Thèse dirigée par **Johannes GEISELMANN**  
et codirigée par **Hidde DE JONG**

Thèse soutenue publiquement le **23 mars 2017**,  
devant le jury composé de :

<b>Mme, Irina MIHALCESCU</b>	Prof.	Univ. Grenoble Alpes	Présidente
<b>M, Matthew SCOTT</b>	Assoc. Prof.	Univ. of Waterloo (CA)	Rapporteur
<b>M, Guy-Bart STAN</b>	Reader	Imp. College London (UK)	Rapporteur
<b>M, Olivier BERNARD</b>	DR	Inria Sophia-Antipolis	Examineur
<b>M, Jean-Luc GOUZÉ</b>	DR	Inria Sophia-Antipolis	Examineur
<b>M, Matthieu JULES</b>	CR	Inra Grignon	Examineur



## Remerciements

Il est probable que cette page, même si elle s'adresse au public restreint des francophones, sera la plus consultée de cette thèse. Cela fait effectivement longtemps que le manuscrit de thèse a perdu le monopole de la communication scientifique. On exige cependant qu'il soit traité avec le même respect qu'il y a un siècle, en faisant comme s'il n'avait rien perdu de sa gloire passée, non sans une certaine forme d'hypocrisie administrative. Ce paradoxe rend généralement cet exercice difficile, mais par sa longueur, le déroulement du doctorat lui-même l'est bien davantage. À ce jour, j'ai par exemple investi plus de 15% de mon existence dans ce projet de recherche. C'est long, surtout pour quelqu'un comme moi qui aime toucher à tout. Inévitablement, de nombreuses rencontres qui ont grandement influencé ce travail se sont succédées, et c'est le rôle de cette page que de leur manifester ma gratitude.

Tout d'abord, je remercie évidemment Marion Joulié, qui même si son engagement date de bien avant ce doctorat, a su rester jusqu'au bout et restera encore je l'espère pour les nombreuses années à venir.

Je remercie particulièrement Valentin Zulkower et Edith Grac qui étaient présents pendant la meilleure période de ce doctorat. Sans nos nombreuses discussions, ce travail n'aurait probablement pas été le même. Viennent ensuite tous les jeunes membres des équipes Ibis et BIOP, notamment les passages remarquables de Çiğdem Ak, Cindy Barillot, Diana Stefan, Bernard Chelli, Ludovic Lancelot, Julien Sauvage, et tous les autres dont les noms ne me viennent pas au moment d'écrire ces lignes. Je n'oublie pas de remercier les (nombreux) jeunes chercheurs de l'équipe Mistis qui ont su m'accueillir lorsque la plupart des gens sus-cités se sont progressivement évaporés.

D'un point de vue plus scientifique, je remercie évidemment mes directeurs Hans Geiselman et Hidde de Jong, mais aussi Corinne Pinel, Irina Mihalcescu, Delphine Ropers, Eugenio Cinquemani, Jean-Luc Gouzé, et Francis Mairet, qui ont apporté des contributions plus que concrètes dans les étapes clés de ce travail. Je remercie pour les mêmes raisons tous les membres des équipes Ibis et BIOP avec qui j'ai pu échanger lors de nos nombreuses réunions hebdomadaires.

Enfin, je tenais également à remercier les communautés de Bioinfo-fr.net et JeBiF – RSG France, en particulier jnsll pour ses corrections des parties françaises de ce manuscrit, mais aussi Chopopope, Kumquatum, Norore, Hautbit, m4rsu et Will qui ont su virtuellement me supporter dans les meilleures et les pires périodes de ce doctorat.

## Abstract

Growth is the most fundamental property of life. Growth consists in the transformation of matter and energy from the environment into diverse organic structures. Interestingly, general growth laws relate the macromolecular composition of the cell to growth rate. These laws are widespread and conserved in different microbial species, suggesting a fundamental principle of design. Recent work has shown that these empirical regularities can be derived from coarse-grained models of resource allocation and explained by the principles of natural selection. However, the vast majority of these studies focus on steady-state growth. Such conditions are rarely found in natural habitats, where microorganisms are continually challenged by environmental fluctuations. The aim of this thesis is to extend the theoretical and experimental studies of microbial growth strategies to changing environments.

Using a self-replicator model, we developed a theoretical framework that encapsulates the main features of growth. We formulate dynamical growth maximization as an optimal control problem that the microbial cell must solve in order to allocate the available resources to the gene expression machinery or to metabolism. Using Pontryagin's Maximum Principle, we have derived a general solution to the optimization problem and we have compared the optimal strategy with possible implementations of growth control in bacterial cells. Our results show that simple control strategies that maximize the growth-rate at steady state are suboptimal for transitions from one growth regime to another. We show that a near-optimal control strategy in dynamical conditions requires information about several, rather than a single, physiological variable. Interestingly, this strategy has structural analogies with the regulation of ribosomal protein synthesis by the signaling molecule ppGpp in the enterobacterium *Escherichia coli*. The strategy involves sensing a discrepancy between the concentrations of precursor metabolites and ribosomes, and the control of the rate of ribosome synthesis in a switch-like manner.

Even though this switch-like ribosome synthesis has been suggested by published data, the phenomenon has never been experimentally confirmed. We therefore measured ribosomal abundance in *Escherichia coli* at the single-cell level during a nutrient upshift. More precisely, we constructed a strain in which a fluorescent marker has been attached to a ribosomal subunit, thus allowing *in-vivo* monitoring of the abundance of ribosomes. We monitored this strain in a microfluidics device designed for long-term imaging of individual cells in a continuous culture, and used this experimental setup to simulate a nutrient upshift by changing the input medium. We developed a Kalman smoothing method for extracting quantitative information about re-

source allocation to ribosome synthesis from the raw data. Even though our preliminary results do not allow to reach a final conclusion, they do suggest the presence of oscillatory patterns after an upshift that are reminiscent of the expected behavior.

Our results demonstrate that the capability of regulatory systems to integrate information about several physiological variables is critical for optimizing growth in a changing environment. The proposed control scheme correctly reproduces the observed growth laws at steady state, but also predicts novel and unexpected behaviors when applied to a dynamical environment. Our improved understanding of the principles that govern the control of bacterial growth could be used for improving biotechnological processes, in particular those that use microorganisms to produce high valuable-added products for the chemical or biomedical industry.

### **Non-technical summary of the work**

Bacteria grow by exploiting matter and energy extracted from their environment. The strategies that enable bacteria to optimize their growth rate have been extensively studied. However, most of these studies were carried out in a constant environment. Here, we construct a simple model of microbial growth and use mathematical methods from optimal control theory to determine how the cells should behave when the environment changes abruptly. We find that microbial cells should adopt an *on-off* strategy: they should alternatively allocate all of their resources to the production of two categories of cellular components, the elements needed for gene expression and the ones needed for the general metabolism of the cell. Our preliminary experiments confirm this prediction. In other words, theory tells us that switch-like mechanisms would optimize the growth of bacterial cells and our experiments validate the theory. This suggests new ways for optimizing biotechnological processes.

## Résumé

Croissance et reproduction sont des mécanismes fondamentaux du vivant. Chez les micro-organismes, ces processus sont couplés dans la transformation des ressources de l'environnement (matière et énergie) en nouvelles structures organiques. Étonnamment, malgré l'extrême diversité des micro-organismes, certaines caractéristiques de leur physiologie suivent des lois de croissance universelles. Cela suggère l'existence de principes fondamentaux, ce qui a été en effet récemment confirmé en montrant que ces lois s'expliquent facilement si l'organisme maximise son taux de croissance dans chaque environnement. Cependant, ces lois ont seulement été étudiées lors de croissances à l'état stationnaire, c'est-à-dire lorsque l'environnement et donc le taux de croissance sont stables. Ces conditions, même si elles peuvent être reproduites en laboratoire, n'ont rien à voir avec les conditions de vie dans lesquelles les organismes ont évolué durant des milliards d'années. Le but de cette thèse est d'étendre dans un contexte d'environnement purement dynamique, l'étude à la fois théorique et expérimentale de ces stratégies de croissance microbienne.

En modélisant la cellule comme un auto-réplicateur, nous cherchons à savoir quelles sont les meilleures stratégies d'allocation des ressources lors de l'adaptation à un nouvel environnement. Ce problème se formule très bien comme un problème de contrôle optimal : la cellule choisit en temps réel comment allouer ses ressources entre la machinerie d'expression génique et le métabolisme. Le meilleur comportement possible, comme le révèle l'application du Principe de Maximisation de Pontryagin, est de successivement orienter toutes les ressources dans chacun des deux secteurs, une stratégie communément appelée *bang-bang*. Mais s'approcher d'un tel contrôle requiert pour la cellule des stratégies de régulation bien plus complexes que celles qui étaient suffisantes pour maximiser le taux de croissance à l'état stationnaire. De manière intéressante, la régulation de la synthèse des ribosomes par le ppGpp chez la bactérie *Escherichia coli* s'avère présenter la structure adéquate. Nous montrons en effet qu'elle permet de détecter rapidement toute incompatibilité entre la concentration de précurseurs et celle des ribosomes, et d'ajuster en conséquence la synthèse de ces derniers pour obtenir un comportement proche de l'optimum mathématique prédit.

Même si de vieilles données le suggèrent, un tel comportement *bang-bang* n'a jamais été totalement confirmé expérimentalement pour la synthèse des ribosomes. Nous mesurons donc l'abondance des ribosomes chez *Escherichia coli* au niveau de cellules individuelles lors d'un changement brutal de milieu de culture. En particulier, nous créons une souche de *E.coli* sur laquelle un rapporteur fluorescent est attaché à l'une des sous-unités ribosomales, permettant ainsi leur quantification *in vivo*. Un appareillage micro-fluidique

nous permet ensuite de contrôler en temps réel le milieu de croissance tout en observant individuellement chaque cellule fluorescente. Nous développons une méthode basée sur le lissage de Kalman qui est capable de reconstruire la façon dont les ressources sont aiguillées vers la synthèse des ribosomes. Même si ces résultats sont préliminaires, ils suggèrent que la concentration des ribosomes oscille après le changement d'environnement, ce qui rappelle une stratégie de type *bang-bang*.

Nos résultats montrent que la capacité des systèmes de régulation à intégrer l'état de différentes variables physiologiques est cruciale dans l'optimisation de la croissance en environnement variable. Au final, nous démontrons que les principes utilisés à l'état stationnaire peuvent, lorsqu'ils sont appliqués en dynamique, générer des comportements inattendus et expliquer plus en détails les stratégies de régulation employées par les micro-organismes.

### **Résumé pour un public non spécialiste**

Bactéries et autres micro-organismes se multiplient en utilisant la matière et l'énergie présentes dans leur environnement. Les stratégies qui leur permettent d'optimiser ce processus ont été longuement étudiées, mais uniquement dans des environnements stables. Ici, nous construisons un modèle simple de la croissance microbienne et utilisons la théorie du contrôle optimal pour déterminer la façon dont les cellules devraient agir lorsque leur environnement est soudainement modifié. Nous mettons en évidence que la meilleure stratégie consiste à aiguiller toutes les ressources disponibles vers la production d'un seul composant de manière alternative, une prédiction que nous tentons également de confirmer de manière expérimentale. Ces résultats pourraient suggérer de nouvelles façons d'optimiser les processus biotechnologiques de production de molécules d'intérêt.





# Contents

<b>1</b>	<b>Introduction</b>	<b>11</b>
1.1	Context . . . . .	13
1.1.1	Self-replication is a resource allocation problem . . . . .	13
1.1.2	Optimizing biomass production yields a competitive advantage . . . . .	14
1.1.3	Growth laws are universal strategies employed by microorganisms . . . . .	15
1.1.4	Static versus dynamical perspective on growth . . . . .	17
1.2	Problem statement . . . . .	20
1.3	Related work . . . . .	21
1.3.1	Modeling growth of microorganisms . . . . .	21
1.3.2	Measuring growth of microorganisms . . . . .	23
1.4	Approach . . . . .	26
<b>2</b>	<b>Dynamical Allocation of Cellular Resources as an Optimal Control Problem</b>	<b>29</b>
2.1	Introduction . . . . .	31
2.2	Results . . . . .	35
2.2.1	Self-replicator model of resource allocation . . . . .	35
2.2.2	Growth-rate maximization of the self-replicator reproduces bacterial growth laws . . . . .	38
2.2.3	Biomass maximization as an optimal control problem . . . . .	42
2.2.4	Solution of the optimal control problem . . . . .	43
2.2.5	Simple feedback control strategies: exploiting information on nutrients or precursors . . . . .	45
2.2.6	A near-optimal feedback control strategy that exploits information on the imbalance between precursors and the gene expression machinery . . . . .	50
2.3	Discussion . . . . .	54
2.4	Methods . . . . .	58
2.4.1	Steady-state analysis of model . . . . .	58

2.4.2	Model fitting . . . . .	59
2.4.3	Solution of optimal control problem . . . . .	60
2.4.4	Specification and analysis of control strategies . . . . .	61
2.5	Supporting information for Chapter 2 . . . . .	63
2.5.1	S1 Text – Model derivation and analysis . . . . .	63
2.5.2	S2 Text – Model parameters . . . . .	71
2.5.3	S3 Text – Solution of optimal control problem . . . . .	74
2.5.4	S4 Text – Kinetic model of the ppGpp system . . . . .	79
2.5.5	S1 Table – Parameter values of the self-replicator model	84
2.5.6	S1 Figure – Simple control strategies for the self-replicator of bacterial growth . . . . .	85
<b>3</b>	<b>Dynamics of resource allocation during an acetate-glucose upshift</b>	<b>87</b>
3.1	Introduction . . . . .	90
3.2	Results . . . . .	92
3.2.1	Experimental design . . . . .	92
3.2.2	Data acquisition . . . . .	94
3.2.3	Data analysis using Kalman smoothing . . . . .	97
3.2.4	Estimation of growth rate and resource allocation profile	99
3.3	Discussion . . . . .	109
3.4	Material and Methods . . . . .	113
3.4.1	Bacterial strain construction . . . . .	113
3.4.2	Growth conditions . . . . .	116
3.4.3	Microfluidic device . . . . .	117
3.4.4	Cell segmentation . . . . .	118
3.4.5	Kalman smoothing . . . . .	119
3.5	Supporting Information for Chapter 3 . . . . .	121
3.5.1	S5 Text – DNA sequences used for the strain construction	121
3.5.2	S6 Text – Strain validation in batch growing conditions	139
3.5.3	S7 Text – Noise estimation in the microscopy experiment	144
3.5.4	S8 Text – Examples of complete analysis for 5 cells . .	149
3.5.5	S2 Figure – Cell categories identified in the analysis . .	155
3.5.6	S3 Figure – Robust statistics for the cell categories . .	156
<b>4</b>	<b>Discussion</b>	<b>157</b>
	<b>Bibliography</b>	<b>164</b>
	<b>List of Figures and Tables</b>	<b>187</b>

# Chapter 1

## Introduction

*"There's an infinity of things that have never been done before, and most of those things are not worth doing."* – Jeremy Fox, *Dynamic Ecology* [1]

---

### Résumé du Chapitre 1: Introduction

Malgré leur apparente simplicité, les microorganismes sont des êtres vivants complexes capables de survivre dans les milieux les plus extrêmes. Une telle ubiquité repose en partie sur leurs excellentes capacités d'adaptation. En effet, *via* l'information stockée dans son ADN, une cellule microbienne est capable, par la production des protéines nécessaires, d'assimiler et de se répliquer à partir d'une grande variété de nutriments. Ce mécanisme, que l'on appelle l'expression génétique, permet donc à la cellule de réguler sa composition en fonction des conditions environnementales. Chaque composant cellulaire est initialement issu des ressources trouvées dans l'environnement. Par l'expression génétique, la machinerie cellulaire résout donc un problème de distribution des ressources : un bien commun (les nutriments) est partagé entre plusieurs milliers de bénéficiaires potentiels (les protéines et autres macromolécules).

Selon quels critères les ressources sont-elles distribuées dans la cellule ? Le comportement de la plupart des systèmes vivants peut s'expliquer par les principes de la sélection naturelle. En définissant des facteurs de *fitness* (adaptation au milieu), on peut généralement identifier des objectifs permettant à l'organisme de maximiser sa descendance, et donc de persister sur le long terme dans un milieu donné. Ces facteurs et surtout leurs importances

relatives, peuvent s'avérer extrêmement complexes à identifier, ce qui est notamment le cas chez les microorganismes. Cependant, de nombreuses études ont pu montrer que la maximisation du taux de croissance (ou taux de répliation) est un facteur qui semble commun à de nombreux microorganismes, et qui permet d'expliquer à lui seul une grande variété de comportements.

Peut-on, à l'aide d'un tel principe universel, formaliser la croissance des microorganismes en lois simples, analogues à celles que l'on peut trouver dans d'autres domaines des sciences comme la physique ? Historiquement, la première tentative d'établissement de lois fondamentales de la physiologie microbienne fut la découverte de la loi de Monod il y a plus de 60 ans (Fig. 1.1 et Eq. 1.1). Elle montre que le taux de croissance d'un microorganisme suit une loi hyperbolique de la concentration du nutriment limitant dans le milieu. Depuis, d'autres lois fondamentales ont pu être identifiées. L'une d'entre elles est une loi de croissance qui décrit la relation linéaire existante entre la concentration en ribosomes dans la cellule et le taux de croissance de l'organisme (Fig. 1.2). De manière étonnante, quels que soient les détails moléculaires qui permettent cette adaptation, cette loi émerge lorsque la cellule cherche à maximiser son taux de croissance dans chaque environnement.

Bien que répandues chez de nombreux microorganismes, ces lois se limitent à un état bien particulier qu'on appelle croissance à l'état stationnaire. Cet état est atteint lorsque l'organisme se multiplie pendant suffisamment longtemps dans un environnement stable (Fig. 1.3). Il s'avère très pratique pour les études théoriques et expérimentales, notamment car les composants de la cellule ne changent plus au cours du temps, ce qui lui vaut aussi l'appellation de croissance équilibrée (*balanced growth*). Mais cet état a beau être pratique, il est très loin des conditions rencontrées par les microorganismes dans leur milieu de vie naturel. Ces derniers sont en effet plutôt soumis à des environnements changeants, où les éléments chimiques indispensables à la croissance sont rares, et sont donc les objets d'une rude compétition. Pour quelles raisons les microorganismes seraient-ils optimisés pour un état de croissance qu'ils n'ont quasiment jamais rencontré au cours de leur évolution ? Si comme on s'y attend ces systèmes sont plutôt adaptés à des environnements variables, ratons-nous quelque chose en ne les étudiant qu'en conditions stables ?

Dans cette thèse, nous nous proposons ainsi d'étendre les études des lois de croissance en adoptant une perspective dynamique, c'est-à-dire, en environnements variables : quelles stratégies dynamiques les microorganismes suivent-ils pour redistribuer leurs ressources lors d'un changement environnemental ? Cette question sera développée en deux grands axes d'étude.

Quelles sont les meilleures stratégies de distribution des ressources si, comme pour les lois de croissance à l'état stationnaire, la production de bio-

masse est le seul critère optimisé par la cellule ? Répondre à cette question est une étape nécessaire pour évaluer dans quelles mesures les critères d'optimisation dynamique peuvent différer de ceux à l'état stationnaire, et donc, de tester si les pressions évolutives s'appliquent effectivement différemment entre environnements stables et dynamiques. Cette preuve de concept reposera sur un modèle de la distribution des ressources dans une cellule microbienne, qui sera utilisé pour déterminer quelle est la manière dynamique optimale pour la cellule de distribuer ses ressources lors d'un changement d'environnement (Chapitre 2).

Les prédictions faites seront ensuite testées expérimentalement. Nous chercherons donc à mesurer la distribution des ressources chez *Escherichia coli* lors d'une transition de croissance (Chapitre 3). Nous montrerons comment l'utilisation de marqueurs fluorescents des ribosomes, d'un dispositif microfluidique de culture continue, et de méthodes originales d'analyse des séries temporelles, permet d'observer et de reconstruire l'allocation des ressources au cours d'une transition brutale du milieu, notamment la transition d'un milieu pauvre vers un milieu riche (*nutrient upshift*).

---

## Beginning of Chapter 1

### 1.1 Context

#### 1.1.1 Self-replication is a resource allocation problem

At every moment of our lives, we are surrounded by billions of microscopic organisms. They sustain themselves by harvesting energy from their environment in the form of organic matter, highly reactive chemicals, or light [2, 3]. While most of our cells are kept in a stable and friendly internal environment (temperature, oxygenation, nutrient abundance, ...), microorganisms have evolved to constantly adapt their physiology to strongly variable conditions [4–10]. Not without success, because microbes have been found to survive almost everywhere, sometimes in the strangest places that were initially thought unsuitable for life [2, 3, 11, 12]. Such ubiquity translates a strong potential for adaptation. The model organism *Escherichia coli*, a bacterium first isolated from our intestinal tract, is capable of growing on dozens of different carbon sources through changes in the expression of thousands of genes [13].

Microorganisms, along with all living cells, are mainly constituted of polymers (DNA, RNA, proteins) consisting of many repeated subunits (nucleotides, amino acids). The information about their sequences is stored in the form of genes, physically located on the DNA. Gene expression is the process by which this information is used in the synthesis of a functional final product (protein or RNA). The gene expression machinery (itself constituted of RNA and proteins) synthesizes gene products after binding to target sequences either by transcription (promoter on the DNA), or translation (ribosome binding site on the messaging RNA). In a given cell, there are roughly as many different target sequences as genes [14], which ensures that not all products are synthesized at the same rate. Furthermore, some of these products specifically bind some promoters and stimulate or repress promoter activity and hence gene expression. These products are called transcription factors and allow, in addition to other mechanisms, the regulation of cell composition.

The reorganization of gene expression controls the abundance of enzymes and ribozymes that catalyze the biochemical reactions allowing the cell to perform different functions. For instance, assimilating a given nutrient involves the synthesis of the corresponding uptake protein and of the metabolic enzymes converting the nutrients into reducing power, nucleotides and amino acids [3]. These precursors are then used to produce new macromolecules, closing the loop of self-replication. In order to optimize the overall process in different environmental conditions, the cell must adjust the abundance of every enzyme involved in this network of biochemical reactions. By doing so, the cell actually solves a resource allocation problem, where a common wealth (precursors) has to be distributed over thousands of potential beneficiaries (proteins and other macromolecules).

### 1.1.2 Optimizing biomass production yields a competitive advantage

Which allocation of resources is optimal for a living cell? Systems capable of Darwinian evolution sustain themselves in the long run by fitting to their environment [15]. Living cells with the best fitness are more likely to survive, replicate, and through long-term competition, eliminate their siblings. The optimization criteria that quantitatively measure fitness are called fitness factors. Whether they apply at the species, the population, the organism, or even the gene level has long been debated [15]. An optimal resource allocation would be a gene expression scheme maximizing the fitness factors. But since fitness factors have to take into account the ability to survive, reproduce, and

compete with other organisms in a given (sometimes changing) environment, they are often context-dependent and highly difficult to identify.

For this reason, there is no consensus about what constitutes the main fitness factors of microorganisms. Competition for common resources seems to favor the maximization of offspring, in the sense that a replicator with more descendants will outnumber its competitors in the long term [15]. It is unclear, however, how this general principle translates into specific features of microbial physiology. For instance, the use of genome-scale metabolic network models has shown that metabolism maximizes the production either of biomass or available energy (ATP) depending on the environmental conditions [16]. Microbes can also either optimize their production yield (making the most of available nutrients at the expense of a lower growth rate) or their production rate (growing quickly while wasting part of the available nutrients), the latter being favored in fluctuating environments [17–19]. Overall, it seems that cell optimality drifts in a multidimensional space and involves making several trade-offs that strongly depend on the environmental context [20].

Nevertheless, numerous studies have shown that considering growth rate maximization confers a good predictive power of microorganism physiology. When cultivated in minimal medium with acetate or succinate, *Escherichia coli* uptake rates are correctly predicted by assuming that the metabolic network operates in a mode that maximizes growth rate [21]. This result is not general though and depends on the carbon source and strain used [22]. But even in other situations, the metabolic networks evolve rapidly towards growth-rate maximization if a constant nutrient-providing environment is maintained [22]. The mutants obtained present modifications in the regulation of the synthesis of metabolic enzymes [23], indicating that not only the efficiency of the enzymes is affected, but also their abundance and thus the cellular composition. Overall, with the currently available knowledge and its limitations (discussed in Chapter 4), focusing on growth-rate maximization sounds like a rational choice to understand the growth strategies of microorganisms [24].

### 1.1.3 Growth laws are universal strategies employed by microorganisms

One of the first attempt to formalize the functioning of microbial physiology into elegant, fundamental laws was the discovery of Monod’s law more than sixty years ago [25]. This empirical relationship states that the growth rate  $\mu$  [div h<sup>-1</sup>] of *Escherichia coli* is a hyperbolic function of the concen-



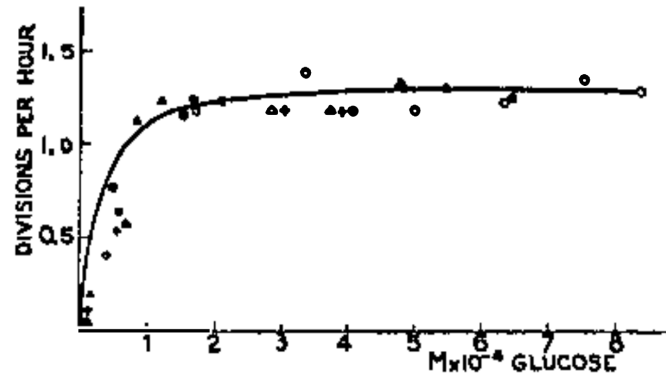


Figure 1.1 – **Monod law, reproduced from [25]**. Monod has been a pioneer in the formalization of microbial growth into fundamental, coarse-grained relationships. This figure displays the steady-state growth rate (in number of divisions per hour) of *Escherichia coli* in a synthetic minimal medium at  $37^{\circ}\text{C}$  containing different concentrations of glucose (a common carbon source). As the concentration of glucose increases, so does the growth rate of *E. coli* by following the hyperbolic relationship described in Eq. 1.1. The solid curve is obtained for  $\mu_K = 1.35 \text{ div h}^{-1}$ , and  $K_C = 0.22 \cdot 10^{-4} \text{ mol L}^{-1}$ .

tration of the limiting nutrient  $C$  [ $\text{mol L}^{-1}$ ], such that

$$\mu = \mu_K \frac{C}{K_C + C}, \quad (1.1)$$

with  $\mu_K$  and  $K_C$  two constants depending on the quality of the nutrient (Fig. 1.1). It is remarkable that the growth rate, a parameter that depends on the use and production of thousands of proteins, can be so easily predicted by a simple relation. What is even more remarkable is that this relation is rather universal. Despite some adjustments (the Droop model being a well-known example [26]), the relation holds for most microorganisms: the growth rate increases with the abundance of the limiting nutrient, up to a point beyond which they cannot grow any faster [27].

The production of every component of the cell is controlled by a complex network of regulatory interactions. But an obvious constraint is that at steady state, the synthesis rate of all individual components must be proportional to the growth rate in order to compensate for growth dilution [25]. Many physiological parameters (like the mass of DNA, RNA and protein) are thus functions of the growth rate alone, regardless of the environmental conditions [28, 29]. These so-called growth laws were carefully measured [29] and are still used today in the quantitative understanding of growth control

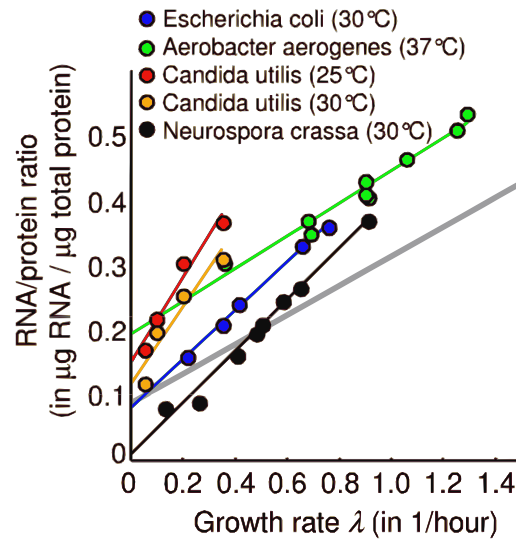


Figure 1.2 – **The growth law of ribosomal abundance (figure reproduced from Fig S1 in [33]).** For a variety of carbon sources and their corresponding steady-state growth rate, the RNA/protein ratio is linearly correlated with the growth rate, a relation that holds for many species of microorganisms. This ratio is correlated with the fraction of ribosome-affiliated proteins, and therefore with the relative abundance of ribosomes in the cell.

in microorganisms [30]. Recently, the topic was revitalized through the work of Scott *et al.* [31]. They focused on the ribosome concentration, a physiological parameter that was long known to vary linearly with the growth rate in microorganisms (Fig 1.2). By measuring it under different environmental perturbations of the protein synthesis machinery, they built a coarse-grained model describing proteome resource allocation [32]. It has allowed to show that, whatever the details of the molecular implementation (differing between organisms), this growth law emerges from the underlying principles of robustness and optimization imposed by natural selection, especially growth-rate maximization, in all conditions [32].

#### 1.1.4 Static versus dynamical perspective on growth

The growth laws cited above apply at steady state, where all intensive properties of the cell are time-invariant [3, 34]. This means that the properties that are independent of the cell volume or the cell mass (temperature, concentrations, ...) are constant over time, even though the cell is growing. It requires that the components of the cell "increase by the same factor over a time interval", which has motivated the use of the term balanced growth [35].

Experimentally, this growth scenario has been used as a standard because it improves reproducibility, in that the results do no longer depend on the precise timing of the samples [3]. It can be easily achieved in the laboratory either in continuous culture, where the substrate is continually supplied [36], or in batch conditions if the substrate is in high excess (Fig 1.3). This approach has been beneficial to mathematical modeling because considering the system at steady state reduces the complexity of the underlying dynamical system governing microbial growth, allowing genome-scale models encapsulating the enzymatic diversity of living cells to be built and analyzed [37].

Although balanced growth is convenient from an experimental and theoretical point of view, it is widely admitted that microorganisms rarely encounter this state in nature [3]. Steady-state growth requires stable conditions over a long period of time, but microorganisms live in environments where the key elements, such as carbon, nitrogen or phosphorus, are quickly depleted by the competitors as soon as they become available [4–6], a consideration that also holds for the lab strains for which growth laws have been established [7–10]. Why would microorganisms be optimal for a state they have barely encountered during their evolution? Are we missing something by studying in stable, unchanging conditions systems that were in fact selected to cope with environmental variations?

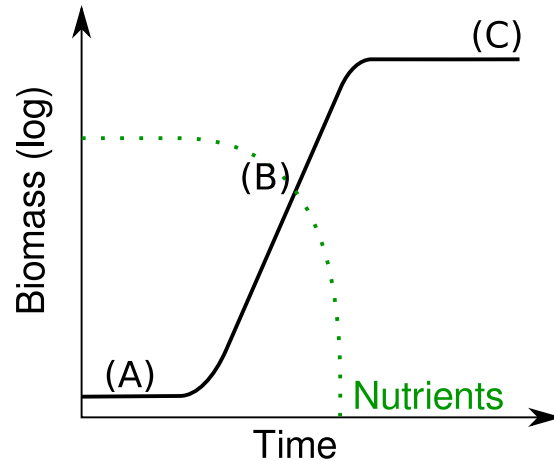


Figure 1.3 – **The different phases of a typical growth curve.** During a typical batch growth scenario, biomass accumulates (black thick line) whereas nutrients are consumed until depletion (green dashed line) [3]. (A) The lag phase is a variable period of time during which the organism adapt to the new medium [38]. It is hard to study experimentally and is known to be affected by the pre-culturing history of the strain [39–41], the magnitude and the rate of the change between the past and present environments [42], and other hard-to-control environmental conditions [43]. (B) The steady-state (or balanced-growth) phase is characterized by an exponential production of biomass. Its characteristics are time-invariant and quite robust accross conditions, which has made it a standard for microbial growth studies [3]. This phase can be extended for hundreds of generations in continuous cultures by the constant renewal of the medium [36, 44, 45]. As represented here, nutrients are quickly depleted in batch conditions, which does not allow to maintain steady-state growth for a long period of time. (C) The stationary phase occurs after the depletion of the limiting nutrients in the medium [3, 46]. In natural conditions, microorganisms usually encounter poor media and spend most of their time in stationary phase [4–6]. Some species that have evolved long-term resistance mechanisms, like sporulation [47] or cannibalism [48], can survive particularly long stationary phases.

## 1.2 Problem statement

Microbial growth is essentially a resource allocation problem that can be summarized in simple, fundamental growth laws. Though established empirically, those laws are elegantly explained if we consider that microorganisms behave in ways that optimize biomass production. However, these laws describe growth in stable environments and it is hard to imagine how natural selection could have resulted in microorganisms that are optimal for conditions they rarely encounter outside the laboratory. This motivates the study of growth laws from a dynamical perspective, that is, in a changing environment, and leads us to the following problem statement: **Which strategies do microorganisms follow in order to dynamically reallocate their resources after a change in the environment?** The development of this question prompts the following investigations.

What are the best strategies of resource allocation if, like for the steady-state growth laws, we continue to assume that biomass production is the only optimization criterion? An answer to this question would allow us to assess to which extent the features of dynamical optimization differ from those of steady-state optimization. In other words, it would test our verbal hypothesis that motivated a dynamical perspective on growth: the expectation that evolutionary pressure applies differently in a changing or stable environment. In order to achieve this, we need to formulate our problem in mathematical terms, through the development of a proof-of-concept model of resource allocation in bacterial cells, and the theoretical determination of the optimal dynamical allocation of resources following a change in the environment.

The model will help us to establish actual resource allocation strategies that optimize biomass production during an environmental change. However, it will not be able to tell us if microorganism do actually optimize such a criterion. We therefore need to measure biomass accumulation and changes in resource allocation following a change in the environment. How could one measure resource allocation during a growth transition? In the light of the discussion in the previous section, this requires monitoring the abundance of key macromolecules in living cells over time, while establishing an experimental set-up that allows to control the inherent variability of dynamical experimental studies of microbial growth.

## 1.3 Related work

### 1.3.1 Modeling growth of microorganisms

My recent experience as a teacher taught me that, as of today, many students in biology do not like mathematics very much. Since mathematics is the main language of modeling, this has often made it difficult for me to convince them that modeling is of key importance for biologists. However, while mathematics is the language, it is not the essence of modeling. Generally speaking, a model depicts and simplifies reality. Maps, sketches, or pictures satisfy this definition, as do graphs, sets of equations, or even the DNA sequences stored in a text file on a computer. In other words, depicting and simplifying reality makes you a modeler, not so much drawing equations on a board. But this does not change the fact that mathematics are powerful tools for formulating and analyzing models [49, 50].

What makes mathematical modeling so useful in microbial growth studies? Microbial physiology results from the interplay of thousands of chemical reactions that are not necessarily relevant in any given situation [3]. These reactions occurs on a wide range of time scales (shorter than 1 second for metabolism, to more than 1 day for the degradation of stable proteins [3]), and are controlled by several layers of regulatory mechanisms that can only be understood through evolutionary considerations [15]. As a consequence, even the simplest verbal hypothesis can resist direct empirical testing. But abstracting away complexity is the purpose of mathematical models. The overwhelming number of variables and the co-existence of several different time scales can be dealt with by choosing the correct framework. When working on a model, microbial behavior is abstracted into a world of clearly stated rules, where it is easier to spell out the logical consequences of the underlying assumptions [49, 50]. Predictions can be made, "unpacked" into our real world, and confronted with experimental testing.

Modeling has proven to be particularly helpful in unveiling how metabolic networks operate. For an increasing number of microorganisms, we are now able to draw quasi-exhaustive maps of their metabolic reactions. For instance, a much-used genome-scale reconstruction of *E. coli* metabolism contains 1366 genes, 2251 metabolic reactions, and 1136 unique metabolites [51]. The number of variables may seem overwhelming, and making sense of this information is not straightforward. Which reactions are important, and in which environmental conditions? Can we predict how perturbations will affect a given metabolic network? Do fundamental regularities exist between different metabolic networks, from different species?

The complexity of the reconstructed metabolic networks does not im-

pede their mathematical study. Constraint-based modeling is a framework that abstracts away the unknown kinetics and represents the metabolic reactions by steady-state fluxes to which physico-chemical constraints can be applied, e.g., compartmentalization, mass conservation, molecular crowding, and thermodynamic directionality [52]. From a mathematical point of view, the model consists of a system of linear equations defining a space of admissible flux distributions that is further reduced by the above-mentioned constraints, which eliminate flux distributions that are unlikely to occur in an environmental condition of interest. The solution space is often further reduced by selecting the flux distributions optimizing a specified objective function, such as the growth rate of the cell. This approach, called flux balance analysis (FBA) [51, 53], has been shown to correctly predict many behaviors of the metabolic network of *E. coli* [21, 54]. It has also proved capable of predicting its long-term adaptation through evolution of the network after a gene deletion [55].

More direct modeling approaches, in the form of large kinetic models, have also been helpful for understanding growth-related processes. Through the construction of a model of 47 differential equations and 193 parameters, Kotte *et al.* have shown how metabolic fluxes are sensed at different locations in central carbon metabolism, and are integrated in a global cellular response by the coupling of enzymatic and transcriptional regulation [56]. Kinetic models of this type are sufficiently detailed to be used as an *in silico* testbed to investigate specific molecular mechanisms and uncover their function in bringing about a cellular response [57]. The most emblematic instance of this approach is probably the recent whole-cell model of *Mycoplasma genitalium* [58]. By aggregating all the available knowledge from more than 900 publications, Karr *et al.* constructed a dynamical model accounting for all the annotated gene functions of this pathogenic organism, and successfully used it to investigate unobserved molecular mechanisms and guide novel experimental analysis. While such approach is not yet applicable to many organisms, it genuinely demonstrates how the increase in computational power could one day guide *in silico* experimentation.

The granularity of the kinetic models cited above is their strength, but also their main weakness. Despite the available knowledge of molecular mechanisms, precisely measuring *in-vivo* kinetic parameters still represents a technological bottleneck [59–61]. While parameter values can be collectively fitted to available data [62, 63], this is known to often produce large parameter uncertainty [64–66]. This is problematic, because the predictions are often particularly sensitive to the parameter values [67–69]. Constraint-based modeling overcomes this limitation by using a mathematical approach that does not require any knowledge about the reaction kinetics. The drawback is that

this approach strongly depends on the constraints used to reduce the space of admissible solutions, e.g. the choice of the objective function that may be tricky, as discussed above [70].

The limitations of constraint-based and detailed kinetic models have motivated the construction of simpler, coarse-grained models. The philosophy is quite different: instead of aggregating all available knowledge into a detailed model, the modeler is concerned about carefully filtering this information to keep the model as simple as possible. The resulting models generally describe cellular functioning on a high level of abstraction, and are particularly valuable when looking for universal laws or principles [31–33]. They can take the form of minimal core models that focus on a given aspect of growth, while still abstracting away the molecular details (see *e.g.*, [71]). They can also be used as proof-of-concepts to submit verbal hypotheses to the logic and rigor of mathematical reasoning [49]. For instance, a simple proof-of-concept model was used to uncover the principles leading to overflow metabolism, a mechanism by which microorganisms switch to inefficient metabolic pathways when growing at high nutrient availability [24]. This paradoxical and widespread phenomenon [72–75] was shown to be easily explained by the fact that high yield pathways also require the synthesis of more enzymes, revealing the occurrence of a cost-benefit trade-off producing the switch when nutrients are no longer the limiting factor. Overall, these models have several advantages for the purpose of our study: they clearly state the underlying assumptions, and are sufficiently tractable to be analyzed by a variety of mathematical tools.

### 1.3.2 Measuring growth of microorganisms

Either to formulate hypotheses or test predictions, data on microbial growth need to be acquired. The most straightforward method is to work at the population level. When working with microorganisms, a clonal population of genetically identical cells can easily be obtained by inoculating a single colony into a growth medium [3]. Over time, this culture can be subjected to different types of measurements: for instance, biomass can be estimated directly through measurements of the dry weight of samples [25], or indirectly through measurements of transmitted or scattered light across the culture [76], allowing to construct growth curves as presented in Fig. 1.3. Other population-wide parameters, like the macromolecular composition [31, 33] or the concentration of metabolites [60] can also be evaluated. They represent averaged estimates from billions of genetically identical cells and are thus usually robust and reliable measurements. In fact, most of *E. coli* parameters that are currently used in today’s growth models have been measured



at the population level, notably through the emblematic work of Bremer and Dennis [29, 77–79].

Nevertheless, some questions cannot be solved at the population level [80–82]. One example are questions about the internal structure of the cell [81]. Bakshi *et al.* used a combination of fluorescent labeling and superresolution imaging to localize the ribosomes, RNA polymerases and DNA in living *E. coli* cells [81]. They showed that despite what was previously thought, most of the translation occurs far from the DNA, on free mRNA molecules that migrate to ribosome-rich regions [81]. Another example concerns questions about single-cell variability [82–84]. Through growth and imaging of single cells in a microfluidic device, Balaban *et al.* showed that a clonal bacterial population can profit from a preexisting heterogeneity to persist when challenged by the presence of an antibiotic, a short-term mechanism that does not involve any genetic mutation [82]. In the long term, heterogeneity has also been shown to help subpopulations to resist lethal stresses, leaving them with enough time to adapt and conquer new ecological niches [83, 84]. In fact, cell heterogeneity seems so crucial for the fitness of microorganisms that diversity-generating mechanisms have been identified [85–87].

In addition to the level of measurement, the cultivation method plays a significant role in the kinds of question that an experiment can answer. With the advent of molecular genetics in the 1970s and 1980s, batch growth conditions have been the preferred choice of most microbiologists [88]. As illustrated in Fig. 1.3, the organism is inoculated in a closed-system and grows until the nutrient is depleted. In a sense, such a condition is close to what microbes encounter in nature, where key nutrients are only available for a short time and quickly depleted [4–6], and is essential to study many questions, *e.g.* diauxic growth [89]. Currently, a strong advantage of batch culturing is the intense parallelization that can be attained using microplates. Indeed, microplates allow to perform dozens to hundreds of growth experiments, each in less than one milliliter, while the optical density or the fluorescence of each culture is automatically monitored. This has proven extremely helpful for screening purposes and the establishment of standard libraries of gene labeling and modification [90, 91]. Parallel cultivation has notably been exploited to construct roughly 4,000 single-cell knock-out mutants of *Escherichia coli*, forming the Keio collection [90]. This mutant collection was extensively used during the last decade to unveil unknown gene functions and test genome-wide effects of gene deletions (as of 2009, more than 4 millions samples issued from this collection had been shared worldwide [92]). Microplates have also helped building a massive library of transcriptional fusions of the green fluorescent protein to each of about 2,000 different promoters in *E. coli* [91]. Interestingly, microplate batch growing conditions have also been used when

applying the above constructions in the inference and analysis of gene regulatory networks [93–97]. But the latter use can be hazardous, because the main drawback of the batch condition is that the culturing system is poorly controlled. Changes of key chemical parameters (*e.g.* pH, pO<sub>2</sub>) have been shown to occur during the whole growth curve [98]. At high density, metabolic by-products can accumulate in the medium and impede growth before nutrients are exhausted [99–101]. This could be dealt with by focusing on the beginning of the culture, but this part has been shown to strongly depend on the preculture history [39–41], and one is never sure when an internal steady state has been reached [102].

For this reason, continuous cultures were developed for microbial studies [44, 102, 103]. Through the permanent renewal of the growth medium, continuous cultures allow for maintenance of a chemically well-controlled environment, allowing for the acquisition of reproducible and reliable data [36, 88]. This has shown to be particularly valuable for comparative omic analysis, for instance the analysis of changes in protein levels [104] or genome-wide transcriptomic changes [105] that occur when *Saccharomyces cerevisiae* is grown in different media. Continuous cultures are more difficult to set-up, however [36, 103]. Even if the working volume can be reduced [106], they usually consume large quantities of medium and are, by design, expected to be more prone to contamination [103]. But recent advances in microfluidic technology have made it possible to set up robust continuous cultures experiments on the microliter scale [45, 82]. For instance, the mother machine [45] allows to perform long-term growth of microorganisms in continuous culture using only a few milliliters of fresh medium per hour. This has enabled to show that *E. coli* growth is remarkably stable on the long term and rather immune to the aging mechanisms that normally affect mother cells in other microorganisms. It is however important to note that at those scales, growth can only be monitored through microscopy analysis, which could unnecessarily complicate studies that do not rely on single-cell measurements.

The experimental approaches briefly reviewed above can be applied to study microbial growth in a dynamical setting [30, 107–109]. For instance, Levy *et al.* [107, 108] applied pulse-like environmental perturbations to a continuous culture of yeast, and measured how transcriptomic reorganization occurs. They observed that the transcription levels are affected before the growth rate of the organism, suggesting a significant role of feed-forward sensing from the environment. Madar *et al.* [109] analyzed the transcriptional reorganization occurring in *E. coli* during the lag phase through the coupling of batch experiments in microplates and single-cell measurements *via* flow cytometry. They showed that *bottleneck enzymes* were produced early in the lag phase, before the cells actually switch to the production of ribosomes and

general metabolic enzymes. As in this PhD project, the studies cited above investigate questions that can only be answered in a dynamical context and they inspired the search for the most appropriate experimental approaches to answer the problem set out in Section 1.2. By taking the best of the works presented above, a part of our study will focus on establishing such experimental conditions.

## 1.4 Approach

Throughout this study, we focus on a specific dynamical growth scenario, namely the case of a nutrient upshift [30, 110–112]. Contrary to steady-state growth, nutrient upshifts and downshifts are frequently encountered in the life cycle of microorganisms [3–6]. In combination, they also provide a good approximation of more complex environments. In nature, nutrient upshifts generally start from stationary phase [4–6], a state in which complex adaptive mechanisms are at work that we would like to sidestep for this study [39–43, 47, 48]. For this reason, we confine the problem to the study of resource allocation during a so-called steady-state-to-steady-state transition. While we expect the starting and ending conditions to be on the growth law presented in Fig. 1.2, we currently have no idea of what happens in the time between.

We start by developing in Chapter 2 a simple proof-of-concept model of resource allocation that evaluates how biomass production can be maximized during an upshift from a medium with low nutrient content to a medium with high nutrient content. The model is an instance of a self-replicator model [24], but focuses on the allocation of resources to only two sectors of the microbial cell: metabolism, taking up and converting nutrients to precursor metabolites, and gene expression, producing macromolecules from the precursors. The model is kept as simple as possible in order to make sure it stays mathematically tractable in a dynamical context. A necessary step is however to verify that, at steady-state, the model account for known growth laws of resource allocation [31, 33].

Using this model, we pose the problem of dynamical optimization as an optimal control problem [113]. To comply with what we know at steady state, the microbial cells are assumed to maximize biomass production, but we adapt the criterion to the dynamical context of an upshift scenario. By using a combination of analytical [114] and numerical optimization [115], we aim to identify a mathematical upper bound for the biomass produced [113]. This theoretical solution can then be used to explore possible regulatory strategies. In particular, we use the bang-bang [113] control solution found

as a benchmark to identify the system variables that need to be sensed in order to optimize resource allocation. This provides a common scale on which the outcomes of different regulatory schemes can be compared during an upshift scenario. This allows us to identify one or several strategies that clearly outperform all the others, giving us testable predictions. Specifically, we show that feedback from the intracellular state turns out to be much more valuable than information from the environment. We investigate if such a strategy could be active in real cells, and identify the widespread ppGpp system [116] as a possible molecular implementation that could control the synthesis of ribosomes in a switch-like manner during a nutrient upshift. Overall, this chapter demonstrates that, if microbial cells actually optimize their biomass production during an upshift, they should dynamically allocate their resources to the gene expression machinery in an on-off manner, making an experimentally testable prediction.

In Chapter 3, we address the challenging problem of experimentally verifying this prediction in *Escherichia coli*. This task requires the development of an experimental set-up allowing the measurement of resource allocation in our dynamical growth scenario, *i.e.* a steady-state-to-steady-state nutrient upshift. From the model developed in Chapter 2, we identified that the predicted on-off pattern should occur on a short time scale after the transition (a couple of generations). This motivates the use of *in-vivo*, high-frequency, single-cell measurements of the concentration of the gene expression machinery. Inspired by the work of Bakshi *et al.* [81], we constructed a strain with a GFP-tagged S2 ribosomal subunit, grew this strain in the mother machine, and monitored expression of the reporter gene using fluorescence microscopy [45]. While originally developed for studies of long-term steady-state growth, we used the mother machine here to perform an instantaneous medium transition in a controlled manner from acetate (a poor carbon source) to glucose (a rich carbon source), in order to obtain a significant difference in growth rates between the two media [117].

After segmentation and cell tracking, this set-up was able to generate fluorescence and size time series for dozens of cells. Using again the self-replicator model as a framework, we showed that these measurements are sufficient for the intended signal reconstructions, in particular the growth rate and the resource allocation profile over time. The method we used is called Kalman smoothing [118, 119]. While well-known in engineering, Kalman smoothing has been rarely used in quantitative biology, but turned out to be particularly suitable for the purpose of the model-based analysis of gene expression data in single cells, as we argue and show on synthetic data in Chapter 3. Finally, the experimental verification or falsification of the optimal control prediction, should allow to conclude whether the biomass

maximization, as considered in Chapter 2, is indeed a cell objective in a dynamical context.

## Chapter 2

# Dynamical Allocation of Cellular Resources as an Optimal Control Problem: Novel Insights into Microbial Growth Strategies

*"Rather than propose a new theory or unearth a new fact, often the most important contribution a scientist can make is to discover a new way of seeing old theories or facts."* – Richard Dawkins, *The Selfish Gene* [15] (preface to 1989 edition).

---

### Résumé du Chapitre 2: Modélisation de l'allocation dynamique des ressources cellulaires comme un problème de contrôle optimal

Ce chapitre est dédié à une approche théorique du problème d'allocation des ressources cellulaires chez les microorganismes. En Section 2.2.1, nous modélisons la cellule comme un auto-réplicateur devant allouer une ressource (les nutriments du milieu) à deux secteurs distincts de macromolécules : la machinerie métabolique qui extrait les nutriments de l'environnement et les convertit en précurseurs utilisables, et la machinerie d'expression génique qui utilise ces précurseurs pour produire de nouvelles macromolécules (Fig. 2.1). Ce modèle très simple comprenant 2 réactions (Eq. 2.1) se traduit mathématiquement par deux équations différentielles ordinaires (Eqs 2.3 et 2.4) faisant

intervenir le paramètre  $\alpha$  qui va devenir clé dans notre étude.  $\alpha$  représente la proportion massique des précurseurs qui sont utilisés pour produire de la machinerie d'expression génique, au détriment de la machinerie métabolique (qui elle, est produite à la proportion  $1 - \alpha$ ). Comme on l'a déjà vu avec les lois de croissance établies à l'état stationnaire (voir Chapitre 1), cette proportion varie avec l'environnement, on s'attend donc à ce que  $\alpha$  en fasse autant, augmentant d'autant que l'environnement devient riche (Fig. 1.2). Cependant, de quelle manière varie-t-il entre deux états de croissance stationnaires, lorsque les composants cellulaires sont en déséquilibre et que la cellule doit adapter sa composition ?

Comme nous l'avons vu dans le Chapitre 1, les lois de croissance à l'état stationnaire s'expliquent facilement si l'on considère que la cellule maximise son taux de croissance. C'est aussi le cas dans notre modèle, et en Section 2.2.2 nous voyons que pour chaque environnement, la croissance de la cellule est maximale pour une valeur unique de  $\alpha$ . Mais surtout, cette valeur que l'on nomme  $\alpha_{opt}^*$  est d'autant plus grande que l'environnement est riche (Fig. 2.2). C'est finalement toutes les lois de croissance à l'état stationnaire qui sont prédites par notre modèle simple, sous la simple hypothèse qu'à tout moment, la cellule maximise son taux de croissance (Fig. 2.3). Peut-on appliquer le même principe, dans un contexte cette fois dynamique, de façon à prédire comment la cellule distribue ses ressources lors d'une transition de croissance ?

Dans ce contexte,  $\alpha$  est désormais une fonction du temps. Nous cherchons donc quelle fonction du temps maximise un objectif donné. Ce problème, formulé en Section 2.2.3, est un problème de contrôle optimal, puisque l'objectif à optimiser est une fonction d'une autre fonction, plus communément appelée fonctionnelle. Nous choisissons comme équivalent dynamique de la maximisation du taux de croissance, la maximisation de la production de biomasse sur un intervalle de temps comprenant un changement abrupt de l'environnement (Eq. 2.9). Nous résolvons ce problème en Section 2.2.4 en couplant l'utilisation du principe du maximum de Pontryagin avec de l'optimisation numérique. La solution optimale obtenue montre que le contrôle  $\alpha$  doit prendre alternativement les valeurs 0 ou 1 jusqu'à ce que le système soit stabilisé sur un nouvel équilibre (Fig. 2.4 et Eq. 2.14). En d'autres termes, la meilleure production de biomasse s'obtient si, à tout moment, la cellule aiguille 100% de ses ressources vers un seul secteur, en alternant entre l'un et l'autre jusqu'à ce que le nouvel état stationnaire de croissance soit atteint. Ce type de solution se rencontre souvent en contrôle optimal, et répond au nom de concept TOR (Tout-Ou-Rien, ou *bang-bang* en anglais). Mais ce contrôle est-il pertinent d'un point de vue biologique ? La cellule a-t-elle suffisamment d'information à sa disposition pour réaliser une telle transition, laquelle re-

pose sur des changements abrupts de l'expression génique à des instants bien précis ?

En Section 2.2.5, nous utilisons le problème de contrôle optimal et sa solution comme un banc d'essai permettant de comparer entre elles différentes stratégies de régulation (Fig. 2.5). En se servant du modèle pour guider notre intuition, et en nous limitant aux solutions qui respectent les lois de croissance stationnaire, nous montrons que deux stratégies de régulation simples sont possibles : soit l'organisme mesure directement la quantité de nutriments dans l'environnement, soit il mesure la concentration en précurseurs dans la cellule. De manière intéressante, ces deux stratégies sont optimales à l'état stationnaire, et donc strictement équivalentes. Cependant, dans le contexte dynamique de notre banc d'essai, la stratégie qui consiste à mesurer les précurseurs est bien plus efficace, même si elle reste évidemment inférieure à la solution optimale (Fig. 2.6).

Finalement, nous montrons en Section 2.2.6 qu'une stratégie plus complexe mesurant à la fois les précurseurs et la concentration en machinerie d'expression génique est tout à fait capable de réaliser un contrôle proche du tout-ou-rien optimal (Eq. 2.19 et Fig. 2.7). Mais surtout, en réutilisant un modèle du système ppGpp, connu pour réguler la synthèse des ribosomes chez la bactérie *Escherichia coli*, nous montrons que ce système répond, au moins structurellement, aux exigences d'une telle stratégie de régulation (Fig. 2.8). Cela éclaire d'un jour nouveau le fonctionnement et l'origine de ce système de régulation largement répandu chez de nombreux microorganismes. En intégrant l'information provenant à la fois de la quantité de ribosomes et de celle des précurseurs, ce dernier est capable d'allouer efficacement les ressources de la cellule lorsque celle-ci doit s'adapter rapidement à un nouvel environnement.

---

## Beginning of Chapter 2

**Important note:** Text and figures in this chapter have been published in Plos Computational Biology [120] under the terms of the [Creative Commons Attribution License](#) (CC-BY 4.0).

### 2.1 Introduction

Microorganisms adapt their physiology to changes in nutrient availability in the environment. This involves changes in the expression of a large



number of genes, encoding proteins with a variety of cellular functions, such as transporters for the uptake of nutrients, enzymes for the conversion of nutrients to energy and building blocks for macromolecules, the components of the transcriptional and translational machinery, and transcription factors to preferentially direct RNA polymerase to specific promoters [3, 14]. Fundamentally, the reorganization of gene expression in response to changes in environmental conditions is a resource allocation problem. It poses the question how microorganisms redistribute their protein synthesis capacity over different cellular functions when constrained by the changing environment.

The mechanisms responsible for resource allocation in microbial cells are usually assumed to have been optimized through evolution, so as to maximize the offspring of cells in their natural environment. How this general principle manifests itself on the level of cellular physiology is not straightforward though. Many studies have reasoned that growth-rate maximization provides a selective advantage to microorganisms, because it allows competitors to be outgrown when resources are scarce. Others have shown, however, that appropriate optimization criteria will depend on the structure of the environment and the ecosystem, as well as on the molecular properties of metabolic pathways [16–20]. For instance, in environments without competition for a shared resource, maximization of growth yield rather than growth rate is expected to provide a selective advantage. Although what counts as optimal is thus context-dependent, growth and evolution experiments in *Escherichia coli* have shown that in certain conditions bacterial metabolism is indeed geared towards growth-rate maximization [21–23].

For this reason, growth-rate maximization is a central hypothesis in a number of recent theoretical studies of resource allocation using coarse-grained models of the cell [24, 32, 33]. The models deliberately reduce the molecular complexity of regulatory networks so as to focus on generic explanatory principles [49]. Along these lines, Molenaar *et al.* developed a series of simple models of the microbial cell, taking into account that growth requires the synthesis of proteins playing a role in metabolism (transporters, enzymes) and gene expression (ribosomes), in varying proportions. Allocation parameters that maximize the growth rate were shown to account, at least in a qualitative way, for the variation of the amount of ribosomal protein as a fraction of total protein in different growth media, and for the occurrence of overflow metabolism above certain growth rates [24]. Using another coarse-grained model of the cell, centered on amino acid supply (metabolism) and demand (protein synthesis), Scott *et al.* derived empirical growth laws with linear relations between the ribosomal protein fraction and the growth rate, in conditions where the nutrient supply or demand are altered [32, 33]. In their model, maximization of growth rate requires maximization of amino

acid flux and is achieved for a specific, unique value of the ribosomal protein fraction. Based on a structurally similar model, Maitra and Dill related optimal resource allocation to the basic constants of the metabolic and gene expression machinery, in particular energy efficiency [121].

The assumption of growth-rate maximization may lead to correct predictions in some situations, but ignores the regulatory mechanisms achieving resource allocation and therefore cannot provide a causal explanation of cellular behavior [89]. Several studies have used coarse-grained models to understand which control strategies microorganisms employ to achieve (optimal) resource allocation [32, 116, 122]. Scott *et al.* have shown that a robust feed-forward control strategy, based on the sensing of the amino acid pool size and the corresponding adjustment of the fraction of ribosomes producing ribosomal proteins, allows the ribosomal protein fraction to be maintained close to its optimal value under a variety of growth conditions [32]. The authors suggest that this control strategy involves the signalling molecule ppGpp, in agreement with conclusions drawn from a recent kinetic model of the regulatory mechanisms achieving optimal adjustment of the ribosomal protein fraction [116]. Weiße *et al.* also developed a coarse-grained model of microbial growth based on resource allocation trade-offs [122]. Without including specific regulatory interactions, the model accounts for the above-mentioned bacterial growth laws, predicts host-circuit interactions in synthetic biology, and relates gene regulation to the nutrient composition of the medium.

The above studies consider resource allocation at steady state, where all intensive variables describing the growing microbial culture, in particular the concentrations of its molecular components, are constant (see [34] for a precise definition of steady-state growth and the closely related notions of balanced and exponential growth). This requires an environment to be stable over a long period of time. Such conditions can be achieved in the laboratory [36], but many microorganisms naturally experience frequently-changing conditions. For example, *E. coli* can cycle between two distinct habitats, the mammalian intestine and the earth's surface (water, sediment, soil) [7]. The bacteria transit through different microenvironments in the intestinal system, where they encounter different mixes of sugars [8]. They are even more challenged in the open environment outside the host, with a greatly fluctuating availability of carbon and energy sources and a large variability in temperature, osmolarity, oxygen, and microbial communities [9, 10].

This situation motivates a dynamical perspective on microbial growth and resource allocation [30, 123–125]. However, fundamental results like the growth laws uncovered for steady-state conditions are still lacking. In particular, extending the results reviewed above to dynamical conditions raises

the following questions: Are control strategies that maximize steady-state growth also optimal in dynamical environments? If this is not the case, then which alternative strategies would be optimal for such conditions? And finally, how do these strategies compare with the regulatory mechanisms that have actually evolved in microorganisms?

The aim of this study is to address the above fundamental questions in a specific dynamical growth scenario, namely a transition between two steady states following an environmental perturbation. In particular, we consider the upshift of a microbial culture from a medium supporting growth at a low rate to a medium supporting growth at a high rate [30]. We develop a coarse-grained model of the cell, inspired by the self-replicator model of Molenaar *et al.* [24], and reformulate our questions in the context of optimal control theory [113] to identify control schemes maximizing biomass production over an interval of time, the dynamical equivalent of growth-rate maximization.

We show that Pontryagin's Maximum Principle suggests that optimal resource allocation after a growth transition is achieved by a bang-bang-singular control law [113], a conjecture confirmed by direct numerical optimization. This optimal solution provides a gold standard against which possible control strategies of the cell can be compared. We consider simple strategies that drive the system to the steady state enabling growth at the maximal rate in the new medium, after the upshift. In a dynamical growth scenario, the strategy sensing the concentration of precursor metabolites emerges as the best candidate, consistent with the analysis of Scott *et al.* that feedforward activation of the rate of synthesis of ribosomal proteins, involving ppGpp-mediated sensing of the amino acid pool [126–128], is the key regulatory mechanism for growth control. It is possible, however, to define a strategy approaching the theoretical optimum even more closely by exploiting information on both the precursor concentration and the abundance of the gene expression machinery. Interestingly, a thorough analysis of the functioning of the ppGpp system, as described by a kinetic model of the synthesis and degradation of this signalling molecule, suggests similarities between our two-variable control strategy and the regulation of the transcription of ribosomal RNA by ppGpp [116].

The results presented here generalize the analysis of control strategies enabling optimal growth of microorganisms from steady-state to dynamical scenarios. The control strategies are formulated in the context of a coarse-grained model of resource allocation, based on minimal assumptions, that accounts for empirical growth laws at steady state. The analysis shows that during growth transitions, control strategies based on information of a single variable are outperformed by systems measuring several variables. This conclusion agrees with the intuition that, in dynamical environments, there may

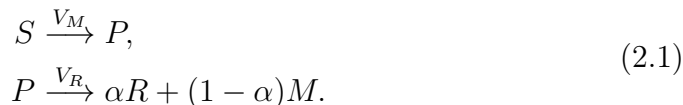
be an evolutionary pressure towards more elaborate sensory systems. From a methodological point of view, our study illustrates how optimal control theory can provide novel insights into complex biological phenomena [129].

## 2.2 Results

### 2.2.1 Self-replicator model of resource allocation

Resource allocation in bacteria involves the distribution of cellular resources (precursor metabolites and energy) over processes supporting maintenance and growth [3]. A simple modelling tool for analyzing resource allocation questions in a precise way are so-called self-replicator models. These models have a long history in various domains of chemistry, biology, physics, and computer science [130], and were recently put to use as an analytical tool in systems biology [24] (see also [131]). We will show that despite their simplicity, which make them tractable for mathematical analysis, self-replicator models are sufficiently expressive to account for empirical observations and make testable predictions.

Bearing in mind that the major constituents of the cell are macromolecules (DNA, RNA, proteins), produced from precursor metabolites, a fundamental resource allocation question is the following: How much of the cellular resources are invested in the making of new macromolecules (gene expression machinery) and how much in performing other functions, in particular producing metabolic enzymes involved in the uptake of nutrients and their conversion to precursor metabolites (metabolic machinery)? In order to address this question, we consider a self-replicating system composed of the gene expression machinery ( $R$ ) and the metabolic machinery ( $M$ ). The system, shown schematically in Fig. 2.1, is thus defined by two macroreactions which are conveniently written as:



The first reaction, catalyzed by  $M$ , converts external substrates ( $S$ ) into precursor metabolites ( $P$ ). The second reaction, catalyzed by  $R$ , converts precursors into macromolecules ( $R$  and  $M$ ). The resource allocation parameter  $\alpha \in [0, 1]$  defines the proportion of precursor mass used for making gene expression machinery as compared to metabolic machinery. We will interchangeably use the symbols  $M$ ,  $R$ ,  $S$ , and  $P$  for the components of the replicators themselves and their total mass [g]. We will denote the rates at which the macroreactions occur by  $V_R$  and  $V_M$  [g h<sup>-1</sup>].

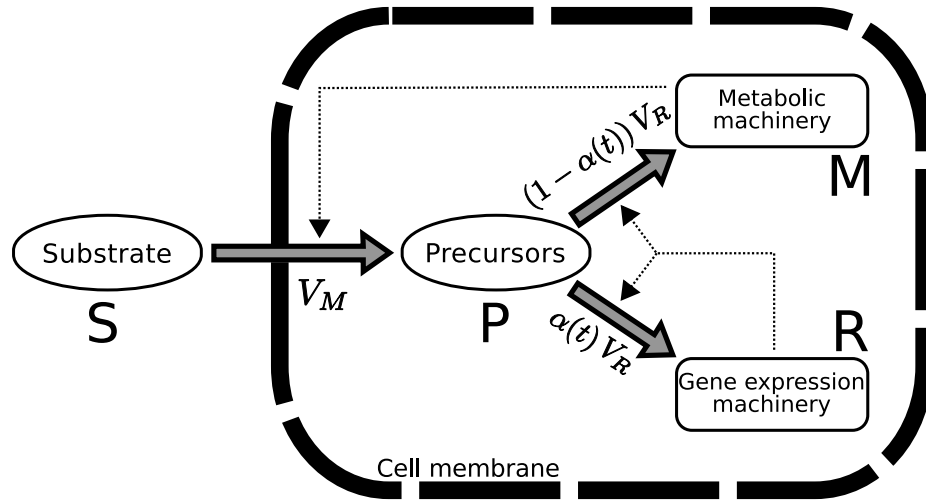


Figure 2.1 – **Self-replicator model of bacterial growth.** External substrates  $S$  enter the cell and are transformed into precursors  $P$  through the action of the metabolic machinery  $M$ . The precursors are used by the gene expression machinery  $R$  to make the proteins composing both the metabolic machinery (transporters, enzymes, ...) and the gene expression machinery itself (RNA polymerase, ribosomes, ...).  $\alpha$  ( $1 - \alpha$ ) is the mass proportion of precursors converted into  $R$  ( $M$ ). Thick arrows denote reactions and thin, dashed arrows denote catalytic activities. The rate of synthesis of precursors and the rate of synthesis of proteins from precursors are denoted by  $V_M$  and  $V_R$ , respectively.

The self-replicator system in Fig. 2.1 is based on a number of simplifying assumptions. First, cell division is not explicitly modelled and replication should therefore be interpreted as the growth of (the mass of) a cell population. This amounts to the assumption that individual cells in a growing population have the same macromolecular composition. Second, degradation of the macromolecules is ignored. In other words, we assume that macromolecules are stable and that their degradation rates are negligible with respect to the rates of other reactions in the system. Third, we consider only two classes of macromolecules ( $R$  and  $M$ ). In particular, we do not assume that an irreducible mass fraction of the precursors is dedicated to cell maintenance [33]. The system could be easily extended to relax the above assumptions, but this would complicate the analysis of the model and obscure the points we want to make.

In what follows, it will be more convenient to describe the quantities in the system as intracellular concentrations rather than as the total mass in

the cell population. To this end, we define the volume  $\text{Vol}$  [L] of the cell population as follows:

$$\text{Vol} = \beta (M + R), \quad (2.2)$$

with  $\beta$  a conversion constant [ $\text{L g}^{-1}$ ] equal to the inverse of the cytoplasmic density. Dividing each variable  $M$ ,  $R$ , and  $P$  by  $\text{Vol}$  yields the concentrations  $m$ ,  $r$ , and  $p$  of metabolic enzymes, ribosomes and other components of the gene expression machinery, and precursor metabolites, respectively [ $\text{g L}^{-1}$ ]. Henceforth, these variables as well as  $\text{Vol}$  and  $\alpha$  will be considered functions of time  $t$  [h].

The dynamics of the self-replicator in Fig. 2.1 can be described by the following system of ordinary differential equations (see S1 Text for the derivation):

$$\frac{dp}{dt} = v_M(s, r) - v_R(p, r) (1 + \beta p), \quad (2.3)$$

$$\frac{dr}{dt} = v_R(p, r) (\alpha(t) - \beta r), \quad (2.4)$$

where  $s$  [ $\text{g L}^{-1}$ ] denote the (extracellular) concentration of substrate.  $v_M(s, r)$  [ $\text{g L}^{-1} \text{h}^{-1}$ ] and  $v_R(p, r)$  [ $\text{g L}^{-1} \text{h}^{-1}$ ] denote the precursor synthesis rate and the macromolecule synthesis rate, respectively. The growth rate  $\mu$  [ $\text{h}^{-1}$ ] of the replicator system is defined as the relative increase of the volume, and can be rewritten with Eqs 2.3-2.4 as proportional to the macromolecule synthesis rate (S1 Text):

$$\mu = \frac{1}{\text{Vol}} \frac{d\text{Vol}}{dt} = \frac{1}{M + R} \frac{d(M + R)}{dt} = \beta v_R(p, r). \quad (2.5)$$

The precursor concentration changes through the joint effect of the precursor synthesis rate  $v_M(\cdot)$ , the macromolecule synthesis rate  $v_R(\cdot)$ , and the rate of growth dilution ( $\beta v_R(\cdot) p$ ). The change in concentration of ribosomes and other components of the gene expression machinery is the net effect of the ribosome synthesis rate ( $\alpha(\cdot) v_R(\cdot)$ ) and the rate of growth dilution ( $\beta v_R(\cdot) r$ ). Remark that it is not necessary to add an equation for  $m$  because it follows from Eq. 2.2 that  $r + m = 1/\beta$ , and therefore  $dm/dt = -dr/dt$ .

We use Michaelis-Menten kinetics to define the synthesis rate of each reaction:

$$v_M(s, r) = k_M m \frac{s}{K_M + s} = k_M (1/\beta - r) \frac{s}{K_M + s}, \quad (2.6)$$

$$v_R(p, r) = k_R r \frac{p}{K_R + p}, \quad (2.7)$$

with rate constants  $k_M, k_R$  [ $\text{h}^{-1}$ ] and half-saturation constants  $K_M, K_R$  [ $\text{g L}^{-1}$ ]. Note that the rate of precursor synthesis is proportional to the concentration of the components of the metabolic machinery, while the macromolecule synthesis rate is proportional to the concentration of the components of the gene expression machinery. These catalytic effects correspond to the dashed arrows in Fig. 2.1. The rate constant  $k_M$  depends both on the quality of the nutrients in the medium (higher  $k_M$  for a richer medium) and on the metabolic efficiency of the macroreaction converting the substrate into precursors (higher  $k_M$  for a more efficient reaction). For convenience, we henceforth assume that the environmental conditions do not change over the time-interval considered, either because  $s$  is constant or because  $s \gg K_M$ , corresponding to a situation in which the substrate is available in excess. In both cases,  $e_M(s) = k_M s / (K_M + s)$  is approximately constant, so that we can write

$$v_M(r) = e_M (1/\beta - r). \quad (2.8)$$

The rate constant  $k_R$  characterizes the efficiency of the gene expression machinery, depending on the elongation rate of ribosomes, among other things. The ratio  $p/K_R$  is an indicator of the saturation of the gene expression machinery by precursors.

The system of Eqs 2.3-2.4 thus has four parameters ( $e_M, k_R, K_R, \beta$ ), one of which characterizes the input from the environment ( $e_M$ ). The order of magnitude of the parameters can be inferred from data in the literature, as explained in S2 Text. Below we use the following values for the parameters  $e_M = 3.6 \text{ h}^{-1}$ ,  $k_R = 3.6 \text{ h}^{-1}$ ,  $K_R = 1 \text{ g L}^{-1}$ ,  $\beta = 0.003 \text{ L g}^{-1}$  (S1 Table). However, it should be emphasized that the conclusions of this paper do not depend on the exact quantitative values of these parameters.

An interesting property of the model is that it is built on minimal assumptions, basically the two macroreactions and the definition of the volume as proportional to the total mass of macromolecules. Like in [24, 32, 123], these assumptions directly lead to the expression of the growth rate in Eq. 2.5, without additional assumptions.

### 2.2.2 Growth-rate maximization of the self-replicator reproduces bacterial growth laws

The nullcline for  $r$  is given by  $r = 0$ ,  $r = \alpha/\beta$ , and  $p = 0$ , while the nullcline for  $p$  is defined by

$$r = \frac{e_M}{\beta \left( e_M + k_R \frac{p}{K_R + p} (1 + \beta p) \right)}.$$



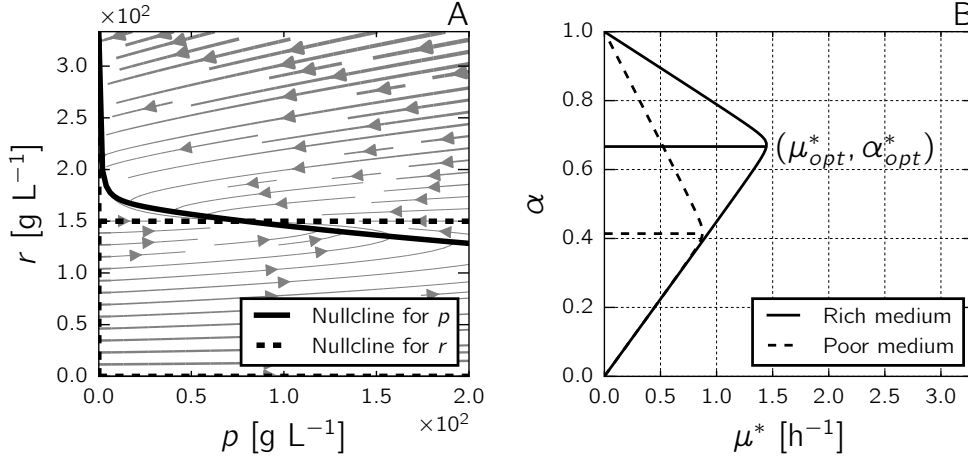


Figure 2.2 – **Analysis of self-replicator model of bacterial growth.** *A*: Phase-plane analysis of the self-replicator model of Eqs 2.3 and 2.4. The nullclines for  $p$  and  $r$  are shown as solid and dashed curves, respectively. Parameter values are  $e_M = 3.6$  h<sup>-1</sup>,  $k_R = 3.6$  h<sup>-1</sup>,  $K_R = 1$  g L<sup>-1</sup>,  $\beta = 0.003$  L g<sup>-1</sup>,  $\alpha = 0.45$ . *B*: Dependence of the growth rate at steady state  $\mu^*$  on the resource allocation parameter  $\alpha$ , for two different environmental conditions (solid line,  $e_M = 4.76$  h<sup>-1</sup>; dashed line,  $e_M = 1.57$  h<sup>-1</sup>, other parameter values are  $k_R = 2.23$  h<sup>-1</sup>,  $K_R = 1$  g L<sup>-1</sup>, and  $\beta = 0.003$  L g<sup>-1</sup>). The maximal growth rate is attained for a unique  $\alpha$ , called  $\alpha_{opt}^*$ .

The nullclines define a single stable steady state  $(p^*, r^*)$  (Fig. 2.2A and *Methods*). At this steady state, the growth rate is constant and denoted by  $\mu^*$ . The nullcline for  $p$  is defined by the environment  $e_M$ . The nullcline for  $r$ , and thus the location of the steady state with the associated growth rate, are given by  $\alpha$ . Fig. 2.2B shows the dependency of the steady-state growth rate  $\mu^*$  on the resource allocation parameter  $\alpha$ . As can be seen,  $\mu^*$  is maximal for a specific, unique value of  $\alpha$ , which we denote  $\alpha_{opt}^*$ . That is, the model predicts that there is a single optimal way to divide the precursor flux over the synthesis of the gene expression machinery and the metabolic machinery. The same result, using a similar model, was obtained by Scott *et al.* [32]. The self-replicator model is simple enough to derive an algebraic expression for computing  $\alpha_{opt}^*$  and the corresponding maximal growth rate  $\mu_{opt}^*$  (*Methods* and S1 Text), which will simplify analysis of the system in later sections.

In order to validate the model, we verified that it can account for data on the macromolecular composition of *E. coli* at steady state [33]. When optimizing  $\alpha$  for different values of  $e_M$  (assuming cells attain maximal growth),



the model predicts a relation between  $\alpha_{opt}^*$  and  $\mu_{opt}^*$  (colored dots and black dashed line in Fig. 2.3A) that is quasi-linear for high growth rates. We compared this prediction with the results of experiments where the relation between the growth rate and the mass ratio of total RNA and protein was determined in different growth media (Fig. 2.3B). In the framework of our model, different media correspond to different values of  $e_M$ , and different total RNA/protein mass ratios to different values of  $\alpha$  (up to a conversion factor), allowing a direct comparison of the model predictions in Fig. 2.3A with the data in Fig. 2.3B (see *Methods*). As can be seen, the model is able to account for the observed quasi-linear relation between the growth rate and the total mass ratio of RNA and protein. Moreover, for realistic values of  $k_R$  and  $e_M$ , a good quantitative fit is obtained (*Methods* and S1 Table).

The data from Scott *et al.* also reveal a second apparently linear relation between the growth rate and the total RNA/protein mass ratio. This relation is obtained when varying, in the same growth medium, the efficiency of protein synthesis by adding different doses of an inhibitor of translation (chloramphenicol) [33]. Using the model, we computed  $\alpha_{opt}^*$  and  $\mu_{opt}^*$ , for constant environment  $e_M$  and different values of the efficiency of protein synthesis  $k_R$  (dashed colored lines in Fig. 2.3A). As can be seen in Fig. 2.3B, the model also captures the second linear relation in the data.

We conclude that the self-replicator model is able to reproduce known observations of resource allocation in bacteria, so-called growth laws [33]. The model is similar to a model recently proposed by Scott *et al.* [32]. Contrary to the latter model, the translation rate is not assumed to be constant in the self-replicator model, but rather depends on precursor abundance, as proposed by the same authors in [132].

The above analysis of bacterial growth has two major limitations. First, the predictions of optimal resource allocation (the value of  $\alpha$  leading to the maximal growth rate) hold at steady state, for a constant environment, whereas most bacteria are not expected to encounter such conditions outside the laboratory. An allocation of resources that is optimal for steady-state growth and constant over time may not be optimal in dynamical growth conditions. Second, while it predicts which value of  $\alpha$  is optimal at steady state, the model says nothing about the strategies that could be used to control resource allocation and set  $\alpha$  to its optimal value. In other words, how could bacterial cells use sensors of changes in their internal state and the environment to optimally adjust  $\alpha$ ? In what follows, we will address the above two questions, after having given a precise statement of the problem of optimal resource allocation in a dynamical environment in the next section.

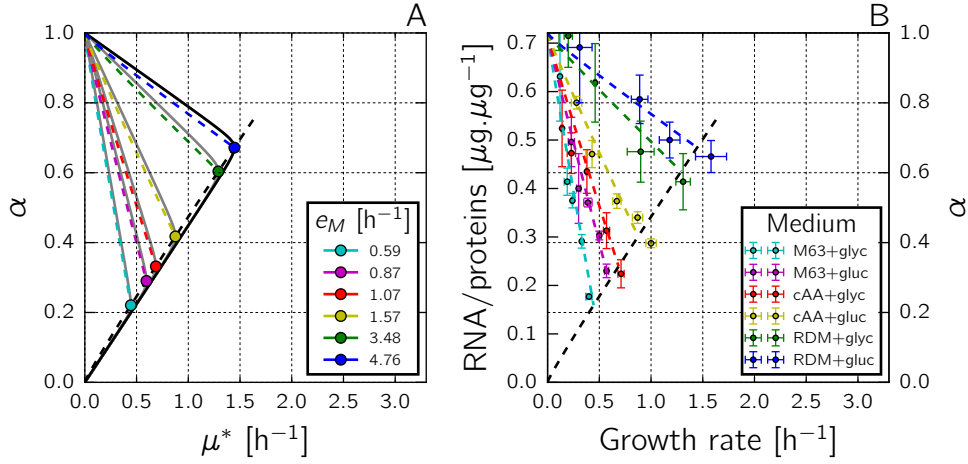


Figure 2.3 – **Self-replicator model accounts for bacterial growth laws.**

*A*: Predicted quasi-linear relation between the maximal growth rate  $\mu_{opt}^*$  and the corresponding optimal resource allocation  $\alpha_{opt}^*$ , for different values of  $e_M$  (different colors). The colored dots indicate  $\alpha_{opt}^*$  and  $\mu_{opt}^*$  for  $k_R = 2.23$  h<sup>-1</sup> and different  $e_M$ , and the dashed black line the relation for all intermediate values of  $e_M$ . The dashed colored lines indicate the relation between  $\alpha_{opt}^*$  and  $\mu_{opt}^*$  obtained when, for a given value of  $e_M$ , the value of  $k_R$  is decreased (lower  $k_R$  leads to lower  $\mu_{opt}^*$ ). The solid grey curves correspond to  $(\mu^*, \alpha)$ -profiles like those shown in Fig. 2.2B. *B*: Measured relation between the total RNA/protein mass ratio and the growth rate, in different growth media with different doses of a translation inhibitor (data from [33]). For each medium, indicated by a color, five different concentrations of inhibitor were used (higher dose leads to lower growth rate). Growth-medium compositions are given in the original publication and error bars represent standard deviations. The dashed black and colored lines are the same as in panel A, indicating the good quantitative correspondence between model predictions and experimental data for the chosen parameter values, obtained by fitting the model to the data points (see *Methods* for details).

### 2.2.3 Biomass maximization as an optimal control problem

A self-replicator at steady state accumulates biomass according to  $\text{Vol}(0) e^{\mu^* t}$ ,  $t \in [0, \tau]$ , when  $\mu^*$  is the growth rate at steady state. The accumulation of biomass is obviously maximal when the growth rate is maximal ( $\mu^* = \mu_{opt}^*$ ). In dynamical conditions, the growth rate is not constant and biomass accumulation is described more generally by:

$$\frac{d\text{Vol}}{dt} = \mu(t) \text{Vol}.$$

In other words, when integrating over the time interval  $[0, \tau]$ :

$$\ln \left( \frac{\text{Vol}(\tau)}{\text{Vol}(0)} \right) = \int_0^\tau \mu(t) dt. \quad (2.9)$$

Since the logarithm is an increasing function, maximizing the biomass produced over  $[0, \tau]$  requires maximization of the right-hand side of the equation.

In a changing environment, maximization of the integral in Eq. 2.9 will generally require the optimal value of  $\alpha$  to be a function of time instead of a specific constant value. This dynamical resource allocation problem can be formulated in a more precise way using concepts from optimal control theory [113]. Let  $J$  be the objective function

$$J(\alpha) = \int_0^\tau \mu(t) dt = \int_0^\tau \beta v_R(p, r) dt,$$

where  $\alpha : \mathbb{R}^+ \rightarrow [0, 1]$  is a time-dependent function. The time evolution of  $p$  and  $r$  is determined by the self-replicator model of Eqs 2.3 and 2.4, and  $p$  and  $r$  thus depend on  $e_M$  and  $\alpha$ . Moreover, let  $\mathcal{U} = \{\alpha : \mathbb{R}^+ \rightarrow [0, 1]\}$  be the set of admissible controls. The optimal dynamical control problem then consists in finding the time-varying function  $\alpha_{opt}(t)$  that maximizes  $J(\alpha)$  over the time-interval  $[0, \tau]$ :

$$\alpha_{opt} = \arg \max_{\alpha \in \mathcal{U}} J(\alpha). \quad (2.10)$$

In what follows, we will simplify the above problem by considering that the environment changes in a step-wise fashion at  $t = 0$ , but remains constant over the time-interval  $[0, \tau]$ , that is,  $e_M(t) = e_M$ . More specifically, we focus on the case of a nutrient upshift, corresponding to a step-wise increase of  $e_M$ . This upshift scenario corresponds to classical experiments in bacterial physiology [110–112], reviewed in [30], and is frequently encountered in the life cycle of a microorganism [3]. Notice that more complex environments can be approximated by a sequence of step-wise nutrient upshifts and downshifts.

### 2.2.4 Solution of the optimal control problem

Optimal dynamical control problems for two-dimensional nonlinear dynamical systems, like the problem of Eq. 2.10, are generally difficult to solve. However, we will show that the class of functions to which  $\alpha_{opt}$  belongs can be identified, and we will use numerical optimization to identify a particular  $\alpha_{opt}$  maximizing  $J$ .

As a preliminary step, in order to simplify the analysis, the variables in the self-replicator model of Eqs 2.3 and 2.4 are made nondimensional, by defining  $\hat{t} = k_R t$ ,  $\hat{p} = \beta p$ , and  $\hat{r} = \beta r$ . This leads to the following ODE system:

$$\frac{d\hat{p}}{d\hat{t}} = (1 - \hat{r}) E_M - (1 + \hat{p}) \hat{r} \frac{\hat{p}}{K + \hat{p}}, \quad (2.11)$$

$$\frac{d\hat{r}}{d\hat{t}} = \hat{r} \frac{\hat{p}}{K + \hat{p}} (\alpha(\hat{t}) - \hat{r}), \quad (2.12)$$

where  $K = \beta K_R$  and  $E_M = e_M/k_R$ . The nondimensional growth rate is given by:

$$\hat{\mu} = \frac{\mu}{k_R} = \frac{\hat{p}}{K + \hat{p}} \hat{r}. \quad (2.13)$$

Notice that the nondimensionalized system depends on a single parameter  $K$ , in addition to the constant environment  $E_M$ , which functions as an input to the system.

Analysis of the nondimensionalized system allows a number of properties of the solution of the optimal control problem of Eq. 2.10 to be derived (*Methods* and [S3 Text](#)). First, by applying a version of the well-known Pontryagin Maximum Principle [114], we can prove that the optimal solution is obtained for an alternating sequence of  $\alpha(\cdot) = 0$  and  $\alpha(\cdot) = 1$ , possibly ending with an intermediate value of  $\alpha(\cdot)$ , corresponding to the optimal steady state  $(\hat{p}(t), \hat{r}(t)) = (\hat{p}_{opt}^*, \hat{r}_{opt}^*)$ , that is, the steady state leading to the optimal growth rate  $\hat{\mu}_{opt}^*$  in the post-upshift environment  $E_M$ . Second, if the optimal solution reaches the optimal steady state for the new environment, then it does so after an infinite number of switches of  $\alpha(\cdot)$  between 0 and 1. Third, this switching behavior is characterized by a so-called switching curve  $\hat{r} = \varphi(\hat{p})$  in the  $(\hat{p}, \hat{r})$ -plane, which passes through  $(\hat{p}_{opt}^*, \hat{r}_{opt}^*)$ . The switching curve divides the phase plane into two regions, such that  $\alpha(\cdot)$  switches to 0 when the system is in the region above  $\varphi$  and to 1 when the system is below  $\varphi$  (black dashed curve in Fig. 2.4A).

In line with these results, we conjecture that the optimal solution consists in a switching transient towards the optimal steady state for the new environment, and remains at this steady state until the next environmental

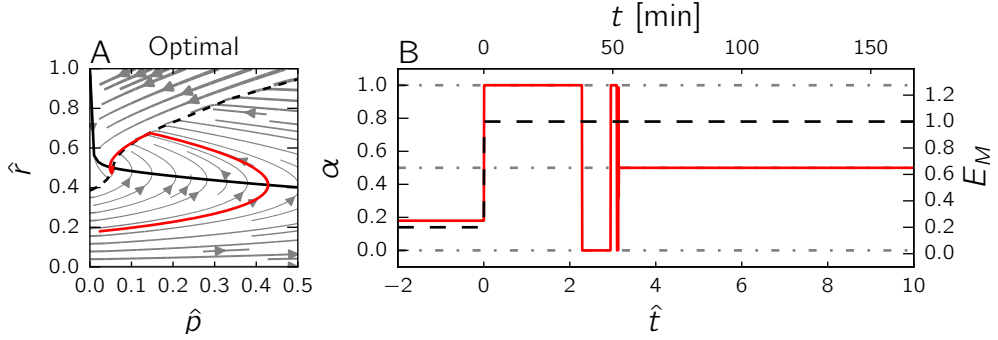


Figure 2.4 – **Optimal control of the self-replicator during a nutrient upshift.** *A*: Optimal trajectory in the phase plane for the nondimensionalized model of Eqs 2.11-2.12, with streamlines. The optimal trajectory is shown as a solid, red curve. The solid, black curve represents the  $\hat{p}$ -nullcline. The dashed, black curve is the switching curve  $\varphi(\hat{p})$ . The optimal solution was obtained by numerical optimization using `bocop` [115] (see *Methods* for details), using the parameter values  $E_M = 1$  and  $K = 0.003$ , and starting from the initial state  $(0.024, 0.18)$  at  $t = 0$  (optimal steady state for  $E_M = 0.2$ ). *B*: Time evolution of the control variable  $\alpha_{opt}(\cdot)$  (thick, red line) and the environment  $E_M$  (dashed, black line).

change. Such a solution is known as a bang-bang-singular solution in the control theory literature [113]. Formally, the solution of Eq. 2.10 can be described as

$$\alpha_{opt}(\hat{t}) = \begin{cases} 0, & \text{if } \hat{r}(\hat{t}) > \varphi(\hat{p}(\hat{t})), \\ 1, & \text{if } \hat{r}(\hat{t}) < \varphi(\hat{p}(\hat{t})), \\ \alpha_{opt}^*, & \text{if } (\hat{p}(\hat{t}), \hat{r}(\hat{t})) = (\hat{p}_{opt}^*, \hat{r}_{opt}^*). \end{cases} \quad (2.14)$$

Notice that the optimal solution involves dynamical feedback from the state of the system to the control variable  $\alpha(\cdot)$ , and is therefore an instance of closed-loop optimization [113].

The optimal control problem of Eq. 2.10 was also solved numerically by a direct method using the `bocop` software [115] (see *Methods* for details). A time discretization allows the problem to be transformed into a nonlinear optimization problem solved here by interior point techniques. The optimal trajectories obtained numerically confirm our conjecture that the optimal control is bang-bang-singular. An example solution, obtained by numerical optimization is shown in Fig. 2.4. At time  $\hat{t} = 0$ ,  $E_M$  jumps from a low to a high value, corresponding to a nutrient upshift (dashed black line in Fig. 2.4B). The optimal solution  $\alpha_{opt}$  consists of a sequence of switches

between  $\alpha = 1$ , corresponding to maximal synthesis of the gene expression machinery, and  $\alpha = 0$ , corresponding to maximal synthesis of the metabolic machinery, until  $(\hat{p}_{opt}^*, \hat{r}_{opt}^*)$  is reached.  $\alpha$  is then set to  $\alpha_{opt}^*$ , the value leading to the maximum growth rate in the new medium (here 0.5, for  $E_M = 1$ ). The sequence of switches of  $\alpha$  in Fig. 2.4B corresponds to successive crossings of the switching curve in Fig. 2.4A. In particular, the switch just after  $\hat{t} = 2$  corresponds to the first crossing of the switching curve; the subsequent switches accumulate around the steady state and are therefore difficult to identify in the plot.

What is the biological relevance of the bang-bang-singular solution maximizing growth of the bacterial self-replicator? In order to answer this question, we will investigate in the next two sections the different ways in which microorganisms could implement or have been shown to implement feedback growth control by sensing the environment and cellular physiology. Although the idealized solution proposed by optimal control theory will obviously not be found in nature, actual control strategies may produce solutions that are close. The optimal solution can thus be used as a gold standard, a benchmark for comparing actual control strategies.

### 2.2.5 Simple feedback control strategies: exploiting information on nutrients or precursors

The control strategies that microbial cells have evolved to bring resource allocation in line with changes in the environment involve a variety of molecular mechanisms [133]. These mechanisms are responsible for sensing the environment and the physiological state of the cell, as well as for adjusting the expression of genes that encode components of the transcriptional and translational machinery, enzymes, transporters, and proteins with other metabolic functions.

In the framework of the self-replicator model of bacterial growth, control strategies take the form of feedback control laws mapping the value of system variables to a value of the control variable  $\alpha(\cdot)$ . In this section, we explore two such strategies, the first exploiting information on the quality and quantity of substrate present in the environment, as reflected in the value of  $E_M$ , and the second using information on the precursor concentration  $\hat{p}$ . The feedback control strategies are graphically displayed in Fig. 2.5, as an extension of the self-replicator of Fig. 2.1. We pose a number of mathematical constraints on the feedback control strategies considered below. First, we require the control laws to be functions of the variables of the self-replicator but not involve derivatives or integrals of these variables. Second, for a constant

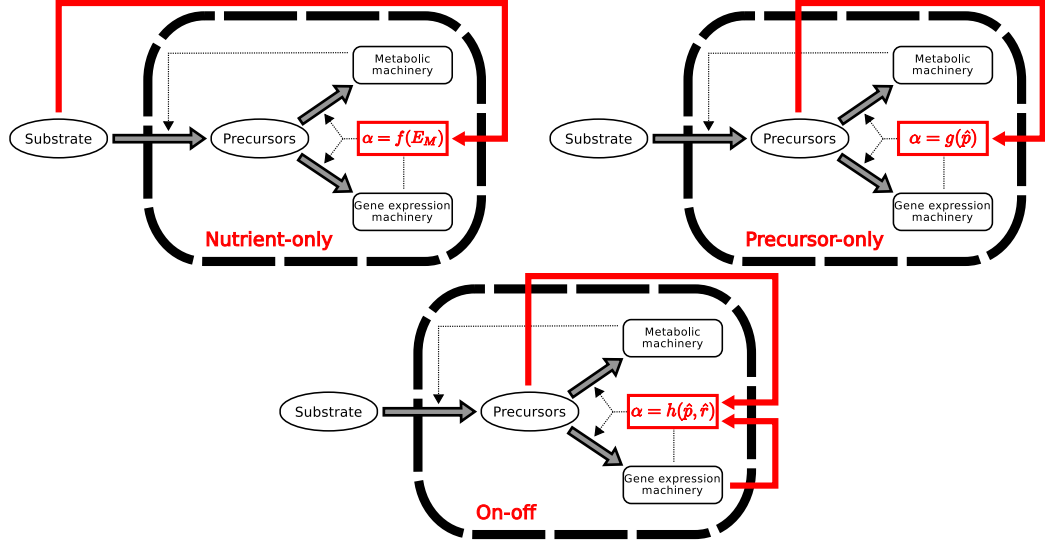


Figure 2.5 – **Alternative strategies for controlling the self-replicator of bacterial growth.** The feedback control strategies, shown in red and superposed on the self-replicator of Fig. 2.1, exploit information on system variables and the environment to adjust the value of  $\alpha$ , and thus the relative allocation of resources to the metabolic machinery and gene expression machinery.

environment  $E_M$ , the control strategies must drive the system to a unique stable and non-trivial steady state, enabling a non-zero growth rate. Third, this steady state must equal the optimal steady state for that environment, given by  $(\hat{p}_{opt}^*, \hat{r}_{opt}^*)$ .

The first control strategy is defined by the function  $f: \mathbb{R}^+ \rightarrow [0, 1]$ , mapping  $E_M$  to  $\alpha$ :

$$\alpha = f(E_M). \quad (2.15)$$

Notice that  $\alpha$  is constant because  $E_M$  is fixed to the value defining the new environment after the upshift. What would be an appropriate choice for  $f$ ? An advantage of the self-replicator model is that the optimal allocation at steady state can be explicitly formulated as a function of  $E_M$  (Eq. 2.22 in *Methods*, with derivation in *S1 Text*). This function is the unique function satisfying all of the above criteria (*S1 Text*). *S1 Figure* plots  $f$  and shows that it is conveniently approximated by a Michaelis-Menten function, *i.e.*,

$$\alpha(\cdot) = \frac{E_M}{E_M + K_{mE}}, \quad (2.16)$$

with the dimensionless half-saturation constant  $K_{mE}$ . The interest of the

approximation is that it demonstrates that the control strategy can be described by a simple and ubiquitous response curve in biochemical kinetics.

As an example of a regulatory system resembling the above control strategy consider the phosphotransferase system responsible for the uptake of glucose, the preferred substrate of *E. coli* [134]. In the presence of glucose, the EIIA<sup>Glc</sup> component of the phosphotransferase system is mostly unphosphorylated, since the phosphate groups are used for the conversion of extracellular glucose to intracellular glucose-6-phosphate. When glucose disappears from the medium, however, the glucose uptake rate decreases and, correspondingly, the phosphorylated fraction of EIIA<sup>Glc</sup> increases. The phosphorylation state of EIIA<sup>Glc</sup> thus provides an indirect read-out of glucose availability. In response to this signal, a variety of metabolic processes are upregulated or downregulated, notably involving the signalling molecule cAMP which activates the pleiotropic transcription factor Crp [134, 135].

How does the control strategy of Eq. 2.15, which we call a nutrient-only strategy, perform in comparison with the optimal solution derived in the previous section? That is, how much biomass does this strategy produce compared with the maximal amount of biomass that can theoretically be obtained after a nutrient upshift? In order to answer these questions, we simulated the response to a sudden upshift of the self-replicator of Eqs 2.11-2.12 controlled by the nutrient-only strategy of Eq. 2.15. The results are shown in Fig. 2.6. Panel A shows the trajectory of the controlled self-replicator system and panel D plots the evolution of the amount of biomass as a fraction of the amount of biomass produced by the optimal strategy. While the system does reach the steady state that is optimal for  $E_M$ , the nutrient-only strategy has poor performance in the transient phase immediately following the nutrient upshift. As can be seen from the solution trajectory in Fig. 2.6A, fixing  $\alpha$  to the value that enables optimal growth at steady state leads to a huge transient overshoot of the precursor concentration. The overshoot reveals that resource allocation is initially suboptimal, with too many resources invested in the metabolic machinery at the expense of the gene expression machinery. This causes a transiently suboptimal growth rate, leading to lower biomass accumulation (Eq. 2.9).

One way to avoid the transient precursor imbalance observed in Fig. 2.6A would be to exploit information on the precursor concentration in the control strategy. The second strategy considered here, which we label a precursor-only strategy, does exactly this: it involves a feedback control law  $g: \mathbb{R}^+ \rightarrow [0, 1]$  mapping  $\hat{p}$  to  $\alpha$ :

$$\alpha = g(\hat{p}). \quad (2.17)$$



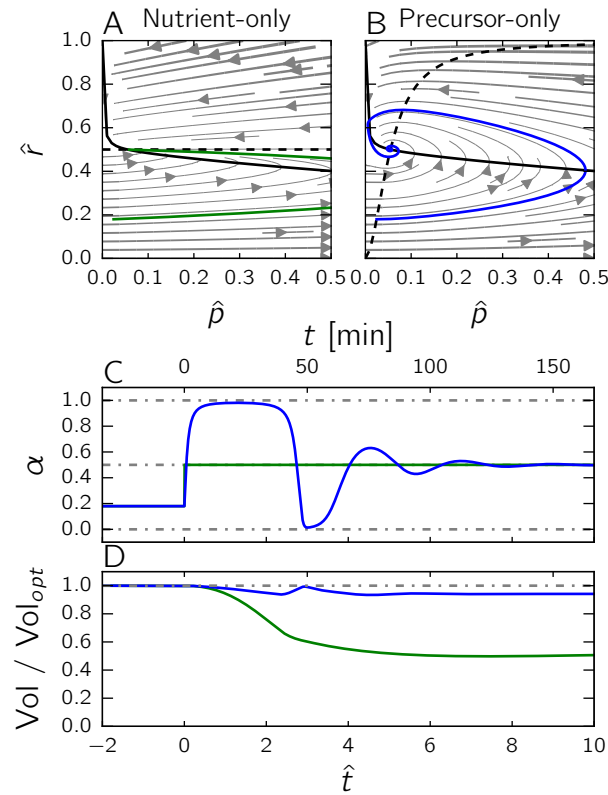


Figure 2.6 – **Comparison of the performance of the nutrient-only and precursor-only strategies after a nutrient upshift.** *A*: Trajectory in the phase plane for the nutrient-only strategy (green curve). The solid, black curve represents the  $\hat{p}$ -nullcline. The dashed, black curve is the  $\hat{r}$ -nullcline. The solution is obtained by numerical simulation of the system of Eqs 2.11-2.12, supplemented with  $\alpha = f(E_M)$  as specified by Eq. 2.27 in the *Methods* section and plotted in S1 Figure. The initial state corresponds to the steady state attained for an environment given by  $0.2 E_M$ . While converging to the new steady state after the upshift, the precursor concentration makes a large overshoot. *B*: As above, but for the precursor-only strategy. The feedback control strategy is now defined by  $\alpha = g(\hat{p})$  as specified by Eq. 2.28 in the *Methods* section and plotted in S1 Figure. The solution trajectory (blue curve) exhibits a lower overshoot. *C*: Evolution of the control variable  $\alpha(\cdot)$  as a function of time, for each of the above two strategies. Notice that in the nutrient-only strategy  $\alpha(\cdot)$  immediately jumps to the optimal value for the post-upshift steady state (green curve), whereas in the precursor-only strategy it depends on the (time-varying) precursor concentration (blue curve). *D*: Evolution of the ratio  $\text{Vol}/\text{Vol}_{opt}$  as a function of time, where  $\text{Vol}$  is the volume of the self-replicator and  $\text{Vol}_{opt}$  the volume of the same replicator following the optimal strategy shown in Fig. 2.4. In all of the above simulations, the parameter values  $E_M = 1$  and  $K = 0.003$  were used.

Since  $\hat{p}$  will vary during the upshift experiment,  $\alpha$  is not constant, contrary to the nutrient-only strategy above. In the *Methods* section, we present a function  $g$  satisfying the requirements listed in the beginning of this section, in particular that the system converge to a stable steady state ensuring maximal growth in the new environment. Moreover, we show that any other choice for  $g$  leads to lower biomass production. The function is plotted in [S1 Figure](#), and as shown in the same panel, is conveniently approximated by a Hill function with cooperativity coefficient 2:

$$\alpha(\cdot) = \frac{\hat{p}^2}{\hat{p}^2 + K_{mp}^2}, \quad (2.18)$$

where  $K_{mp}$  is a dimensionless half-saturation constant.

While converging to the same steady state, this second strategy, which we will refer to as the precursor-only strategy, performs much better than the nutrient-only strategy after an upshift, as shown in [Fig. 2.6](#). We simulated the response to a nutrient upshift of the self-replicator of [Eqs 2.11-2.12](#) with the precursor-only strategy of [Eq. 2.17](#). The relative biomass increases by 51% and reaches 94% of the biomass produced by the optimal control strategy (the theoretical maximum). The precursor-only strategy notably avoids the inefficient transient accumulation of precursors directly after the nutrient upshift, by alternately investing more resources in gene expression (consumption of precursors) and metabolism (production of precursors). In this respect, the oscillatory time profile of  $\alpha$  ([Fig. 2.6C](#)) is somewhat reminiscent of the bang-bang-singular control in the solution of the optimal control problem ([Fig. 2.4B](#)).

Both strategies, nutrient-only and precursor-only, drive the self-replicator towards the same steady state. Whereas the two strategies are thus indistinguishable when the analysis is restricted to steady state, the precursor-only strategy is shown to perform much better in a dynamical upshift scenario, in the sense that the biomass produced is much closer to that produced by the optimal strategy. Several authors have concluded that control strategies based on precursor sensing are key for maintaining optimal growth at steady state. *Scott et al.* argue that a strategy similar to the precursor-only approach above allows robust control of amino acid supply and demand, resulting in optimal steady-state growth over a range of nutrient conditions [[32](#)]. They associate this strategy with ppGpp-mediated control of the synthesis of ribosomal proteins [[126–128](#)]. The signalling molecule ppGpp accumulates in response to an increase in the level of uncharged tRNA, when amino acid concentrations in the cell drop. This causes ribosomes to "stall" and leads to RelA-mediated conversion of GTP to ppGpp, the molecular details

of which are still subject of debate [128, 136]. Since ppGpp inhibits the transcription of ribosomal RNAs [137], the concentration of the latter decreases, leading to more inactive ribosomal proteins and, through a well-characterized post-transcriptional autoregulatory mechanism, a lower synthesis rate of ribosomal proteins [127, 138]. Our analysis adds to the above study a novel insight: measuring precursors does not only enable resource allocation control to achieve maximal growth at steady state, but is also a good strategy in a dynamical context.

While the precursor-only strategy is thus seen to lead to good results, Fig. 2.6D shows that there remains room for improvement. It seems reasonable to expect that control strategies exploiting information of not just a single variable, but several variables simultaneously, could further improve the performance of the self-replicator during a growth transition.

### 2.2.6 A near-optimal feedback control strategy that exploits information on the imbalance between precursors and the gene expression machinery

In the quest for further improvements, a natural starting-point would be to consider the curve defining the optimal steady states  $(\hat{p}_{opt}^*, \hat{r}_{opt}^*)$  for different environments  $E_M$ . This curve is defined by a function mapping  $\hat{p}^*$  to  $\hat{r}^*$ , which is actually the same as the function  $g$  introduced in the precursor-only strategy (*Methods* and *S1 Figure*), given that at steady state  $\hat{r} = \alpha$  (Eq. 2.12). The curve can be seen as representing an optimal balance between precursors and the gene expression machinery, in the sense that the maximal growth rate attainable for a given precursor concentration  $\hat{p}$  requires a concentration  $\hat{r}$  of ribosomes and other components of the gene expression machinery equal to  $g(\hat{p})$ . If either  $\hat{r} > g(\hat{p})$  or  $\hat{r} < g(\hat{p})$ , the growth rate is suboptimal.

These considerations suggest an intuitive control strategy, namely to avoid an imbalance between  $\hat{p}$  and  $\hat{r}$  at all times, and remain as close as possible to the curve defined by  $g$ . In particular, when the gene expression machinery is more abundant than what is optimal given the available precursors ( $\hat{r} > g(\hat{p})$ ), its synthesis is switched off ( $\alpha = 0$ ). Conversely, when  $\hat{r} < g(\hat{p})$ , synthesis of the gene expression machinery is switched on. This strategy thus tries to restore "as quickly as possible" the optimal balance between precursors  $\hat{p}$  and the gene expression machinery  $\hat{r}$ , giving rise to a

so-called on-off control strategy:

$$\alpha = h(\hat{p}, \hat{r}) = \begin{cases} 0, & \text{if } \hat{r} > g(\hat{p}), \\ 1, & \text{if } \hat{r} < g(\hat{p}), \\ \alpha_{opt}^* & \text{if } (\hat{p}, \hat{r}) = (\hat{p}_{opt}^*, \hat{r}_{opt}^*). \end{cases} \quad (2.19)$$

As shown in the *Methods* section, the on-off strategy drives the system to a stable steady state ensuring growth at the maximal rate. Notice that, contrary to the strategies discussed in the previous section, the value of  $\alpha$  selected by the on-off strategy depends on both  $\hat{p}$  and  $\hat{r}$  (Fig. 2.5). It thus uses more information on the state of the system than the nutrient-only and precursor-only strategies.

Fig. 2.7 shows the performance of the on-off strategy after a nutrient upshift, as compared to the precursor-only strategy. The transition is seen to be nearly perfect, in the sense that 98% of the optimal biomass is produced by the strategy. The time course of  $\alpha$  in panel *D* is very similar to the optimal time course obtained by numerical optimization, shown in Fig. 2.4*B*, and clearly brings out the bang-bang-singular nature of the solution. These results show that a strategy exploiting complete information on the internal state of the self-replicator can lead to near-optimal performance, outcompeting a strategy that uses partial information on the internal state (precursor abundance only).

Are microbial cells equipped with mechanisms implementing a strategy similar to the on-off strategy? A possible candidate would again be the ppGpp system. A kinetic model of ppGpp metabolism and the regulation of the synthesis of ribosomal proteins was recently presented by Bosdriesz *et al.* [116]. The model proved capable of accounting for a range of experimental data, including the steady-state concentration of ppGpp as a function of the growth rate [29] and the dynamical response of ppGpp to a nutrient upshift or downshift [139]. A major conclusion of the model is that the steady-state concentration of ppGpp exhibits a strongly ultrasensitive response to deviations of the ribosomal protein fraction from the optimal ribosomal protein fraction at a given growth rate. These deviations from optimality, in turn, lead to a switch-like response of the synthesis rate of ribosomal proteins (Fig. 4 in Bosdriesz *et al.* [116]).

How does this mechanistic model of ppGpp regulation relate to the on-off strategy presented above? In order to answer this question, we first need to find a correspondence between the variables  $p$  and  $r$  of our coarse-grained model and the concentrations of molecular species in the kinetic model of Bosdriesz *et al.* This is rather straightforward to achieve, by equating  $p$  to the total amino acid concentration and  $r$  to the ribosome concentration.

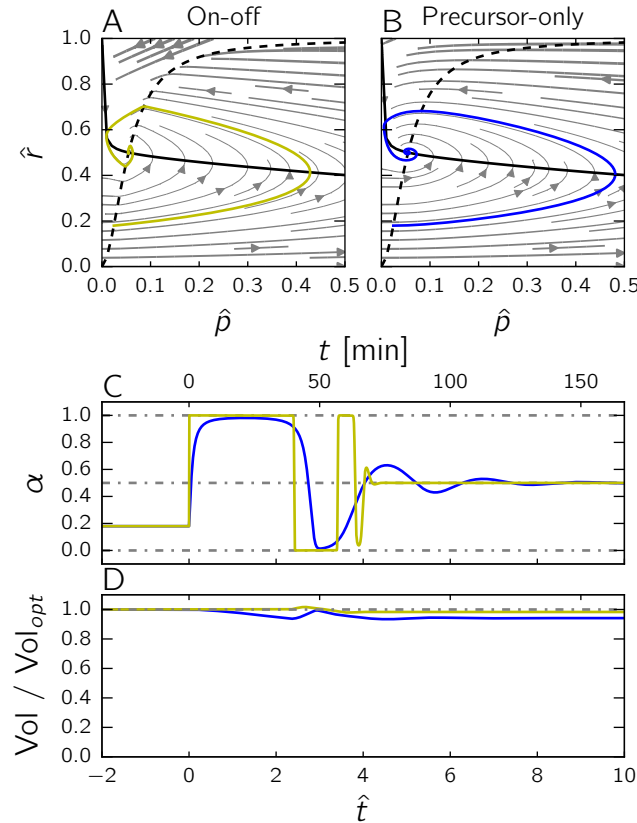


Figure 2.7 – **Comparison of the performance of the precursor-only and the on-off strategies after a nutrient upshift.** *A*: Trajectory in the phase plane for the on-off strategy (yellow curve). The solid, black curve represents the  $\hat{p}$ -nullcline and the dashed, black curve the function  $g$ . The solution is obtained by numerical simulation of the system of Eqs 2.11-2.12, supplemented with the equation  $\alpha = h(\hat{p}, \hat{r})$  defined in Eq. 2.19 and plotted in Fig. 2.8A. The initial state corresponds to the optimal steady state attained for an environment given by  $0.2 E_M$ . *B*: Trajectory in the phase plane for the precursor-only strategy (same as in Fig. 2.6B, added for comparison). *C*: Evolution of the control variable  $\alpha$  for each strategy as a function of time. Both strategies stabilize the system at the optimal steady state, but only the on-off strategy (yellow curve) exhibits bang-bang behavior. *D*: Evolution of the ratio  $\text{Vol}/\text{Vol}_{opt}$  for the on-off and precursor-only strategies as a function of time, where  $\text{Vol}$  is the volume of the self-replicator and  $\text{Vol}_{opt}$  the volume of the same replicator following the optimal strategy shown in Fig. 2.4. The final values of  $\text{Vol}/\text{Vol}_{opt}$  attained by the two strategies are 0.9831 and 0.9413, respectively. The on-off strategy is thus hardly distinguishable from the optimal control strategy in the plot. In all of the above simulations, the parameter values  $E_M = 1$  and  $K = 0.003$  were used.

Second, [S4 Text](#) shows that by making two simplifying assumptions, ppGpp can be expressed as a function of the total amino acid concentration and the ribosome concentration. In particular, we assume that concentrations of all individual amino acids are equal, and that the concentrations of charged tRNAs and ppGpp evolve fast relative to the dynamics of the amino acid and ribosome concentrations. The third step consists in positing an explicit relation between ppGpp and  $\alpha$ , based on the regulatory action of ppGpp on the transcription of ribosomal RNA [\[137\]](#):

$$\alpha(\cdot) = \frac{K_I}{K_I + \text{ppGpp}(\cdot)}, \quad (2.20)$$

with  $K_I$  a Michaelis-Menten inhibition constant [ $\mu\text{mol L}^{-1}$ ] and ppGpp the (time-varying) intracellular concentration of ppGpp [ $\mu\text{mol L}^{-1}$ ].

The response function for ppGpp thus obtained and evaluated for a range of amino acid and ribosome concentrations is represented in [Fig. 2.8](#), and visually compared with the on-off strategy. As can be seen, the two response surfaces are very similar. In other words, the ultrasensitive response of the synthesis rate of ribosomal proteins to the suboptimal allocation of cellular resources, derived from a model of the molecular mechanisms involved in the synthesis, degradation, and regulatory action of ppGpp [\[116\]](#), implements a control strategy that is close to the optimal predicted by a control-theoretical analysis of the self-replicator. While the role of ppGpp in maintaining optimal resource allocation was already pointed out by Scott *et al.* and Bosdriesz *et al.*, the latter studies were restricted to optimizing steady-state growth. A major insight from the analysis in this section is that this conclusion seems to carry over to dynamical scenarios as well. Fundamentally, the analysis suggests that the ppGpp system is a likely candidate to fulfill this role because it integrates information on the imbalance between precursor concentration and abundance of the gene expression machinery.

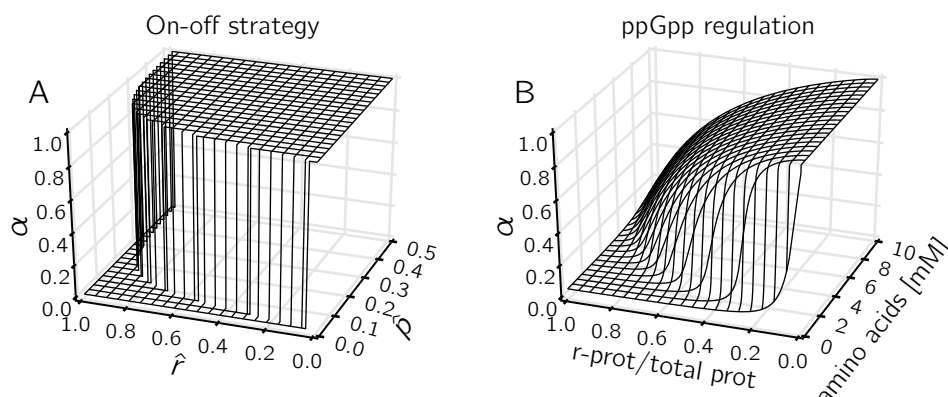


Figure 2.8 – **ppGpp regulation implements an on-off control strategy of resource allocation.** *A*: Response surface of the on-off control strategy, defined by  $\alpha = h(\hat{p}, \hat{r})$  in Eq. 2.19. *B*: Response surface of the ppGpp control strategy, as defined by Eq. 2.20 and the simplified kinetic model defining ppGpp in terms of the total amino acid concentration and the ribosomal protein fraction (S4 Text). The shape of the response surface of the ppGpp control strategy is seen to be in very good agreement with the on-off strategy leading to near-optimal performance of the self-replicator during a nutrient upshift.

## 2.3 Discussion

Quantitative growth laws are empirical regularities pointing at fundamental properties of microbial life [31]. Recent work has led to the precise theoretical formulation of growth laws and has shown that they can be derived from basic assumptions on the molecular processes responsible for the assimilation of nutrients and their conversion to biomass [24, 32, 116, 121, 122]. The growth laws are uniquely defined under the hypothesis that microorganisms allocate resources in such a way as to maximize their growth rate. Several of the above-mentioned studies have analyzed feedback control strategies on the molecular level enabling cells to achieve optimal resource allocation in a robust manner. The control strategies exploit information on the physiological state of the cell to adjust the (relative) rate of synthesis of different classes of proteins (ribosomes, metabolic enzymes, ...). Whereas the growth laws describe microbial growth at steady state, most microorganisms live in complex, continuously changing environments. Despite some precursory work [123, 124], questions about the dynamics of microbial growth remain largely unanswered: Which resource allocation schemes are optimal in changing environments? Which dynamical control strategies lead to (near-)optimal



resource allocation? How do these strategies compare with those actually implemented by microorganisms?

We have addressed the above questions by means of a self-replicator model of microbial growth, which, like other coarse-grained models of bacterial growth [24, 32, 121], is capable of reproducing the growth laws at steady state (Fig. 2.3). A first major contribution of our work is to show that, in the case of a dynamical upshift scenario, optimal production of biomass requires a bang-bang-singular resource allocation scheme (Fig. 2.4). That is, the optimal self-replicator should iteratively allocate all of its resources to the gene expression machinery (bang control input) and the metabolic machinery (another bang control input), until the steady state enabling maximal growth in the post-upshift environment is reached, corresponding to a trade-off in the allocation of resources to the two processes (singular control input).

Bang-bang phenomena are widespread in a variety of life processes. Applications of optimal control theory to reproductive strategies in insects [140], the development of intestinal crypts [141], and the activation of metabolic pathways [142, 143] have led to bang-bang or bang-bang-singular strategies. In optimal control problems, such a solution arises with systems where the differential equations are linear in the control variable (in our case,  $\alpha(\cdot)$ ). Examples of applications that are close to the problem considered here are the control of gene expression for adaptation to environmental changes [109, 123], and the allocation of resources between nutrient uptake and growth in microorganisms [124, 144]. Whereas the former applications focus on minimization of response times, the latter also optimize biomass during a growth transition, using a different model, not derived from first principles as in this study. However, the optimal solution of the corresponding optimal control problem is also bang-bang-singular, thus showing that our conclusions are robust to model variations.

Our second major contribution is the assessment of how different feedback control strategies perform with respect to each other and to the gold standard determined from optimal control theory. We show that the precursor-only and nutrient-only strategies, both of which drive the self-replicator to the steady state with maximal growth rate in a static environment, perform quite differently in a dynamical upshift scenario (Fig. 2.6). While the precursor-only strategy is better than the nutrient-only strategy in a dynamical environment, it is in turn outperformed by a so-called on-off strategy, which achieves a near-perfect growth transition by exploiting information on the imbalance between the precursor concentration and the abundance of the gene expression machinery (Fig. 2.7). The superior performance of the on-off strategy can be intuitively explained by the fact that during a growth transition the two variables are not fully correlated, which means that sensing



both instead of either one provides additional information in a dynamical context.

Interestingly, the on-off strategy is based on a feedback control law that very much resembles the response function for ppGpp-mediated regulation of the synthesis of ribosomal RNAs in *E. coli* [116]. The role of ppGpp in controlling microbial growth has been amply documented [126–128]. For example, Potrykus *et al.* observed that in cells without ppGpp (ppGpp<sup>0</sup> mutants) the RNA/protein mass ratio, a proxy for our resource allocation variable  $\alpha$ , does not change with the growth rate, which has led these authors to conclude that ppGpp is the major source of growth-rate control in *E. coli* [145]. The central importance of ppGpp in the reallocation of gene expression resources in *E. coli* following changes in nutrient availability has also been mapped with higher resolution, using genome-wide transcriptome studies [146, 147]. In nearly all bacterial species examined so far, ppGpp is known to accumulate in response to an increase in the level of uncharged tRNA [148], although the molecular details of ppGpp metabolism and the range of other functions of the alarmone may greatly vary across species [128, 148, 149]. While it has thus been well-established that regulation by ppGpp is an evolutionary conserved mechanism of growth control in the bacterial cell, our analysis provides a new perspective by suggesting that ppGpp enables optimal reallocation of resources after a growth transition, dynamically maximizing the accumulation of biomass.

The model on which the above results are based is built from first principles by distinguishing two fundamental cellular processes: metabolism (converting nutrients to precursors) and gene expression (converting precursors to the proteins that make up biomass) (Fig. 2.1). Despite its simplicity, our self-replicator model is capable of reproducing the empirical growth laws and of making testable predictions on the time-course profile of the resource allocation variable  $\alpha$  and on the concentrations  $p$  and  $r$  of components of the gene expression machinery and metabolic machinery, respectively (see Fig. 2.8 and below). The model can be easily extended with more details on protein synthesis, central carbon and energy metabolism, stress systems, or cell membranes, but this would make the mathematical analysis of the model dynamics and the optimal control problem more complicated. Notice, however, that the direct numerical approach for solving the optimal control problem remains applicable, even for more fine-grained models (Fig. 2.4, see also [125]).

The comparison of different control strategies during a classical growth transition should be interpreted carefully, in a qualitative rather than quantitative manner. Whereas the differences in performance based on the biomass ratio  $\text{Vol}/\text{Vol}_{opt}$  of the control strategies are robust, the absolute numbers for

the biomass ratio will depend on details of the growth experiment chosen and the exact parameter values. Another implicit assumption in the analysis of the control strategies is that the costs of their molecular implementation can be neglected. This is not true in general, since every control strategy requires resources to be diverted towards the synthesis of sensory systems and regulatory proteins, with possibly detrimental effects on growth. In other words, a control strategy entails a trade-off between the growth burden of regulation and the growth benefit of the improved capability to adapt to changes in the environment [150, 151]. The analysis of control strategies could be refined by adding a reaction to the self-replicator that models the loss of resources incurred by regulatory strategies. While in general the growth burden of a control strategy requiring information on several aspects of cellular physiology is expected to be higher, notice that a single regulatory system may be capable of sensing more than one variable. For example, we show that ppGpp levels in the cell carry information on both the metabolic and the gene expression state (Fig. 2.8), thus integrating several signals in a cost-efficient manner.

The model predictions for the dynamical adaptation of resource allocation after a nutrient upshift suggest several interesting experimental tests. In particular, the switching profile of the resource allocation variable  $\alpha$  is a promising candidate for experimental validation. The most straightforward option would be direct measurement of the synthesis rate of ribosomal proteins, using a translational fusion of a fluorescent reporter with a ribosomal protein [81, 136]. However, a more indirect approach based on the quantification of ppGpp concentrations in the cell or the activity of the ribosomal RNA (rRNA) promoters would also be a possibility. Interestingly, some data are already available in the literature. For instance, Gausing has reviewed data on the synthesis of ribosomal proteins after a nutrient upshift, showing that the synthesis rate goes through "a series of rapid changes" resembling oscillations [152]. Later work attributed this pattern to regulation on the transcriptional level [153]. Friesen *et al.* observed oscillatory patterns in ppGpp concentrations after a nutrient upshift, with an initial response resembling bang control for an upshift to a particularly rich medium [154]. Murray *et al.* also present data on the ppGpp concentration after a nutrient upshift [139], but with a lower temporal resolution and no clear oscillatory pattern. All of the above measurements were carried out on the population level, which means that switching patterns may be obscured by desynchronisation of the individual cells. More sophisticated experimental set-ups are necessary for the decisive validation of the model predictions, allowing gene expression in single cells to be followed over time in tightly regulated growth conditions [155, 156]. In addition, the model could be validated on other

dynamical scenarios, for example nutrient downshifts [139, 157].

Apart from its interest for fundamental science, resource allocation is also a critical question in biotechnology, where there exists an inherent trade-off between the maximization of yield and productivity [158]. High yield means that most of the substrate is converted to a metabolite, peptide or recombinant protein of interest, but this leads to low productivity if the remaining nutrient influx is insufficient to sustain population growth. Engineered control of resource allocation may help in establishing the right trade-off, the most profitable balance between yield and productivity, in a biotechnological process. Such a trade-off could be attained either in steady-state conditions (the incoming nutrient flux is optimally distributed over growth and production) or in dynamical conditions (alternating utilization of the incoming nutrient flux for growth or production) [159–161]. When extended with heterologous metabolic pathways, the self-replicator models used in this study would provide an adequate *in-silico* test bed for the rapid screening and comparison of alternative control strategies in bioprocess engineering.

## 2.4 Methods

### 2.4.1 Steady-state analysis of model

The nondimensional version of the model, given by Eqs 2.11-2.12, was used for a steady-state analysis of the self-replicator. Eqs 2.11-2.12 were derived from the original model of Eqs 2.3-2.4 by means of the following rescalings:

$$\hat{p} = \beta p, \quad \hat{r} = \beta r, \quad \hat{t} = k_R t, \quad E_M = e_M/k_R, \quad K = \beta K_R.$$

As shown in S1 Text, for a constant environment  $E_M$  and constant resource allocation  $\alpha$ , the system has two steady states: a trivial unstable steady state  $(\hat{p}^*, \hat{r}^*) = (0, 1)$ , allowing no growth in the absence of precursors, and a steady state with a positive growth rate given by

$$(\hat{p}^*, \hat{r}^*) = \left( \frac{(1 - \alpha) E_M - \alpha + \sqrt{[(1 - \alpha) E_M - \alpha]^2 + 4\alpha(1 - \alpha) E_M K}}{2\alpha}, \alpha \right). \quad (2.21)$$

The two eigenvalues of the Jacobian matrix evaluated at  $(\hat{p}^*, \hat{r}^*)$  are negative (S1 Text), so that this steady state is stable.

The growth rate at steady state, as a function of  $\hat{p}^*$  and  $\hat{r}^*$ , is given by Eq. 2.13, which we repeat here for clarity:

$$\hat{\mu}^* = \frac{\hat{p}^*}{K + \hat{p}^*} \hat{r}^*.$$

Evaluating  $d\hat{p}/dt = 0$  at  $(\hat{p}^*, \hat{r}^*)$  allows  $\hat{r}^*$ , and therefore  $\hat{\mu}^*$ , to be written as a function of  $\hat{p}^*$  (S1 Text). Accordingly, we can compute  $\partial\hat{\mu}^*/\partial\hat{p}^*$  and, when setting this partial derivative to 0, determine the maximum growth rate at steady state  $\mu_{opt}^*$  and the optimal resource allocation  $\alpha_{opt}^*$  bringing about this maximal growth rate. As shown in S1 Text,  $\mu_{opt}^*$  and  $\alpha_{opt}^*$  can be written as explicit functions of either the environment  $E_M$ :

$$\alpha_{opt}^* = \frac{E_M + \sqrt{K E_M}}{E_M + 2\sqrt{K E_M} + 1}, \quad \hat{\mu}_{opt}^* = \frac{E_M}{E_M + 2\sqrt{K E_M} + 1}, \quad (2.22)$$

or the precursor abundance  $\hat{p}_{opt}^*$ :

$$\alpha_{opt}^* = \frac{\hat{p}_{opt}^*}{\hat{p}_{opt}^* + \frac{K}{K + \hat{p}_{opt}^*}(1 + \hat{p}_{opt}^*)}, \quad \hat{\mu}_{opt}^* = \frac{\hat{p}_{opt}^{*2}}{\hat{p}_{opt}^{*2} + 2K\hat{p}_{opt}^* + K}. \quad (2.23)$$

The above equations were used for the derivation of the control strategies (see below).

### 2.4.2 Model fitting

As can be seen by comparing Figs 2.3A and 2.3B, growth-rate maximization in the self-replicator model leads to a good qualitative correspondence with the growth laws. In order to determine if a good quantitative fit of the model with the data from Scott *et al.* [33] can be obtained, for reasonable parameter values, we estimated  $e_M$  and  $k_R$  in Eqs 2.3-2.4 from the measured RNA/protein mass ratios. At steady state, the RNA/protein mass ratio can be interpreted as proportional to  $\hat{r}^*$  (and thus  $\alpha_{opt}^*$ ), with an unknown (dimensionless) proportionality constant  $\gamma$  (see [33] for details on the use of the RNA/protein mass ratio as a proxy for the ribosomal protein mass fraction):

$$\hat{r}^* = \alpha_{opt}^* = \gamma \frac{\text{RNA mass}}{\text{protein mass}}. \quad (2.24)$$

Reformulating Eq. 2.22 in terms of the original parameters  $e_M$  and  $k_R$ , which have physical dimensions facilitating the biological interpretation of their values, we obtain a straightforward relation between  $e_M$ ,  $k_R$ ,  $K$ ,  $\alpha_{opt}^*$  and  $\mu_{opt}^*$ :

$$\alpha_{opt}^* = \frac{e_M + \sqrt{K e_M k_R}}{e_M + 2\sqrt{K e_M k_R} + k_R}, \quad \mu_{opt}^* = \frac{e_M k_R}{e_M + 2\sqrt{K e_M k_R} + k_R}. \quad (2.25)$$

Eqs 2.24-2.25 were used to estimate values of  $k_R$  and  $\gamma$ , as well as  $e_M$  for each of the six growth conditions, from the measurements of the growth rate

and the RNA/protein mass ratio. The value  $K$  was not estimated from the experimental data, but set to a value inferred from the literature (S1 Text). The optimization process was carried out by means of the differential evolution algorithm of Storn and Price [162]. The results are shown in Fig. 2.3B, while the estimated parameter values are summarized in S1 Table. The parameter values are in very good agreement with order-of-magnitude values determined from the literature (S2 Text and S1 Table).

### 2.4.3 Solution of optimal control problem

The optimal control problem of Eq. 2.10 consists in identifying the function  $\alpha_{opt}(t)$  that maximizes the integral of the growth rate  $\hat{\mu}$  over an interval  $[0, \tau]$ . In order to solve this problem, we first redefined it over an infinite horizon (*i.e.*,  $\tau \rightarrow \infty$ ) in order to avoid boundary effects occurring over finite time intervals, in particular the depletion of precursors just before reaching  $\tau$ . With  $\mathcal{U} = \{\alpha : \mathbb{R}^+ \rightarrow [0, 1]\}$  the set of admissible controls, the full optimization problem for the nondimensionalized system is given by

$$\max_{\alpha \in \mathcal{U}} J(\alpha) \equiv \int_0^{\infty} \hat{r}(\hat{t}) \frac{\hat{p}(\hat{t})}{K + \hat{p}(\hat{t})} d\hat{t}. \quad (2.26)$$

Since  $J(\alpha)$  diverges, we actually consider overtaking optimality: A solution is overtaking optimal if its performance index catches up with the performance index of any other solution ([114], see S3 Text for details).

Necessary conditions on optimal trajectories can be obtained by the Infinite Horizon Maximum Principle [114], an extension of the well-known Pontryagin Maximum Principle. Analysis of the Hamiltonian of the system of Eqs 2.11-2.12 and the associated adjoint system shows that the optimal trajectory is a concatenation of bang arcs ( $\alpha(\cdot) = 0$  or  $\alpha(\cdot) = 1$ ) and possibly a singular arc corresponding to the optimal steady state  $(\hat{p}(t), \hat{r}(t)) = (\hat{p}_{opt}^*, \hat{r}_{opt}^*)$ , that is, the steady state leading to the optimal growth rate  $\hat{\mu}_{opt}^*$  in the new environment after the upshift (S3 Text). Moreover, from the Kelley condition [163], we can show that if the optimal trajectory has a singular arc, then it must enter this singular arc *via* a chattering arc, *i.e.*, with an infinite number of switches of  $\alpha(\cdot)$  between 0 and 1 (S3 Text). The chattering arc is characterized by a switching curve  $\hat{r} = \varphi(\hat{p})$  in the  $(\hat{p}, \hat{r})$ -plane, which passes through  $(\hat{p}_{opt}^*, \hat{r}_{opt}^*)$ . The switching curve divides the phase plane into two regions, such that  $\alpha(t)$  switches to 0 when the system is in the region above  $\varphi$  and to 1 when the system is below  $\varphi$  (S3 Text and Fig. 2.4).

The above results have led to the conjectured optimal solution of Eq. 2.14. In parallel, we numerically solved the problem of Eq. 2.26 by a direct method

using the bocop software [115]. A time discretization allows the optimal control problem to be transformed into a nonlinear optimization problem, solved here by interior point techniques. A discretization by a Lobatto IIIc formula (6th order) was used with 4000 time steps, and the relative tolerance for the NLP solver was set to  $10^{-14}$ . The optimal trajectories thus obtained are composed of a chattering arc followed by a steady state corresponding to the singular arc (Fig. 2.4). The switching curve  $\varphi(\hat{p})$  was computed from numerical simulations with different initial conditions.

#### 2.4.4 Specification and analysis of control strategies

As described in the *Results* section, we are interested in control strategies satisfying the following conditions:

- (C1) The control laws are static functions of the system variables (as opposed to, for instance, functions that depend on derivatives or integrals of the variables).
- (C2) For any given constant environment  $E_M$ , they drive the self-replicator system towards a unique stable steady state that is not trivial, *i.e.*, with nonzero growth rate.
- (C3) This steady state corresponds to the optimal steady state  $(\hat{p}_{opt}^*, \hat{r}_{opt}^*)$ , allowing growth at the maximal rate  $\mu_{opt}^*$ .

It can be directly verified from the functions  $f$ ,  $g$ , and  $h$  defining the nutrient-only, precursor-only, and on-off control strategies (Eqs 2.15, 2.17, and 2.19) that they are indeed static functions of the system variables (or the system input, in the case of the nutrient-only strategy). Here we show that the other two conditions are also satisfied for all three strategies.

Following Eq. 2.15, the nutrient-only strategy is defined by  $\alpha = f(E_M)$ , so that  $\alpha$  is constant after the upshift. As shown above and in [S1 Text](#), this means that the system controlled by the nutrient-only strategy has a single nontrivial stable steady state (Condition C2). In addition, in this case the optimal steady state is attained for  $\alpha_{opt}^*$  defined as in Eq. 2.22, and the following function  $f$  therefore guarantees Condition C3:

$$f(E_M) = \frac{E_M + \sqrt{KE_M}}{E_M + 2\sqrt{KE_M} + 1}. \quad (2.27)$$

In [S1 Text](#), it is shown that Eq. 2.27 is the only definition of  $f$  satisfying all conditions. [S1 Figure](#) shows a plot of  $f(E_M)$  together with a biologically plausible Michaelis-Menten approximation (Eq. 2.16).

The full specification of the precursor-only strategy demands an expression for the function  $g$  in Eq. 2.17. Recall that Eq. 2.23 defines  $\alpha_{opt}^*$  in terms

of the precursor concentration  $\hat{p}_{opt}^*$ , which leads us to propose the following function  $g$ :

$$g(\hat{p}) = \frac{\hat{p}}{\hat{p} + \frac{K}{K+\hat{p}}(1 + \hat{p})}. \quad (2.28)$$

As shown in [S1 Text](#) by computing the Jacobian, the system given by Eqs [2.11-2.12](#) and [2.28](#) has a single nontrivial stable steady state for any environment  $E_M$  (Condition C2). Moreover, Eq. [2.28](#) guarantees this steady state to be optimal (Condition C3). This can be seen by noting that at steady state,  $d\hat{r}/dt = 0$  implies  $\hat{r}^* = g(\hat{p}^*)$  (Eq. [2.12](#)). In order for the self-replicator to attain a maximal growth rate at steady rate, Eq. [2.23](#) needs to be satisfied, which is the case for the above choice of the function  $g$ . Like for  $f$ , Eq. [2.28](#) is the only choice for  $g$  satisfying C1-C3. [S1 Figure](#) shows a plot of  $g(\hat{p})$  together with a biologically plausible Hill approximation (Eq. [2.18](#)).

The on-off control strategy is defined in Eq. [2.19](#) and repeated below:

$$h(\hat{p}, \hat{r}) = \begin{cases} 0, & \text{if } \hat{r} > g(\hat{p}), \\ 1, & \text{if } \hat{r} < g(\hat{p}), \\ \alpha_{opt}^*, & \text{if } (\hat{p}, \hat{r}) = (\hat{p}_{opt}^*, \hat{r}_{opt}^*). \end{cases} \quad (2.29)$$

This strategy drives the system to a single steady state, because the  $\hat{p}$ -nullcline crosses the function  $g(\hat{p})$  only once, as shown graphically in [Fig. 2.7A](#). In [S1 Text](#) we argue that this steady state is stable, by taking into account so-called sliding modes on the switching curve [\[164\]](#) (Condition C2). Moreover, the steady state coincides with the optimal steady state  $(\hat{p}_{opt}^*, \hat{r}_{opt}^*)$  by construction, so that Condition C3 is satisfied as well. [Fig 2.8A](#) shows a plot of  $h(\hat{p}, \hat{r})$ .

Note that since  $h(\cdot)$  is discontinuous, numerical instabilities occur during simulations. We therefore used the following continuous approximation of this function:

$$\frac{g(\hat{p})^{100}}{g(\hat{p})^{100} + \hat{r}^{100}}, \quad \text{if } \hat{r} \neq g(\hat{p}). \quad (2.30)$$

The approximation causes  $\alpha$  to take intermediate values (instead of 0 or 1) just before reaching the optimal steady state in [Fig 2.7C](#). For numerical simulations of the ODE system, we used the CVODE solver [\[165\]](#) from SUNDIALS 2.6.2 [\[166\]](#).



## 2.5 Supporting information for Chapter 2

### 2.5.1 S1 Text – Model derivation and analysis

#### Model formulation

The time evolution of the total mass of each component of the self-replicator can be written as follows:

$$\begin{aligned}\frac{dP}{dt} &= V_M(t) - V_R(t), \\ \frac{dM}{dt} &= (1 - \alpha(t)) V_R(t), \\ \frac{dR}{dt} &= \alpha(t) V_R(t),\end{aligned}\tag{2.31}$$

where  $P$ ,  $M$ ,  $R$  [g] denote the total mass of precursors, metabolic machinery and gene expression machinery, respectively.  $V_M$  [g h<sup>-1</sup>] is the rate of production of precursors by metabolism and  $V_R$  [g h<sup>-1</sup>] the rate of utilisation of precursors for gene expression.

Dividing the mass variables by the total time-varying volume  $\text{Vo1}(t)$  of the system, we obtain the concentration variables  $p = P/\text{Vo1}$ ,  $m = M/\text{Vo1}$ ,  $r = R/\text{Vo1}$  [g L<sup>-1</sup>]. The dynamics of the concentration variables then follows with Eq. 2.31:

$$\begin{aligned}\frac{dp}{dt} &= \frac{V_M(t)}{\text{Vo1}} - \frac{V_R(t)}{\text{Vo1}} - \frac{1}{\text{Vo1}} \frac{d\text{Vo1}}{dt} p, \\ \frac{dm}{dt} &= (1 - \alpha(t)) \frac{V_R(t)}{\text{Vo1}} - \frac{1}{\text{Vo1}} \frac{d\text{Vo1}}{dt} m, \\ \frac{dr}{dt} &= \alpha(t) \frac{V_R(t)}{\text{Vo1}} - \frac{1}{\text{Vo1}} \frac{d\text{Vo1}}{dt} r.\end{aligned}\tag{2.32}$$

At this point, we define  $v_M = V_M/\text{Vo1}$  and  $v_R = V_R/\text{Vo1}$  [g L<sup>-1</sup> h<sup>-1</sup>] as the mass fluxes per unit volume. Moreover, with the definition of the volume in terms of the total protein mass in Eq. 2.2 of the main text, that is,  $\text{Vo1} = \beta (M + R)$ , we find that

$$\frac{1}{\text{Vo1}} \frac{d\text{Vo1}}{dt} = \frac{\beta}{\text{Vo1}} \frac{d(M + R)}{dt} = \beta \frac{V_R(t)}{\text{Vo1}} = \beta v_R(t).\tag{2.33}$$

This leads to the system

$$\frac{dp}{dt} = v_M(t) - v_R(t) (1 + \beta p),\tag{2.34}$$

$$\frac{dr}{dt} = v_R(t) (\alpha(t) - \beta r),\tag{2.35}$$



where the equation for  $m(t)$  is omitted since by construction  $r(t)+m(t) = 1/\beta$  and  $dr/dt + dm/dt = 0$ .

As stated in the main text, we use Michaelis-Menten kinetics to express  $v_M$  and  $v_R$  in terms of the system variables:

$$\begin{aligned} v_M(t) &= m(t) k_M \frac{s(t)}{K_M + s(t)} = \left( \frac{1}{\beta} - r(t) \right) e_M(t), \\ v_R(t) &= r(t) k_R \frac{p(t)}{K_R + p(t)}, \end{aligned}$$

with rate constants  $k_M, k_R$  [ $\text{h}^{-1}$ ] and half-saturation constants  $K_M, K_R$  [ $\text{g L}^{-1}$ ].  $s(t)$  is an exogenous variable representing the nutrient concentration in the external medium. We simplify  $v_M(t)$  by defining the environmental input  $e_M(t) = k_M s(t)/(K_M + s(t))$ . Throughout the paper, as explained in the main text, we assume the environment is constant, *i.e.*,  $e_M(t) = e_M$ .

Finally, the growth rate  $\mu$  [ $\text{h}^{-1}$ ] is defined as the relative increase of the volume of the self-replicator. From Eq. 2.33, it follows that:

$$\mu(t) = \frac{1}{\text{Vol}} \frac{d\text{Vol}}{dt} = \beta v_R(t). \quad (2.36)$$

### Nondimensionalization of the system

For the sake of simplifying the proofs and derivations below, we define the following nondimensional variables:

$$\hat{p} = \beta p, \quad \hat{r} = \beta r, \quad \hat{t} = k_R t.$$

When injecting these into Eq. 2.34, we obtain

$$\frac{k_R}{\beta} \frac{d\hat{p}}{d\hat{t}} = \left( \frac{1}{\beta} - \frac{\hat{r}}{\beta} \right) e_M - \frac{\hat{r}}{\beta} k_R \frac{\hat{p}}{\beta K_R + \hat{p}} (1 + \hat{p}),$$

which simplifies to

$$\frac{d\hat{p}}{d\hat{t}} = (1 - \hat{r}) \frac{e_M}{k_R} - \hat{r} \frac{\hat{p}}{\beta K_R + \hat{p}} (1 + \hat{p}).$$

In a similar manner, we derive the time evolution of the nondimensional  $\hat{r}$ , and thus obtain the system

$$\begin{aligned} \frac{d\hat{p}}{d\hat{t}} &= (1 - \hat{r}) E_M - (1 + \hat{p}) \frac{\hat{p}}{K + \hat{p}} \hat{r}, \\ \frac{d\hat{r}}{d\hat{t}} &= (\alpha - \hat{r}) \frac{\hat{p}}{K + \hat{p}} \hat{r}, \end{aligned} \quad (2.37)$$

with the lumped parameters  $E_M = e_M/k_R$  and  $K = \beta K_R$ . The corresponding nondimensionalized growth rate is given by

$$\hat{\mu} = \frac{\mu}{k_R} = \frac{\hat{p}}{K + \hat{p}} \hat{r}. \quad (2.38)$$

### Steady-state growth of the self-replicator

If we suppose  $E_M > 0$ ,  $K > 0$  and  $\alpha \in ]0, 1[$ , there is a trivial unstable steady state at  $(0, 1)$ . A second steady-state exists for the point in which  $\hat{r}^* = \alpha$  and  $\hat{p}^*$  is a root of the following polynomial:

$$\alpha \hat{p}^2 + (\alpha - (1 - \alpha) E_M) \hat{p} - (1 - \alpha) E_M K.$$

If we keep the only admissible root for this polynomial (*i.e.*, for which  $\hat{p} \geq 0$ ), the second steady state is given by

$$(\hat{p}^*, \hat{r}^*) = \left( \frac{(1 - \alpha) E_M - \alpha + \sqrt{[(1 - \alpha) E_M - \alpha]^2 + 4\alpha(1 - \alpha) E_M K}}{2\alpha}, \alpha \right). \quad (2.39)$$

We can determine the stability of this steady state by looking at the Jacobian matrix  $J$  of the ODE system:

$$J = \begin{pmatrix} -\frac{\hat{r}}{K + \hat{p}} \left[ \hat{p} + (1 + \hat{p}) \frac{K}{K + \hat{p}} \right] & -E_M - (1 + \hat{p}) \frac{\hat{p}}{K + \hat{p}} \\ (\alpha - \hat{r}) \hat{r} \frac{K}{(K + \hat{p})^2} & (\alpha - 2\hat{r}) \frac{\hat{p}}{K + \hat{p}} \end{pmatrix}. \quad (2.40)$$

Evaluated at the point  $(\hat{p}^*, \hat{r}^*)$ , the Jacobian matrix becomes

$$J_{(\hat{p}^*, \hat{r}^*)} = \begin{pmatrix} -\frac{\alpha}{K + \hat{p}^*} \left[ \hat{p}^* + (1 + \hat{p}^*) \frac{K}{K + \hat{p}^*} \right] & -E_M - (1 + \hat{p}^*) \frac{\hat{p}^*}{K + \hat{p}^*} \\ 0 & -\alpha \frac{\hat{p}^*}{K + \hat{p}^*} \end{pmatrix}.$$

Since  $\hat{p}^*$ ,  $\alpha$ ,  $E_M$ ,  $K > 0$ , the two eigenvalues are negative and therefore the steady state  $(\hat{p}^*, \hat{r}^*)$  is stable (see also the streamlines in Figure 2.2A in the main text). It means that for fixed environmental conditions  $E_M$  and resource allocation  $\alpha$ , the self-replicator converges towards a steady state in which the concentration variables are constant.

One can now easily derive the steady-state growth rate, denoted  $\hat{\mu}^*$ . By substituting Eq. 2.38 into the first ODE of the system of Eq. 2.37, we find at steady state:

$$\left( \frac{d\hat{p}}{dt} \right)_{(\hat{p}^*, \hat{r}^*)} = 0 = (1 - \alpha) E_M - (1 + \hat{p}^*) \hat{\mu}^*,$$

which by means of Eq. 2.39 gives the following relation:

$$\begin{aligned}\hat{\mu}^* &= \frac{(1-\alpha) E_M}{1+\hat{p}^*} \\ &= \frac{2\alpha(1-\alpha) E_M}{(1-\alpha) E_M + \alpha + \sqrt{[(1-\alpha) E_M - \alpha]^2 + 4\alpha(1-\alpha) E_M K}}.\end{aligned}\quad (2.41)$$

Finally, we can transform this expression to obtain

$$\hat{\mu}^* = \begin{cases} \frac{(1-\alpha) E_M + \alpha - \sqrt{[(1-\alpha) E_M - \alpha]^2 + 4(1-\alpha)\alpha E_M K}}{2(1-K)} & \text{for } K \neq 1, \\ \frac{\alpha(1-\alpha) E_M}{\alpha + (1-\alpha) E_M} & \text{for } K = 1. \end{cases}\quad (2.42)$$

This function of  $\alpha$  is plotted in Figure 2.2B in the main text.

### Maximization of growth rate at steady state

We are interested in the steady state at which growth occurs at the maximum rate. The growth rate at steady state  $\hat{\mu}^*$  is given by

$$\hat{\mu}^* = \frac{\hat{p}^*}{K + \hat{p}^*} \hat{r}^*.\quad (2.43)$$

From the first ODE of the system of Eq. 2.37, we have

$$\hat{r}^* = \frac{E_M}{E_M + \frac{\hat{p}^*}{K + \hat{p}^*} (1 + \hat{p}^*)}.\quad (2.44)$$

Substituting Eq. 2.44 into Eq. 2.43, we obtain

$$\hat{\mu}^* = \frac{E_M \hat{p}^*}{\hat{p}^{*2} + (E_M + 1) \hat{p}^* + E_M K}.\quad (2.45)$$

The value of  $\hat{p}^*$  maximizing  $\hat{\mu}^*$  can be determined from

$$\frac{\partial \hat{\mu}^*}{\partial \hat{p}^*} = \frac{E_M (E_M K - \hat{p}^{*2})}{(\hat{p}^{*2} + (E_M + 1) \hat{p}^* + E_M K)^2},\quad (2.46)$$

by looking at the values of  $\hat{p}^*$  for which this derivative equals 0. It follows that  $\hat{\mu}^*$  is maximal for

$$\hat{p}^* = \hat{p}_{opt}^* = \sqrt{K E_M}.\quad (2.47)$$

By substituting  $\hat{p}_{opt}^*$  and  $\alpha_{opt}^*$  for  $\hat{p}^*$  and  $\hat{r}^*$ , respectively, in Eq. 2.44, we obtain the resource allocation maximizing the growth rate

$$\alpha_{opt}^* = \frac{E_M + \sqrt{K E_M}}{E_M + 2\sqrt{K E_M} + 1}.\quad (2.48)$$

Finally, injecting this result into Eq. 2.43 we obtain the optimal steady-state growth rate:

$$\hat{\mu}_{opt}^* = \frac{E_M}{E_M + 2\sqrt{K E_M} + 1}. \quad (2.49)$$

In addition, by using Eq. 2.47, we can write  $\alpha_{opt}^*$  and  $\hat{\mu}_{opt}^*$  as a function of  $\hat{p}_{opt}^*$  only:

$$\alpha_{opt}^* = \frac{\hat{p}_{opt}^*}{\hat{p}_{opt}^* + \frac{K}{K + \hat{p}_{opt}^*}(1 + \hat{p}_{opt}^*)}, \quad \hat{\mu}_{opt}^* = \frac{\hat{p}_{opt}^{*2}}{\hat{p}_{opt}^{*2} + 2K\hat{p}_{opt}^* + K}. \quad (2.50)$$

### Analysis of the control strategies

In this section, we derive the main results for the functions  $f$ ,  $g$ , and  $h$  defining the nutrient-only, precursor-only, and on-off control strategies. For each of these, we prove that the Conditions C1, C2 and C3 from the *Methods* section are satisfied, which we repeat here for clarity:

- (C1) The control laws are static functions of the system variables (as opposed to, for instance, functions that depend on derivatives or integrals of the variables).
- (C2) For any given constant environment  $E_M$ , they drive the self-replicator system towards a unique stable steady state that is not trivial, *i.e.*, with nonzero growth rate.
- (C3) This steady state corresponds to the optimal steady state  $(\hat{p}_{opt}^*, \hat{r}_{opt}^*)$ , allowing growth at the maximal rate  $\mu_{opt}^*$ .

**Nutrient-only strategy** The nutrient-only strategy is defined by:

$$\alpha = f(E_M) = \frac{E_M + \sqrt{K E_M}}{E_M + 2\sqrt{K E_M} + 1}. \quad (2.51)$$

It drives the system to the optimal steady state by measuring the environment  $E_M$ . Note that Condition C1 is satisfied by definition.

By injecting Eq. 2.51 into Eq. 2.37, the ODE system under the control of  $f$  becomes:

$$\begin{aligned} \frac{d\hat{p}}{dt} &= (1 - \hat{r}) E_M - (1 + \hat{p}) \frac{\hat{p}}{K + \hat{p}} \hat{r}, \\ \frac{d\hat{r}}{dt} &= (f(E_M) - \hat{r}) \frac{\hat{p}}{K + \hat{p}} \hat{r}. \end{aligned} \quad (2.52)$$

Since  $E_M$  is constant on the interval of interest (starting right after the upshift), we are in the case of Section S1 Text (*i.e.*,  $\alpha$  constant). In particular,

the system has two steady states: a trivial unstable one at  $(0, 1)$  (with zero growth), and a stable one defined by Eq. 2.39 (Condition C2). Since  $f(E_M) = \alpha_{opt}^*$ , we conclude from the derivations in Section S1 Text that the stable steady state is optimal for every environment  $E_M$  (Condition C3).

It is interesting to note that the expression in Eq. 2.51 is the only function  $f(E_M)$  satisfying C1-C3. We can prove this statement by contradiction. Assume a control strategy  $c(E_M)$  satisfying C1-C3, and different from  $f(E_M)$ , *i.e.*, there exists  $E_M = E_{M_1}$  such that  $c(E_{M_1}) \neq f(E_{M_1})$ . In this environment, the system reaches a steady state  $(\hat{p}_1^*, \hat{r}_1^*)$  with  $\hat{r}_1^* = c(E_{M_1}) \neq f(E_{M_1})$ . However, by Eq. 2.48 the optimal value for  $\hat{r}^*$  in this environment is given by  $f(E_{M_1})$ . So, the control law  $c(E_M)$  does not drive the system to the optimal steady state in this environment, in contradiction with Condition C3.

**Precursor-only strategy** The precursor-only strategy is defined by:

$$\alpha = g(\hat{p}) = \frac{\hat{p}}{\hat{p} + \frac{K}{K+\hat{p}}(1 + \hat{p})}. \quad (2.53)$$

Here as well, C1 is satisfied by construction.

The ODE system under the control of  $g$  becomes

$$\begin{aligned} \frac{d\hat{p}}{dt} &= (1 - \hat{r}) E_M - (1 + \hat{p}) \frac{\hat{p}}{K + \hat{p}} \hat{r}, \\ \frac{d\hat{r}}{dt} &= (g(\hat{p}) - \hat{r}) \frac{\hat{p}}{K + \hat{p}} \hat{r}. \end{aligned} \quad (2.54)$$

The nullcline for  $\hat{p}$  remains unchanged and is defined by

$$\frac{d\hat{p}}{dt} = 0 \Leftrightarrow \hat{r} = \frac{E_M}{E_M + \frac{\hat{p}}{K+\hat{p}}(1 + \hat{p})}, \quad (2.55)$$

while the nullcline for  $\hat{r}$  is

$$\frac{d\hat{r}}{dt} = 0 \Leftrightarrow \begin{cases} \hat{p} = 0, \\ \hat{r} = 0, \\ \hat{r} = \frac{\hat{p}}{\hat{p} + \frac{K}{K+\hat{p}}(1 + \hat{p})}. \end{cases} \quad (2.56)$$

Hence, we also have a trivial unstable steady state at  $(0,1)$  (with zero growth). The second steady state is obtained from Eqs 2.55-2.56:

$$\frac{E_M}{E_M + \frac{\hat{p}^*}{K+\hat{p}^*}(1 + \hat{p}^*)} = \frac{\hat{p}^*}{\hat{p}^* + \frac{K}{K+\hat{p}^*}(1 + \hat{p}^*)},$$

which we rearrange into

$$\hat{p}^* E_M + \frac{K}{K + \hat{p}^*} (1 + \hat{p}^*) E_M = \hat{p}^* E_M + \frac{\hat{p}^*}{K + \hat{p}^*} (1 + \hat{p}^*) \hat{p}^*.$$

This leads to

$$\hat{p}^* = \sqrt{K E_M},$$

and therefore

$$\hat{r}^* = g(\hat{p}^*) = \frac{\sqrt{K E_M}}{\sqrt{K E_M} + \frac{K}{K + \sqrt{K E_M}} (1 + \sqrt{K E_M})} = \frac{E_M + \sqrt{K E_M}}{E_M + 2\sqrt{K E_M} + 1}.$$

From Eqs 2.47-2.48, we recognize the optimal steady state for the environment  $E_M$ , validating Condition C3. We now look for the stability of this (optimal) steady state by deriving the Jacobian of this system:

$$J = \begin{pmatrix} -\frac{\hat{r}}{K + \hat{p}} \frac{\hat{p}^2 + 2K\hat{p} + K}{\hat{p} + K} & -E_M - \frac{\hat{p}}{K + \hat{p}} (1 + \hat{p}) \\ \frac{\hat{r}}{K + \hat{p}} \left[ \frac{K}{K + \hat{p}} (g(\hat{p}) - \hat{r}) + \hat{p}K \frac{\hat{p}^2 + 2\hat{p} + K}{(\hat{p}^2 + 2K\hat{p} + K)^2} \right] & \frac{\hat{p}}{K + \hat{p}} (g(\hat{p}) - 2\hat{r}) \end{pmatrix}. \quad (2.57)$$

Evaluated at  $(\hat{p}^*, \hat{r}^*) = (\sqrt{K E_M}, g(\sqrt{K E_M}))$ , the Jacobian  $J_{(\hat{p}^*, \hat{r}^*)}$  becomes

$$\begin{pmatrix} -\frac{\sqrt{E_M}}{\sqrt{K + \sqrt{E_M}}} & -E_M - \frac{\sqrt{E_M}}{\sqrt{K + \sqrt{E_M}}} (1 + \sqrt{K E_M}) \\ \frac{\sqrt{E_M}}{\sqrt{K + \sqrt{E_M}}} \frac{K E_M + 2\sqrt{K E_M} + K}{K(E_M + 2\sqrt{K E_M} + 1)^2} g(\sqrt{K E_M}) & -\frac{\sqrt{E_M}}{\sqrt{K + \sqrt{E_M}}} g(\sqrt{K E_M}) \end{pmatrix}. \quad (2.58)$$

Since  $K$ ,  $E_M$ , and  $g(\sqrt{K E_M}) > 0$ , it follows immediately that the real part of the eigenvalues of this matrix are both negative.<sup>1</sup> Hence, the non-trivial steady state is stable, completing the proof of Condition C2.

Here again, it is interesting to observe that the expression in Eq. 2.53 is the only function  $g(\hat{p})$  satisfying C1-C3. This can be proven in a similar way as for  $f$ .

**On-off strategy** The on-off strategy is defined by:

$$\alpha = h(\hat{p}, \hat{r}) = \begin{cases} 0, & \text{if } \hat{r} > g(\hat{p}), \\ 1, & \text{if } \hat{r} < g(\hat{p}), \\ \alpha_{opt}^*, & \text{if } (\hat{p}, \hat{r}) = (\hat{p}_{opt}^*, \hat{r}_{opt}^*). \end{cases} \quad (2.59)$$

$h$  is a static function of  $\hat{p}$  and  $\hat{r}$  (Condition C1).

1. Notice that the eigenvalues  $\lambda_1$  and  $\lambda_2$  of  $J_{(\hat{p}^*, \hat{r}^*)}$  satisfy the inequalities  $\text{Tr}(J) = \lambda_1 + \lambda_2 < 0$  and  $\det(J) = \lambda_1 \lambda_2 > 0$ .

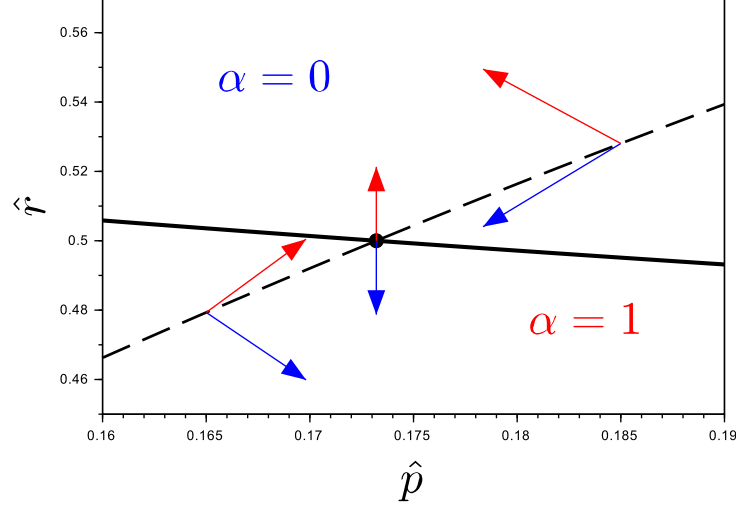


Figure 2.9 – **Local stability of the on-off strategy.** The on-off strategy sets  $\alpha$  to a value of 0 (1) when  $\hat{r} > g(\hat{p})$  ( $\hat{r} < g(\hat{p})$ ). The solid, black curve is the  $\hat{p}$ -nullcline. The dashed, black curve is the curve  $\hat{r} = g(\hat{p})$ . The arrows represent the vector fields for  $\alpha = 0$  (in blue) and  $\alpha = 1$  (in red). The intersection of the  $\hat{p}$ -nullcline and the curve  $\hat{r} = g(\hat{p})$  corresponds to a unique non-trivial stable steady state, which is equal to  $(\hat{p}_{opt}^*, \hat{r}_{opt}^*)$  by Eq. 2.60.

As a consequence, the ODE system under the control of  $h$  is given by

$$\begin{aligned} \frac{d\hat{p}}{d\hat{t}} &= (1 - \hat{r}) E_M - (1 + \hat{p}) \frac{\hat{p}}{K + \hat{p}} \hat{r}, \\ \frac{d\hat{r}}{d\hat{t}} &= (h(\hat{p}, \hat{r}) - \hat{r}) \frac{\hat{p}}{K + \hat{p}} \hat{r}. \end{aligned} \quad (2.60)$$

Notice that the system has a discontinuous right-hand side, due to the fact that  $\alpha$  switches between 0 and 1 on  $\hat{r} = g(\hat{p})$ . Fig. 2.9 shows the dynamics of the system in the phase plane. Due to the direction of the vector fields relative to  $\hat{r} = g(\hat{p})$ , a *sliding mode* occurs on the latter curve [164]. The system is seen to evolve towards a locally asymptotically stable steady state, which is the single non-trivial steady state (Condition C2). This steady state coincides with the intersection of  $\hat{r} = g(\hat{p})$  and the  $\hat{p}$ -nullcline, which is the steady state  $(\hat{p}_{opt}^*, \hat{r}_{opt}^*)$  allowing maximal growth, thus verifying Condition C3.

## 2.5.2 S2 Text – Model parameters

Most of the conclusions of this paper are parameter-independent in the range of physically admissible values. The exact parameter values used in the simulations aim to represent relevant orders of magnitude. They were derived from the available literature on fast-growing bacteria (mostly *Escherichia coli*). This document describes this derivation for each parameter. In the *Methods* section of the main text, we also describe how some of them were validated by fitting the model to available experimental data (Fig. 2.3 in the main text).

The model parameters (Eqs. 2.3-2.4 in the main text) are listed in the table below:

Name	Unit	Description
$e_M$	$\text{h}^{-1}$	Constant characterizing nutrient composition of medium
$k_R$	$\text{h}^{-1}$	Rate constant of macromolecular synthesis
$K_R$	$\text{g L}^{-1}$	Half-saturation constant of macromolecular synthesis
$\beta$	$\text{L g}^{-1}$	Inverse of the cellular density of macromolecules
$\alpha$	–	Resource allocation parameter

The values derived below are summarized in S1 Table.

### $e_M$

By definition,  $e_M$  is the effective turnover of the metabolic macroreaction producing precursors from external substrates, obtained by dividing the reaction rate  $v_M$  by the enzyme concentration  $m$  (Eq. 2.6). The unit of  $e_M$  is  $\text{min}^{-1}$ , and can be decomposed as follows:

$$[e_M] = \frac{[\text{mass of metabolic product}]}{[\text{mass of enzyme M}] \cdot [\text{time}]} = \frac{1}{[\text{time}]}.$$

Note that  $e_M = k_M s / (K_M + s)$  where  $k_M$  is a rate constant, indicating the maximal rate of conversion of external nutrients to precursor metabolites.  $e_M$  will thus vary with the concentration  $s$  of the external nutrients and the kind of nutrient. For example, the precursor mass that can be produced from 1 g of glucose is higher than that produced from 1 g of acetate.

How can we find a typical value for  $k_M$ , and thus for  $e_M$  (both have the same order of magnitude if we suppose that the reaction is not operating far below saturation, that is,  $e_M \approx k_M$ )? A reasonable estimate for  $k_M$  can be obtained from the turnover numbers of reactions involved in the synthesis of charged tRNA, since the latter are directly consumed by the most abundant part of the gene expression machinery, the ribosomes.



Ref. [167] provides a typical value for such a reaction, catalyzed by glutaminyl-tRNA synthetase:  $k_{cat, GlnRS} = 3.2 \text{ s}^{-1}$ , indicating that on average 3.2 glutaminyl-tRNA molecules are produced per glutaminyl-tRNA synthetase molecule per second. After conversion to mass units using molar weight from [168], this yields

$$k_{cat, GlnRS} = \frac{3.2 \cdot 147}{64.4 \cdot 10^3} \approx 10^{-3} \text{ g of glutaminyl-tRNA} \cdot \text{g of enzyme}^{-1} \cdot \text{s}^{-1}.$$

We therefore take

$$k_M \approx 3.6 \text{ h}^{-1},$$

and thus obtain an upper bound for  $e_M$  in our simulations.

### $k_R$

$k_R$  is the mass rate constant describing the maximal rate of conversion of precursors to macromolecules [ $\text{h}^{-1}$ ]. As for  $e_M$ , we can decompose this into

$$[k_R] = \frac{[\text{mass of macromolecules}]}{[\text{mass of gene expression machinery}] \cdot [\text{time}]} = \frac{1}{[\text{time}]}$$

To obtain an order of magnitude for the mass of macromolecules, we focus on proteins since they are the most abundant macromolecules in the cell [30]. The dimensional analysis of  $k_R$  thus becomes:

$$\begin{aligned} [k_R] &= \frac{[\text{protein mass produced}]}{[\text{ribosomal mass}] \cdot [\text{hour}]} \\ &= \frac{[\text{moles of protein}] \cdot [\text{protein molar mass}]}{[\text{moles of ribosome}] \cdot [\text{ribosome molar mass}] \cdot [\text{hour}]} \\ &= \frac{[\text{moles of amino acids}] \cdot [\text{molar mass of amino acids}]}{[\text{moles of ribosome}] \cdot [\text{hour}] \cdot [\text{ribosome molar mass}]} \\ &\approx \frac{[\text{maximal protein elongation rate}] \cdot [\text{molar mass of amino-acids}]}{[\text{ribosome molar mass}]} \end{aligned}$$

The values in the last equality are available from the literature [30, 132, 169, 170]. We obtain

$$k_R \approx \frac{10 \cdot 100}{10^6} \cdot 3600 \approx 3.6 \text{ h}^{-1}.$$

This value is comparable with the translational capacity  $k_T$ , in  $\mu\text{g}$  of protein per  $\mu\text{g}$  of ribosomal protein per hour, given by Scott *et al.* [33]:

$$k_T = \frac{4.5 \mu\text{g of protein} / \mu\text{g of RNA} / \text{h}}{0.76 \mu\text{g of ribosomal protein} / \mu\text{g of RNA}} = 5.9 \text{ h}^{-1}.$$

$K_R$ 

A value for the parameter  $K_R$ , representing the half-saturation constant of macromolecular synthesis, is more difficult to obtain from the literature. However, assuming that ribosomes operate close to saturation (80% over a range of growth rates [30]), we find that  $K_R \approx 0.25 p$ , with  $p$  the total amino acid concentration. The total concentration of amino acids in the cell is around  $150 \text{ mmol L}^{-1}$  [60], which with a mean molecular weight of  $118.9 \text{ g mol}^{-1}$  for amino acids [169], yields a mass concentration of  $17.8 \text{ g L}^{-1}$ . These considerations led to the following order of magnitude for  $K_R$ :

$$K_R \approx 1 \text{ g L}^{-1}.$$

 $\beta$ 

$\beta$  is the inverse of the cellular density of macromolecules, which has been shown constant during balanced growth over a large range of growth rates [77], and there is some data suggesting that  $\beta$  varies little during growth transitions as well [171]. From [172, 173] we take the following typical value for  $\beta$ :

$$\beta \approx \frac{1}{300} \approx 0.003 \text{ L g}^{-1}.$$

 $E_M$  and  $K$ 

From the values of the parameter in the dimensional model, one can deduce the parameters in the nondimensional model used in the simulations:

$$E_M = \frac{e_M}{k_R} = \frac{3.6}{3.6} = 1 \quad , \quad K = \beta K_R = 3 \cdot 10^{-3} \cdot 1 = 0.003.$$

### 2.5.3 S3 Text – Solution of optimal control problem

#### Statement of the problem

We consider the dimensionless system defined by Eqs 2.11-2.12 in the main text, which are here repeated for clarity:

$$\begin{aligned}\frac{d\hat{p}}{d\hat{t}} &= (1 - \hat{r}) E_M - (1 + \hat{p}) \hat{r} \frac{\hat{p}}{K + \hat{p}}, \\ \frac{d\hat{r}}{d\hat{t}} &= \hat{r} \frac{\hat{p}}{K + \hat{p}} (\alpha(\hat{t}) - \hat{r}).\end{aligned}\tag{2.61}$$

As stated in the section *Biomass maximization as an optimal control problem* in the main text, the objective of this study is to maximize the growth rate on an interval  $[0, \tau]$  after a nutrient upshift. With Eq. 2.38 in S1 Text, we have

$$\hat{\mu} = \hat{r} \frac{\hat{p}}{K + \hat{p}}.$$

In order to avoid boundary effects occurring over finite time intervals, notably the depletion of precursors just before  $\tau$ , we solve the optimal control problem over an infinite horizon ( $\tau \rightarrow \infty$ ). Consider the set of admissible controls

$$\mathcal{U} = \{\alpha : \mathbb{R} \rightarrow [0, 1] \mid \alpha(\cdot) \text{ measurable}\}.$$

The optimization problem can then be stated as follows:

$$\alpha_{opt} = \arg \max_{\alpha \in \mathcal{U}} J(\alpha) \equiv \int_0^{+\infty} \hat{r}(\hat{t}) \frac{\hat{p}(\hat{t})}{K + \hat{p}(\hat{t})} d\hat{t},\tag{2.62}$$

where  $(\hat{p}(\hat{t}), \hat{r}(\hat{t}))$  is the unique solution of Eq. 2.61 starting at a given point  $(\hat{p}_0, \hat{r}_0) \in \Omega \equiv \mathbb{R}_*^+ \times (0, 1)$  for a given control  $\alpha \in \mathcal{U}$ .

Given that the performance index  $J(\alpha)$  diverges, we actually consider *overtaking optimality* [114]. Consider the performance index of the trajectory  $x(\cdot)$  emanating from  $x_0$  and generated by  $u(\cdot)$  defined for any  $T \geq 0$  by

$$J_T(x_0, u(\cdot)) = \int_0^T f_0(x(t), u(t), t) dt.$$

A trajectory  $x^*(\cdot)$  emanating from  $x_0$  and generated by  $u^*(\cdot)$  is said to be *overtaking optimal* if for any other trajectory  $x(\cdot)$  emanating from  $x_0$  and generated by  $u(\cdot)$  the following holds

$$\liminf_{T \rightarrow \infty} \{J_T(x_0, u^*(\cdot)) - J_T(x_0, u(\cdot))\} \geq 0.$$

Roughly speaking, a trajectory is *overtaking optimal* if "the performance index catches up with the performance index of any other trajectory" [114].

### Maximum Principle

Necessary conditions on optimal trajectories can be obtained by the Infinite Horizon Maximum Principle [114]. Let  $H(\hat{p}, \hat{r}, \lambda_p, \lambda_r, \lambda_0, \alpha)$  be the Hamiltonian of the system, defined by

$$H(\cdot) \equiv \lambda_p E_M (1 - \hat{r}) - \hat{r} \frac{\hat{p}}{K + \hat{p}} [\lambda_p (1 + \hat{p}) + \lambda_r \hat{r} + \lambda_0] + \alpha \lambda_r \hat{r} \frac{\hat{p}}{K + \hat{p}}.$$

Moreover, let  $\alpha$  be an optimal control, and  $\hat{x}(\cdot) = (\hat{p}(\cdot), \hat{r}(\cdot))$  the associated trajectory. Then, there exists  $\lambda_0 \leq 0$  and an absolutely continuous map  $\lambda = (\lambda_p, \lambda_r) : [0, +\infty) \rightarrow \mathbb{R}^2$  such that  $(\lambda, \lambda_0) \neq 0$ , and

$$\dot{\lambda}_p = -\frac{\partial H}{\partial \hat{p}} = \hat{r} \frac{K}{(K + \hat{p})^2} [\lambda_p (1 + \hat{p}) + \lambda_r (\hat{r} - \alpha) + \lambda_0] + \hat{r} \frac{\hat{p}}{K + \hat{p}} \lambda_p, \quad (2.63)$$

$$\dot{\lambda}_r = -\frac{\partial H}{\partial \hat{r}} = \lambda_p E_M + \frac{\hat{p}}{K + \hat{p}} [\lambda_p (1 + \hat{p}) + \lambda_r (2\hat{r} - \alpha) + \lambda_0]. \quad (2.64)$$

The maximization condition is given by:

$$\alpha(\hat{t}) \in \arg \max_{v \in [0,1]} H(\hat{x}(\hat{t}), \lambda(\hat{t}), \lambda_0, v), \quad (2.65)$$

almost everywhere on  $[0, +\infty)$ .

An extremal trajectory is a quadruplet  $(\hat{x}(\cdot), \lambda(\cdot), \lambda_0, \alpha(\cdot))$  satisfying Eqs 2.61-2.65. The extremal is said to be normal (resp. abnormal) if  $\lambda_0 < 0$  (resp.  $\lambda_0 = 0$ ). In the normal case, we normalize the adjoint vector so that  $\lambda_0 = -1$ .

From Eq. 2.65, it follows that the control strategy is given by the sign of the *switching function*  $\phi(\cdot) \equiv \lambda_r \hat{r} \hat{p} / (K + \hat{p})$ , that is,

$$\begin{cases} \alpha = 1 & \iff \phi(\cdot) > 0, \\ \alpha = 0 & \iff \phi(\cdot) < 0. \end{cases}$$

Finally, given that the system is autonomous, the Hamiltonian is conserved along any extremal trajectory.

### Characterization of singular arcs

Whenever  $\phi$  is vanishing over a time interval, we say that the trajectory is *singular*. We will now characterize such trajectories. If  $I = [\hat{t}_1, \hat{t}_2]$  is a

singular arc, we have  $\phi(\hat{t}) = \dot{\phi}(\hat{t}) = 0$ , for all  $\hat{t} \in [\hat{t}_1, \hat{t}_2]$ , that is,  $\lambda_r(\hat{t}) = 0$  and  $\dot{\lambda}_r(\hat{t}) = 0$ .

For abnormal extremal trajectories, we get  $\lambda_p(\hat{t}) = 0$ , in contradiction with the Maximum Principle, so there is no singular arc. An abnormal extremal trajectory is therefore a concatenation of bang arcs.

For normal extremal trajectories, using additionally that  $H$  is constant along an extremal trajectory, we obtain that  $\lambda_p$  is constant along a singular arc. By combining  $\dot{\lambda}_p = 0$  and  $\dot{\lambda}_r = 0$ , we obtain  $\hat{p}(\hat{t}) = \sqrt{E_M K} = \hat{p}_{opt}^*$ . Using  $d\hat{p}/d\hat{t} = 0$ , we finally get  $\hat{r}(\hat{t}) = \hat{r}_{opt}^*$ . Thus, the singular arc is the optimal steady state, corresponding to a singular control  $\alpha(\hat{t}) = \alpha_{opt}^*$ , with  $\alpha_{opt}^*$  depending on  $E_M$  (S1 Text).

A necessary condition of optimality for a singular arc is given by the Kelley condition [163]. We must differentiate  $\phi$  with respect to  $\hat{t}$  until  $\alpha$  appears in the derivative. Along a singular arc, we obtain for  $q = 2$ :

$$(-1)^q \frac{\partial}{\partial \alpha} \frac{d^{2q}}{d\hat{t}^{2q}} \phi(\hat{t}) < 0,$$

satisfying the Kelley condition necessary for optimality. Given that the singular arc is of second order, an optimal trajectory can enter into the singular arc only by a *chattering arc* (also called the Fuller's phenomenon, *i.e.*, an arc with an infinite number of switches [163, 174]).

### Analysis of the adjoint system

Recalling that a switch corresponds to a change of sign of  $\lambda_r$ , the analysis of the adjoint system (Eqs 2.63-2.64) may be useful to characterize the switches of extremal trajectories.

First, for the abnormal case, we can easily determine in the phase-plane the possible transitions between the four regions defined by the axes (see Fig. 2.10). A trajectory can cross at most twice the  $\lambda_p$ -axis, so we conclude that an abnormal extremal cannot have more than two switches. Thus, an abnormal extremal is a concatenation of at most three bang arcs ( $\alpha(t) = 0$  or  $\alpha(t) = 1$ ). When  $\alpha(t) = 0$  or  $\alpha(t) = 1$  for a long time, the growth rate tends to zero. We therefore conclude that abnormal extremal trajectories are not optimal.

Secondly, for the normal case, after the first switch, a trajectory with two consecutive switches in the regions  $\{(\hat{p}, \hat{r}) \in \Omega \mid \hat{p} < \hat{p}_{opt}^*\}$  or  $\{(\hat{p}, \hat{r}) \in \Omega \mid \hat{p} > \hat{p}_{opt}^*\}$  is not possible, as shown in Fig. 2.10. Therefore, such a trajectory is not optimal given that it does not fulfill the conditions given by the Maximum Principle. We conclude that if the optimal trajectory has

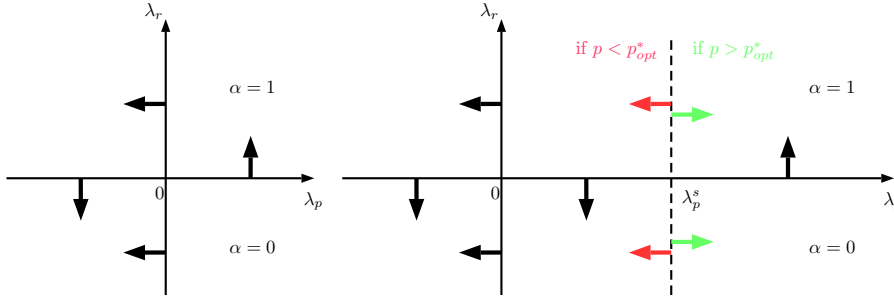


Figure 2.10 – **Transitions between regions in the phase-plane for the adjoint system.** A switch occurs when a trajectory crosses the  $\lambda_p$ -axis. Left: abnormal case. An extremal trajectory cannot have more than two switches. Right: normal case.  $(\lambda_p^s, 0)$  corresponds to the singular arc. After the first switch, an extremal trajectory cannot have two consecutive switches if it stays in the region  $\{(\hat{p}, \hat{r}) \in \Omega \mid \hat{p} < \hat{p}_{opt}^*\}$  or  $\{(\hat{p}, \hat{r}) \in \Omega \mid \hat{p} > \hat{p}_{opt}^*\}$ .

a concatenation of bang arcs, the switches must alternately occur in the regions  $\{(\hat{p}, \hat{r}) \in \Omega \mid \hat{p} < \hat{p}_{opt}^*\}$  and  $\{(\hat{p}, \hat{r}) \in \Omega \mid \hat{p} > \hat{p}_{opt}^*\}$ .

### Optimal trajectories

From the Maximum Principle, we have shown that the optimal trajectory is a concatenation of bang arcs ( $\alpha(t) = 0$  or  $\alpha(t) = 1$ ) and possibly a singular arc corresponding to the optimal steady state  $(\hat{p}(\hat{t}), \hat{r}(\hat{t})) = (\hat{p}_{opt}^*, \hat{r}_{opt}^*)$ . Moreover, if the optimal trajectory has a singular arc, it must enter it through a chattering arc (*i.e.*, with an infinite number of switches between  $\alpha = 0$  and  $\alpha = 1$ ).

These elements motivate the supposition that optimal solutions consist in a transient (chattering arc) towards the optimal steady state, after which they remain there (until the next change of environment). The chattering arc can be characterized by a switching curve  $\hat{p} \mapsto \varphi(\hat{p})$  which passes through the optimal steady state. Defining  $A_0$  and  $A_1$  the regions above and below  $\varphi$  in the  $(\hat{p}, \hat{r})$ -plane, respectively, we conjecture that the following feedback control law is optimal:

$$\begin{cases} \alpha(\hat{t}) = 0 & \text{if } (\hat{p}(\hat{t}), \hat{r}(\hat{t})) \in A_0, \\ \alpha(\hat{t}) = 1 & \text{if } (\hat{p}(\hat{t}), \hat{r}(\hat{t})) \in A_1, \\ \alpha(\hat{t}) = \alpha_{opt} & \text{if } (\hat{p}(\hat{t}), \hat{r}(\hat{t})) = (\hat{p}_{opt}^*, \hat{r}_{opt}^*). \end{cases} \quad (2.66)$$

Loosely speaking, the chattering arc corresponds to a spiral composed of bang arcs wrapping around the optimal steady state, where the switches

alternatingly occur in the regions  $\{(\hat{p}, \hat{r}) \in \Omega \mid \hat{p} < \hat{p}_{opt}^*\}$  and  $\{(\hat{p}, \hat{r}) \in \Omega \mid \hat{p} > \hat{p}_{opt}^*\}$ , in line with the analysis of the adjoint system. This is a first hint that the proposed control strategy is optimal. Moreover, our conjecture is also in line with the *turnpike property*: Trélat and Zuazua [175] have shown that, for quite a generic class of systems, the optimal strategy consists in staying at the optimal steady state (after a short transient).

As explained in the *Methods* section of the main text, we numerically solved the optimal control problem by the direct method using the `bocop` software [115]. It is important to stress that the optimization process was performed without any preliminary assumptions on the characteristics of the optimal trajectory. The fact that the numerical solution verifies the Maximum Principle (*i.e.*, the singular arc corresponds to the optimal steady state) and the Kelley condition (*i.e.*, the presence of a chattering arc) tends to confirm that the control strategy of Eq. 2.66 is optimal. As an aside, we note that due to the fact that numerical optimization was performed for a finite horizon, we actually obtained a second chattering arc escaping from the singular arc at the end of the simulation. This is a classical property of the turnpike strategy: the optimal trajectory leaves the optimal steady state just before the end of the time interval of interest, in our case consuming almost all precursors. This arc was removed from the plot in Fig. 2.4, because it does not occur with an infinite horizon and is therefore a numerical artifact for this study.

### 2.5.4 S4 Text – Kinetic model of the ppGpp system in *Escherichia coli*

The recently published model of Bosdriesz *et al.* [116] provides a synthesis of the currently available knowledge of the ppGpp regulatory system. Through the mechanisms of ppGpp production and degradation, it describes regulation of the synthesis of ribosomal RNA. We explain below how we use the model to compare the action of the ppGpp system with the on-off control strategy. The denomination of variables and parameters follows the Supporting Information of [116], and is reproduced in Table 2.1 in order to make the text self-contained.

The evolution of the cellular concentration of ppGpp is described in [116] by

$$\frac{d\text{ppGpp}}{dt} = v_{\text{RelA}}(r_{t,\text{tot}}) + v_{\text{spoT}} - k_{\text{spoT}} \cdot \text{ppGpp}, \quad (2.67)$$

where  $v_{\text{spoT}}$  and  $k_{\text{spoT}}$  are constants (see Table 2.1), and  $v_{\text{RelA}}$  is a function of  $r_{t,\text{tot}}$ , the total concentration of "stalled" ribosomes:

$$v_{\text{RelA}}(r_{t,\text{tot}}) = k_{\text{RelA}} \cdot \text{RelA}_{\text{tot}} \cdot \frac{r_{t,\text{tot}}}{K_{D,\text{RelA}} + r_{t,\text{tot}}}. \quad (2.68)$$

The amount of stalled ribosomes is determined by the equilibrium between charged and uncharged tRNA,  $t_{ai}$  and  $t_i$ , in the cell:

$$r_{t,\text{tot}} = \sum_i r_{ti} = \sum_i r_i \frac{t_i/\kappa_t}{1 + t_{ai}/\kappa_{ta} + t_i/\kappa_t}, \quad (2.69)$$

which can be rewritten as

$$r_{t,\text{tot}} = \sum_i r_i \frac{t_i/\kappa_t}{1 + (0.5r - t_i)/\kappa_{ta} + t_i/\kappa_t}, \quad (2.70)$$

using the assumption that  $t_{tot,i} = t_{ai} + t_i = 0.5 \cdot r$ .  $r_i$  denotes the concentration of ribosomes recognizing amino acid  $i$ . Finally, with  $r = \sum_i r_i$  the total ribosome concentration and  $a_i$  the concentration of amino acid  $i$ , the dynamics of the charged tRNA concentration is described by

$$\frac{dt_{ai}}{dt} = v_{tai}(a_i, t_i) - f_i \cdot v_{\text{ribosome}}(t_i, r), \quad (2.71)$$

with  $v_{tai}(a_i, t_i)$  the synthesis rate of charged tRNA, and  $f_i \cdot v_{\text{ribosome}}(t_i, r)$  their consumption via protein synthesis. In particular,

$$v_{tai}(a_i, t_i) = k_{Si} \cdot S_{tot,i} \cdot \frac{t_i a_i}{t_i K_{Mai} + a_i K_{Mti} + t_i a_i}, \quad (2.72)$$

$$v_{\text{ribosome}}(t_i, r) = k_{rib} \cdot r \cdot \left( 1 + \sum_i \left[ f_i \cdot \left( 1 + \frac{t_i}{\kappa_t} \right) \frac{\kappa_{ta}}{0.5 \cdot r - t_i} \right] \right)^{-1} \quad (2.73)$$



For comparison with our framework, we need ppGpp as a direct function of the total amino acid concentration  $a = \sum_i a_i$  (a proxy for precursors) and total ribosome concentration  $r$  (a proxy for gene expression machinery). To this end, we made two additional assumptions:

- (A1) All concentrations specific to one type of amino acid  $i$  ( $a_i$ ,  $t_{ai}$ ,  $t_i$ ,  $r_i$ ) are in the same proportion  $f_i = f = 1/20$  with respect to the total concentrations ( $a$ ,  $t_a$ ,  $t$ ,  $r$ ).
- (A2) We apply a quasi-steady-state approximation (QSSA) to the dynamics of the concentration of the charged tRNAs ( $t_{ai}$ ) and the concentration of ppGpp (ppGpp). That is, the dynamics of these variables are assumed fast relative to the dynamics of the amino acid concentrations ( $a_i$ ) and the total ribosome concentration ( $r$ ).

Using (A2), we can rewrite Eq. 2.71 as follow:

$$v_{tai}(a_i, t_i) = f_i \cdot v_{ribosome}(t_i, r), \quad (2.74)$$

which, using (A1) and Eqs 2.72 and 2.73, leads to:

$$k_{Si} \cdot S_{tot,i} \cdot \frac{t_i a_i}{t_i K_{Mai} + a_i K_{Mti} + t_i a_i} = f_i \cdot k_{rib} \cdot r \cdot \left(1 + \frac{\kappa_{ta}}{0.5r - t_i} \cdot \left(1 + \frac{t_i}{\kappa_t}\right)\right)^{-1}. \quad (2.75)$$

By rearranging both sides of the equation,  $t_i$  can be expressed as a function of  $a_i$  and  $r$ , which yields:

$$\begin{aligned} At_i^2 + Bt_i + C &= 0, \quad \text{with} \\ A &= \frac{k_{Si} S_{tot,i} a_i}{f_i k_{rib} r} \left(\frac{\kappa_{ta}}{\kappa_t} - 1\right) + K_{Mai} + a_i, \\ B &= \frac{k_{Si} S_{tot,i} a_i}{f_i k_{rib} r} (0.5r + \kappa_{ta}) + a_i K_{Mti} - 0.5r (K_{Mai} + a_i), \\ C &= -0.5r a_i K_{Mti}, \end{aligned} \quad (2.76)$$

and therefore

$$t_i(a_i, r) = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}.$$

It is not difficult to show that the only solution on  $[0, 0.5r]$  is

$$t_i(a_i, r) = \frac{-B + \sqrt{B^2 - 4AC}}{2A}. \quad (2.77)$$

From this result, we obtain  $r_{t,tot}$  as a function of  $a_i$  and  $r$ , by applying (A1) to Eq. 2.70:

$$r_{t,tot}(t_i, r) = r \cdot \frac{t_i/\kappa_t}{1 + (0.5r - t_i)/\kappa_{ta} + t_i/\kappa_t}, \quad (2.78)$$

and substituting  $t_i$  by the expression of Eq. 2.77.

Finally, we apply (A2) to Eq. 2.67 and obtain the final expression giving the concentration of ppGpp as a function of the total amino acid and ribosome concentrations:

$$\text{ppGpp}(a_i, r) = \frac{1}{k_{spoT}} \left( k_{RelA} \cdot RelA_{tot} \cdot \frac{r_{t,tot}(a_i, r)}{K_{D,RelA} + r_{t,tot}(a_i, r)} + v_{spoT} \right). \quad (2.79)$$

This function is represented in Fig. 2.11 with parameters taken from Table 2.1.

The plotted surface of the function resembles the inverse of the on-off control strategy in Fig. 2.8, as expected, bearing in mind that ppGpp has an inhibitory effect on the synthesis of ribosomal RNA. We assumed a Michaelis-Menten inhibition for the regulatory effect of ppGpp on rRNA synthesis, and thus indirectly on the synthesis of ribosomal proteins [127, 138]:

$$\alpha(\text{ppGpp}) = \frac{K_I}{K_I + \text{ppGpp}}. \quad (2.80)$$

The inhibitory constant  $K_I$  lies in the dynamical range of variation of ppGpp. In Fig. 2.8 in the main text, we took  $K_I = 10 \mu\text{M}$ .

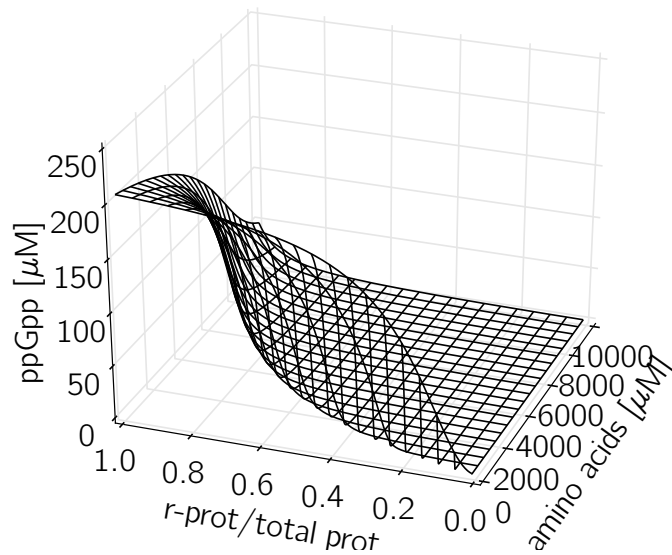


Figure 2.11 – **ppGpp concentration is a function of total ribosome and amino acid concentrations.** We assume the dynamics of ppGpp to be fast on the time-scale of changes in the ribosome and amino acid concentrations. The concentration of ppGpp can thus be expressed as a function of the latter two variables, using the model of Bosdriesz *et al.* [116]. Parameters are taken from Table 2.1.

Symbol	Value	Unit	Description
$a_i$	–	$\mu\text{M}$	Concentration of aa $i$ (not incorporated in protein)
$t_{ai}$	–	$\mu\text{M}$	Concentration of tRNA charged with aa $i$
$t_i$	–	$\mu\text{M}$	Concentration of free tRNA conjugate to aa $i$
$t_{tot,i}$	$0.5 \cdot r$	$\mu\text{M}$	Total concentration of tRNA conjugate to aa $i$
$r_i$	–	$\mu\text{M}$	Total concentration of ribosome with an A-site for aa $i$
$r_{ti}$	–	$\mu\text{M}$	Ribosomes with uncharged tRNA in an A-site for aa $i$
ppGpp	–	$\mu\text{M}$	Concentration of ppGpp
$a$	$\sum_i a_i$	$\mu\text{M}$	Total concentration of aa (not incorporated in protein)
$t_a$	$\sum_i t_{ai}$	$\mu\text{M}$	Total concentration of tRNA charged with aa
$t$	$\sum_i t_i$	$\mu\text{M}$	Total concentration of free tRNA
$r_{t,tot}$	$\sum_i r_{ti}$	$\mu\text{M}$	Total concentration of uncharged tRNA bound to ribosomes
$r$	$\sum_i r_i$	$\mu\text{M}$	Total concentration of ribosomes
$v_{RelA}$	–	$\mu\text{M}/\text{s}$	Rate of RelA-catalyzed ppGpp synthesis
$v_{SpoT}$	$10^{-3}$	$\mu\text{M}/\text{s}$	Rate of ppGpp synthesis by SpoT
$v_{tai}$	–	$\mu\text{M}/\text{s}$	Rate of amino-acyl tRNA $i$ synthetase
$v_{ribosome}$	–	$\mu\text{M}/\text{s}$	Total rate of protein synthesis
$k_{rib}$	20	$\text{s}^{-1}$	$k_{cat}$ of protein elongation
$k_{RelA}$	75	$\text{s}^{-1}$	$k_{cat}$ of ppGpp synthesis by RelA
$K_{D,RelA}$	0.26	$\mu\text{M}$	Michaelis constant of RelA-catalyzed ppGpp production
$RelA_{tot}$	1/15	$\mu\text{M}$	RelA concentration
$k_{SpoT}$	$\ln(2)/30$	$\text{s}^{-1}$	Rate of ppGpp degradation by SpoT
$\kappa_t$	500	$\mu\text{M}$	Dissociation constant of uncharged tRNA-ribosome complex
$\kappa_{ta}$	1	$\mu\text{M}$	Dissociation constant of charged tRNA-ribosome complex
$k_{Si}$	100	$\text{s}^{-1}$	$k_{cat}$ of aminoacyl-tRNA synthetase
$S_{tot,i}$	1	$\mu\text{M}$	Total concentration of aminoacyl-tRNA synthetase for aa $i$
$K_{Mai}$	100	$\mu\text{M}$	Michaelis constant of aa-tRNA synthetase for amino acids
$K_{Mti}$	1	$\mu\text{M}$	Michaelis constant of aa-tRNA synthetase for uncharged tRNA
$f_i$	1/20	–	Proportion of aa $i$ in proteins (Assumption A1)

Table 2.1 – Parameters and variables reused from Bosdriesz *et al.* [116]. The abbreviation aa denotes amino acids.

### 2.5.5 S1 Table – Parameter values of the self-replicator model

Parameter	Unit	Literature value	Fitted value
$\gamma$	–	No value	1.39
$k_R$	$\text{h}^{-1}$	3.6	2.23
$e_M$ for M63+glycerol	$\text{h}^{-1}$	< 3.6	0.587
$e_M$ for M63+glucose	$\text{h}^{-1}$	< 3.6	0.867
$e_M$ for cAA+glycerol	$\text{h}^{-1}$	< 3.6	1.07
$e_M$ for cAA+glucose	$\text{h}^{-1}$	< 3.6	1.57
$e_M$ for RDM+glycerol	$\text{h}^{-1}$	< 3.6	3.48
$e_M$ for RDM+glucose	$\text{h}^{-1}$	< 3.6	4.76
$\beta K_R$	–	0.003	Not fitted

Table 2.2 – **Parameter values of self-replicator model** The parameter values in the model were obtained by fitting Eq. 2.25 to the data of Scott et al [33] (Fig. 2.3 in the main text), as described in the *Methods* section. They are compared with order-of-magnitude estimates from the literature (S2 Text).

### 2.5.6 S1 Figure – Simple control strategies for the self-replicator of bacterial growth

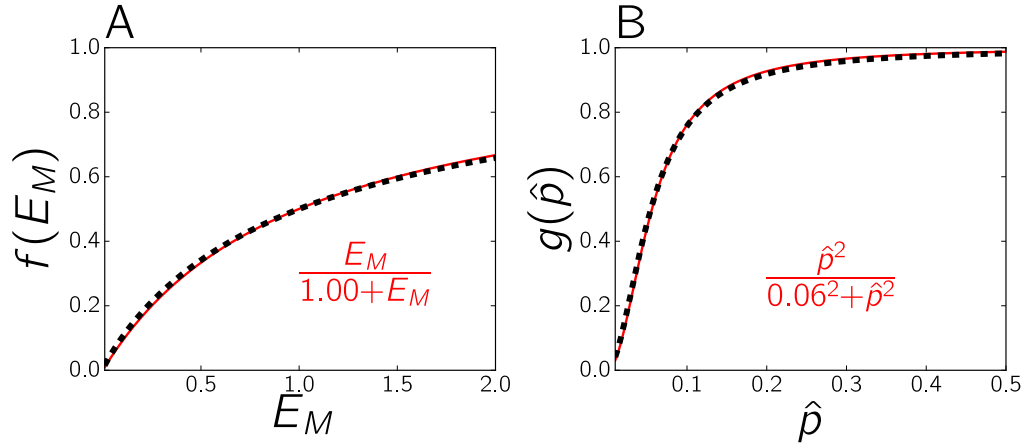


Figure 2.12 – **Simple control strategies for the self-replicator of bacterial growth.** *A*: Nutrient-only strategy:  $\alpha = f(E_M)$ . The dashed, black curve is the (unique) strategy driving the system exactly to the optimal steady state, that is, the state in which growth occurs at the maximal rate supported by  $E_M$ . The function  $f$  is defined by Eq. 2.27 in the *Methods* section of the main text. The solid, red curve is an approximation of this function by the simple Michaelis-Menten curve of Eq. 2.16, with  $K_{mE} = 1.0$ . *B*: Precursor-only strategy:  $\alpha = g(\hat{p})$ . The dashed, black curve is the (unique) strategy driving the system exactly to the optimal steady state. The function  $g$  is defined by Eq. 2.28 in the *Methods* section of the main text. The solid, red curve is an approximation of this function by the simple Hill curve of Eq. 2.18, with  $K_{mp} = 0.06$  and a cooperativity coefficient 2.



## Chapter 3

# Dynamics of resource allocation during an acetate-glucose upshift

*"Oh you like beetles? Wonderful! Then that's what you'll be killing for the next 50 years." – Zach Weinersmith, *Biologists are weird* [176].*

---

### Résumé du Chapitre 3: Mesures dynamiques de l'allocation des ressources lors d'une transition acétate vers glucose

Dans ce chapitre, nous nous intéressons à la validation expérimentale des prédictions faites au cours du Chapitre 2. Nous avons en effet pu voir que, si l'on suppose qu'un microorganisme optimise sa production de biomasse lors d'une transition de croissance, on peut s'attendre à ce que ce dernier distribue ses ressources cellulaires entre la production de machinerie d'expression génique et de machinerie métabolique en suivant un schéma tout-ou-rien (*bang-bang*). Le but est donc ici de mesurer précisément la dynamique de  $\alpha$  lors d'une transition de croissance contrôlée.

La Section 3.2.1 s'attache à définir les conditions expérimentales dans lesquelles le comportement prédit lors du Chapitre 2 pourra être observé. En utilisant le modèle d'auto-réplicateur (Eqs 3.1 et 3.2), nous montrons que l'allocation des ressources  $\alpha(\cdot)$  peut être reconstruite au cours de la transition si l'on dispose de mesures dynamiques du taux de croissance et de la concentration en machinerie d'expression génique. Cependant, le comportement tout-ou-rien a une dynamique bien particulière qui peut facilement être



offusquée au niveau d'une population de cellules si ces dernières ne sont pas synchrones. De fait, cette question ne peut être résolue qu'en obtenant des mesures à haute fréquence et au niveau de cellules individuelles, ce qui nous a conduits à utiliser des rapporteurs fluorescents de l'abondance d'une des machineries. En particulier, nous construisons une souche d'*Escherichia coli* chez laquelle la sous-unité S2 du ribosome a été fusionnée avec une protéine fluorescente (GFP) (voir Section 3.4.1 des matériels et méthodes). Couplée à de la vidéomicroscopie, cette souche nous permet de quantifier en temps réel la quantité de ribosomes au niveau de cellules uniques, ce qui procure un excellent moyen d'obtenir la concentration en machinerie d'expression génique (le ribosome étant son composant principal). Le taux de croissance quant à lui, peut être facilement reconstruit en mesurant l'évolution de la taille des cellules dans le temps.

Nous définissons également dans la Section 3.2.1 les conditions de croissance adéquates. Nous nous focalisons sur le cas particulier d'une transition d'un milieu de croissance pauvre vers un milieu riche. Il est important que les bactéries soient en phase de croissance dans chaque milieu du fait de l'existence d'une physiologie bien particulière propre à la non-croissance. Les cellules doivent donc croître pendant suffisamment longtemps sur le premier milieu, sans épuiser celui-ci, puis une transition rapide doit être effectuée vers le milieu riche. Nous avons utilisé pour cela la *Mother Machine*, un dispositif microfluidique à la base mis au point pour maintenir et observer les cellules bactériennes en état de croissance stationnaire pendant de longues périodes. Au sein de ce dispositif, les cellules sont en contact avec du milieu frais qui circule et leur procure de nouveaux nutriments durant toute la durée de l'expérience (voir Section 3.4.3). En changeant simplement le type de milieu injecté dans le dispositif, nous pouvons l'utiliser pour modifier en quelques secondes les nutriments auxquels les bactéries ont accès, et donc réaliser la transition de manière robuste et contrôlée. Plus particulièrement, le milieu pauvre utilisé contient de l'acétate comme seule source de carbone, tandis que le milieu riche contient essentiellement du glucose (Fig. 3.1).

De telles expériences s'avèrent malheureusement pointues et, au cours du déroulement de cette thèse, nous n'avons pu en réaliser qu'un petit nombre, dont seule celle présentée ici s'est avérée exploitable. Celle-ci a cependant subi de nombreux problèmes expérimentaux qui, s'ils n'ont pas empêché complètement son exploitation, ont posé des problèmes pour l'analyse et l'interprétation des résultats. Ces problèmes expérimentaux sont abordés dans la Section 3.2.2, et sont largement discutés en Section 3.3. En effet, suite à un problème logiciel et à ce qui semble être une contamination par un bactériophage, nous n'avons pas pu obtenir des séries temporelles aussi longues qu'initialement prévu. Même si nous disposons néanmoins des données se si-

tuant autour de la transition, le fait de ne pas avoir atteint un nouvel état de croissance stationnaire sur glucose va s'avérer être un inconvénient majeur pour l'analyse des résultats. De plus, nous n'avons pu analyser qu'une partie des données, en nous focalisant sur une seule cellule par canal de la *Mother Machine* (Fig. 3.2). Au final, il faut retenir de ces sections que cette expérience et son analyse sont largement préliminaires, et seront consolidées dans le futur par de nouveaux travaux.

La Section 3.2.3 présente la façon dont nous pouvons reconstruire  $\alpha(\cdot)$  lors de la transition à partir des données obtenues. Pour cela, nous développons une méthode basée sur le lissage de Kalman, qui utilise une approche probabiliste et bayésienne pour reconstruire de manière robuste un signal à partir de mesures bruitées. Nous montrons comment l'information dont nous disposons sur le déroulement de l'expérience peut être injectée dans l'analyse sous la forme d'un *prior* bayésien afin de faciliter la reconstruction, ce qui s'avère particulièrement utile sur nos données. De plus, à partir de données synthétiques générées grâce au modèle développé lors du Chapitre 2, nous constatons que la méthode est capable de capturer le profil oscillatoire généré par le comportement tout-ou-rien que nous attendons (Fig. 3.5).

Finalement, nous appliquons cette méthode à la reconstruction du taux de croissance puis de  $\alpha(\cdot)$ , et présentons les résultats dans la Section 3.2.4. En moyenne, le taux de croissance présente le comportement attendu (Figs 3.3 et 3.4), avec des valeurs à l'état de croissance stationnaire qui sont proches de celles connues pour *E. coli*, et une augmentation rapide au moment de la transition vers le milieu riche. Les résultats pour la reconstruction de  $\alpha(\cdot)$  sont cependant plus mitigés. Même si un comportement oscillatoire est observé pour chaque cellule (Figs 3.6 et 3.8), il n'est pas certain que ce dernier ne soit pas un artefact dû à une mauvaise calibration de la méthode. Il aurait fallu en effet obtenir de longs états stationnaires de croissance avant et après la transition pour pouvoir réaliser cette calibration dans de bonnes conditions. Néanmoins, si l'on s'intéresse au comportement médian des cellules (Fig. 3.7), on constate que la première oscillation survenant après la transition semble conservée dans la population (Fig. 3.8). Ces résultats s'avèrent finalement encourageants, et incitent à pousser plus en avant cette étude, notamment via l'acquisition de nouvelles données pour lesquelles les problèmes cités auront pu être corrigés (voir discussion en Section 3.3).

---

## Beginning of Chapter 3

### 3.1 Introduction

Most studies of the growth of microorganisms have been done at steady state. This is a reasonable and logical choice given that at steady state, the behavior of microorganisms is reproducible, which helps uncovering simple theories to apprehend their inherent complexity. But while this condition can be easily achieved in the laboratory, microorganisms naturally spend very little time in steady state. This motivated the construction of a new theoretical framework in Chapter 2 to study growth laws during growth transitions.

The principles underlying regulatory processes in dynamical conditions appear to be different from those applying at steady state. We showed that regulatory strategies approaching the theoretical optimal, given by bang-bang control of gene expression machinery, have to be able to effect abrupt and strong variations in the activity of a gene. It was seen that for implementing such strategies the regulatory system needs to be capable of sensing the internal state of the cell and not just the environment. A near-optimal transition also requires information of several different variables, whereas for maintaining a steady state with maximal biomass accumulation a single variable turned out to be sufficient in our theoretical framework. Interestingly, we showed that the ppGpp regulatory system of *E. coli* fulfills the requirement of such a regulatory strategy, by implementing an on-off strategy for regulating the synthesis of ribosomes. However, although this provides circumstantial evidence that bacteria control resource allocation in a manner consistent with theoretical optimality, experimental data are necessary to decide if an on-off strategy is at work in *E. coli* during growth transitions.

Unfortunately, far more information is available on ribosome abundance at steady state than during growth transitions [33, 152]. The main reason for this bias on steady-state conditions is that, from an experimental point of view, growth transitions are hard to control and may depend on the history of the culture [39–43]. However, as we discussed in Section 2.3, the few studies that have been reported seem to be consistent with our predictions, in the sense that they indicate that during growth transitions, the synthesis rate of ribosomes oscillates [152, 153] and the ppGpp concentration manifests a rapid succession of increases and decreases [139, 154]. But we cannot decisively validate or disprove the model predictions from measurements that were carried out at the population level, where it is inherently hard to identify switching patterns. What is needed for the verification of our predictions are dynamical single-cell measurements of the ribosome concentration during

well-controlled growth transitions. While the observation of an on-off strategy would show that simple models of the type introduced in Chapter 2 are instrumental in gaining a better understanding of microbial physiology, its refutation would also raise interesting questions. If maximization of growth rate and biomass accumulation have been retained by natural selection in *E. coli*, which factors would prevent the cell from behaving optimally from a theoretical point of view?

The aim of this chapter is to measure *in vivo*, during a growth transition, the resource allocation profile  $\alpha(\cdot)$  and compare its dynamics with the gold standard established in Chapter 2. To this end, we have measured ribosomal abundance of *E. coli* at the single-cell level during a nutrient upshift. More precisely, we constructed a strain in which a fluorescent marker has been attached to a ribosomal subunit, thus allowing the *in-vivo* monitoring of (changes in) the abundance of ribosomes. In collaboration with Irina Mihalcescu of the Laboratoire Interdisciplinaire de Physique, we cultivated this *E. coli* strain in a microfluidic device, allowing the long-term imaging of individual cells in well-controlled conditions, notably involving a classic upshift experiment from growth in minimal medium with acetate to minimal medium with glucose. We then developed a Kalman smoothing method, in collaboration with Eugenio Cinquemani of Inria Grenoble – Rhône-Alpes, to reconstruct  $\alpha(\cdot)$  from the estimates of the variations of the growth rate and the relative ribosomal synthesis rate. While the preliminary results presented here do not allow a decisive validation of the expected behavior, among other things due to the difficulties that were encountered during the experiments and the image analysis, we believe they are promising as a first step towards the better understanding of global resource allocation during growth transitions.

For clarity, the model of resource allocation during growth transitions developed in Chapter 2 is reproduced here, where dotted variables represent time derivatives:

$$\dot{p}(t) = e_M(t) \cdot (1/\beta - r(t)) - \frac{k_R \cdot p(t)}{K_R + p(t)} \cdot r(t) (1 + \beta p(t)), \quad (3.1)$$

$$\dot{r}(t) = \frac{k_R \cdot p(t)}{K_R + p(t)} \cdot r(t) (\alpha(t) - \beta r(t)). \quad (3.2)$$

In this form, it contains 4 variables ( $e_M(t)$ ,  $p(t)$ ,  $r(t)$ ,  $\alpha(t)$ ) and 3 parameters ( $\beta$ ,  $k_R$ ,  $K_R$ ).  $e_M(t)$  [ $\text{min}^{-1}$ ] is an indicator of the richness of the environment.  $p(t)$  [ $\text{g.L}^{-1}$ ] is the precursor concentration inside the cell.  $r(t)$  [ $\text{g.L}^{-1}$ ] is the concentration of gene expression machinery inside the cell.  $\alpha(t)$  [ $\emptyset$ ] is the fraction of resources allocated to the synthesis of gene expression machinery.  $k_R$  [ $\text{min}^{-1}$ ] is the rate constant of macromolecular synthesis.  $K_R$  [ $\text{g.L}^{-1}$ ] is the

half-saturation constant of macromolecular synthesis.  $\beta$  [L.g<sup>-1</sup>] is the inverse of the cellular density of macromolecules, assumed to be constant.

## 3.2 Results

### 3.2.1 Experimental design

How can one measure the resource allocation variable in *E. coli* cells? There is no direct way to quantify  $\alpha(\cdot)$  in an experiment, mostly because of its abstract nature. To correctly reconstruct it, one has to know the value of every single term in the equations in which it appears. In the model of Eqs 3.1-3.2,  $\alpha$  only appears in Eq. 3.2. We can therefore identify  $\alpha(t_k)$  for each time  $t_k$  when the terms  $\beta r(t_k)$ ,  $\dot{r}(t_k)$ , and  $\frac{k_R p(t_k)}{K_R + p(t_k)} \cdot r(t_k)$  are known, or equivalently, the six individual components  $r(t_k)$ ,  $\dot{r}(t_k)$ ,  $p(t_k)$ ,  $k_R$ ,  $K_R$ ,  $\beta$  they are composed of. Even when assuming that the parameters  $k_R$ ,  $K_R$ , and  $\beta$  are known, or at least easy to estimate independently, reconstructing the dynamics of  $\alpha$  would require estimation of the (changes in) concentration of the gene expression machinery ( $r$ ,  $\dot{r}$ ) and the precursor concentration  $p$ .

Are such measurements feasible in practice? The abundance of ribosomes can be quantified, for instance, by measuring the total RNA of the cell [33], or using radioactive markers [152, 153], fluorescent labels [81], or mass spectrometry techniques [177]. However, only the use of fluorescent proteins complies with our need for real-time single-cell quantification. Obtaining an estimate of precursor abundance is even more challenging, since there is no clear proxy for the totality of precursors in the cell and, while absolute quantification of all internal metabolites has been achieved in *E. coli* [60], this method is also not suitable for a dynamical estimation in individual cells.

The problem of quantifying  $\alpha(\cdot)$  can be simplified by applying a transformation of the model of Eqs 3.1-3.2. Taking into account that by construction the growth rate  $\mu(t)$  is given by

$$\mu(t) = \beta \frac{k_R \cdot p(t)}{K_R + p(t)} \cdot r(t),$$

we can rewrite Eq. 3.2 as follows

$$\dot{r}(t) = \mu(t) \left( \frac{\alpha(t)}{\beta} - r(t) \right). \quad (3.3)$$

The problem of estimating  $\alpha(\cdot)$  is thus equivalent to the problem of estimating  $r(\cdot)$ ,  $\dot{r}(\cdot)$ ,  $\mu(\cdot)$  and  $\beta$ . If we are satisfied with estimating  $\alpha$  up to a constant

factor  $\beta$ , dynamically measuring the ribosome concentration and the growth rate would make  $\alpha(\cdot)/\beta$  identifiable.

Given that fluorescent reporters of ribosomal proteins can provide information on both  $r(\cdot)$  and  $\dot{r}(\cdot)$  [178], the above reformulation of the problem is promising for the purposes of our study. Moreover, since the predicted bang-bang profiles in gene activity are easily hidden at the population level if cells are not synchronous, which is generally the case, expression of the ribosomal genes need to be monitored at high sampling density on the single-cell level, which is also possible with fluorescent reporters. Inspired by previous work [81], we therefore constructed an *E. coli* strain in which the S2 ribosomal subunit, encoded by the gene *rpsB*, has been tagged with a green fluorescent protein (GFP). The strain with the chromosomal *rpsB-gfp* fusion produces fluorescent ribosomes that are quantifiable *in vivo* in single cells while not affecting the growth rate (see Material and Methods 3.4.1 and [81]). Monitoring single cells of this strain, by time-lapse fluorescence microscopy, thus makes it possible to estimate  $r(\cdot)$  and  $\dot{r}(\cdot)$  and reconstruct  $\mu(\cdot)$ .

The growth conditions also need to be carefully chosen for this study. In natural environments, for many bacteria one of the most frequently encountered growth transitions, or at least the transition on which selection is expected to operate, is the transition from stationary phase in a nutrient-deprived medium [179] to (exponential) growth after the renewed availability of nutrients. The transition from stationary to exponential phase is difficult to study though as many hard-to-control parameters play a role, including the duration of the nutrient stress, the composition of the medium before growth arrest, and the accumulation of waste products in the medium [30]. As a consequence, growth resumption only occurs after a lag phase of variable duration [39–43], which is not included in the model. For this reason, we decided to focus on a transition from exponential growth on a poor carbon source (acetate) to exponential growth on a rich carbon source (glucose) in a well-defined minimal medium. By construction, this steady-state-to-steady-state transition requires a long acquisition time before and after the shift, respectively, to ensure that the cells have no memory of their physiological state before the transition and have enough time to reach the new steady state after the transition. The use of the mother machine, a microfluidic device that was designed to sustain exponential growth for long periods of time, is therefore a good choice for this purpose. It allows to maintain a constant flow of fresh medium and to effect fast transitions by switching the medium (see Material and Methods 3.4.2 and 3.4.3). The complete experimental plan is summarized in Fig. 3.1.

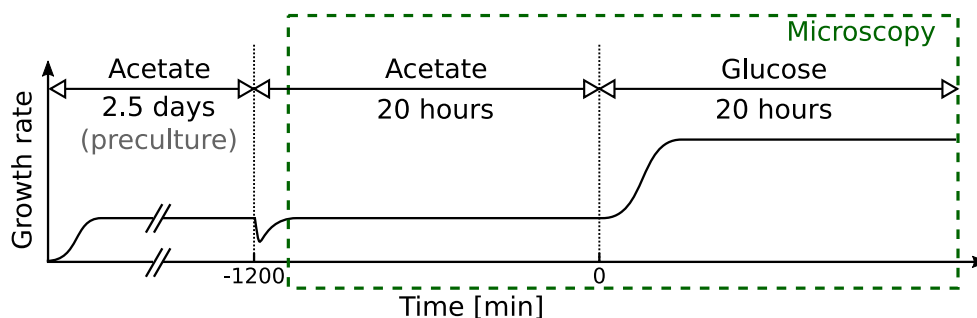


Figure 3.1 – **Schematic outline of the upshift experiment.** The goal is to measure the fluorescence and length/area of *E. coli rpsB-gfp* cells during an acetate-to-glucose upshift. We use M9 minimal medium supplemented with 0.2% of acetate or glucose (see Material and Methods 3.4.2). A preculture was started from glycerol stock for 2.5 days on 0.2% acetate in batch condition (shake flask). The day of the experiment, the cells were injected into the mothermachine and fed by a constant flow of fresh 0.2% acetate medium (see Material and Methods 3.4.3). Fluorescence and phase contrast images were taken every 5 minutes. After 20 h, the feeding media was switched to 0.2% glucose and maintained for 20 h while continuing image acquisition. Time 0 corresponds to the moment of the acetate-glucose upshift.

### 3.2.2 Data acquisition

An experiment following the plan of Fig. 3.1 was carried out, over a period of 40 h, but encountered a number of difficulties. First, the motor displacing the microscope along the Z-axis intermittently stalled, thus deactivating the autofocus and the Z-compensation for the tilt of the microfluidic device in the XY plane. As a consequence, until the experimenter manually intervened, many of the images acquired were out of focus. This is the reason why data points between -720 and -150 min are not exploitable, but fortunately the experiment was planned in such a way that when it really mattered, notably around the growth transition, someone was present to monitor the microscope. Second, at around 7 h after the nutrient upshift, the bacteria started to die for reasons that are unknown, the most plausible hypothesis being a phage contamination (see Discussion 3.3). Data analysis was therefore limited to 550 min after the upshift, when roughly 3/4 of the cells were still growing. For these and other reasons (see Section 3.3), the experiment will have to be repeated, but there was no time left for this in the framework of my PhD thesis.

The acquired data consisted of phase contrast and fluorescence images of 60 wells in total, in four different fields. While quite a few image analysis



programs have been reported in the literature [155, 180–182], and some of these specifically address the mother machine design of the microfluidic device [45, 183], they all presented limitations when applied to our data. For instance, the segmentation algorithms experienced difficulties due to the low resolution of the camera, and the phase contrast images had a superposed reflection band due to the microfluidic device. Moreover, the fluorescence density was concentrated in hot spots (Fig. 3.2) and its intensity varied during the experiment. While this was expected, since ribosomes are localized in the cell poles [81] and ribosomal abundance is known to vary with the growth rate [33], it complicated automatic segmentation on the fluorescence images.

For these reasons, the image analysis carried out for this chapter has been much simplified. First, we have focused on the cell at the bottom of each well, since this avoids the tracking of individual cells across several generations and ensures that the descendance of this cell can be followed throughout the experiment. Second, segmentation was done by manually selecting two pixels, one at each pole of the cell. These pixels were used to create a rectangle surrounding the cell (Fig. 3.2 and Material and Methods 3.4.4). After background correction (Material and Methods 3.4.4), the fluorescence intensity in units RFU was evaluated for each pixel in the rectangle, as well as the length of the cell, defined by the distance in pixels between the poles (Fig. 3.2). The fluorescence density for the entire cell [RFU/pixel/cell] was computed by dividing the sum of the fluorescence intensities of the pixels in the rectangle by the total number of pixels in the rectangle. Although we are well aware that the above procedure can be improved on many counts (Section 3.3), we nevertheless believe that it provides a reasonable approximation of the quantities of interest and a valid starting-point for the estimation of the growth rate and the resource allocation profile in the remainder of this chapter.

In total, we obtained time courses of fluorescence density [RFU/pixel/cell] and the length for 68 bacterial cells (Fig. 3.2). The fluorescence density appears constant during growth on acetate (before time 0) and immediately increases when the carbon source is switched to glucose. Inspection of the cellular length shows that the division frequency abruptly increases after the nutrient upshift, corresponding to a higher growth rate. Unfortunately, while steady-state growth on acetate was reached before the upshift, the fluorescence density profile suggests that the experiment did not continue long enough to ensure that a new steady state on glucose was reached. However, the data before and after the nutrient upshift are exploitable. What do we observe if we estimate the growth rate and the resource allocation profile from these data?



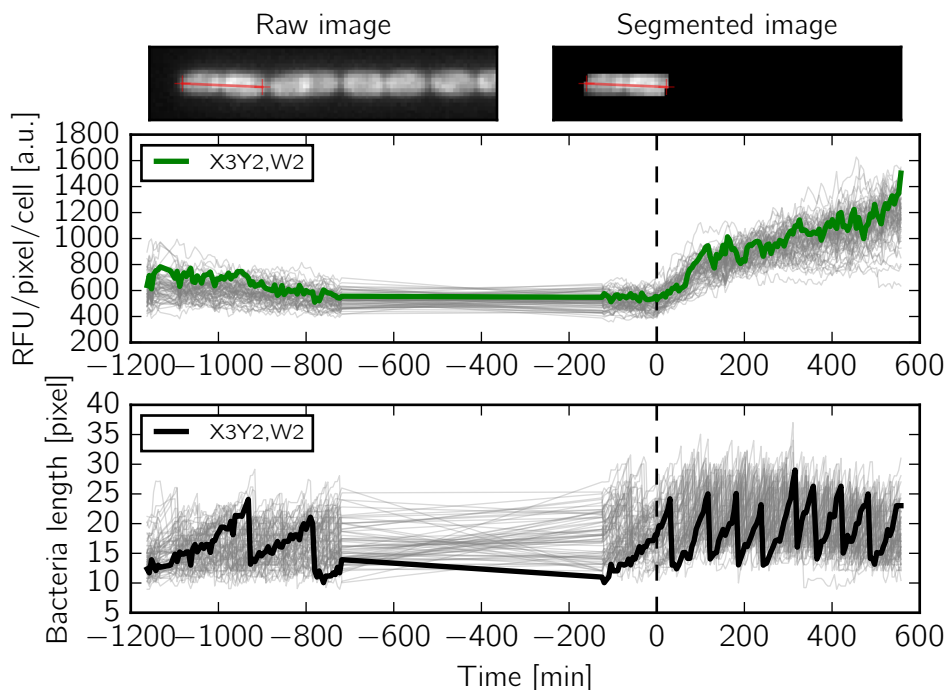


Figure 3.2 – **Results of data acquisition.** We imaged 6 fields (X[1-3]Y[1-2]) each containing 15 wells (W[0-14]) for  $\sim 40$  h. From the 90 wells, 68 were suitable for further analysis (the others being empty, out of frame for some period of time, or plugged). For all wells, data points are missing in the interval between  $[-720, -150]$  because the camera was out of focus. We stopped the analysis at 550 min, when about 3/4 of the bacteria were still growing, before the entire population died within a few hours for an unknown reason. The image labeled "Raw image" is the last image analyzed for the highlighted well. The bacterium on the left of this image was manually segmented by selecting two pixels at the poles on the fluorescence images (red cross). A 6-pixel-wide rectangular mask was computed for each image, resulting in the "Segmented image" on the right. Fluorescence intensities are expressed in Relative Fluorescence Units (RFU) on a 16-bit image and were corrected for camera background, but not autofluorescence background (Material and Methods 3.4.4 and Discussion 3.3). The fluorescence intensity of the cell, expressed in units RFU/pixel/cell, was computed by dividing the sum of the fluorescence intensities of the pixels in the rectangle by the total number of pixels in the rectangle. The cell length is the distance in pixels between the two poles (red line). The thick lines in green and black highlight the time-varying length and fluorescence density for the cell visible in the top images, labeled "X3Y2, W2". The vertical dashed lines represent the time of the upshift from growth on acetate to growth on glucose.

### 3.2.3 Data analysis using Kalman smoothing

As we presented in section 3.2.1, our goal is to reconstruct the signal  $\alpha(\cdot)$  during a growth transition. By modeling *E. coli* as a cylinder, the volume can be assumed directly proportional to the length of the cell. If we assume a negligible background of autofluorescence (see Material and Methods 3.4.4), the fluorescence concentration in the cell can be assumed proportional to the ribosome concentration. More precisely, based on the available data, we have the following measurement model at each time-point  $t_k$ ,  $0 \leq k \leq N - 1$ :

$$L(t_k) = \lambda \cdot V(t_k) + \epsilon_k, \quad (3.4)$$

$$F(t_k) = \gamma \cdot r(t_k) + \eta_k, \quad (3.5)$$

where  $L(t_k)$  and  $F(t_k)$  are the length and fluorescence density in Relative Fluorescence Units per pixels measured at time  $t_k$ , respectively, and  $V(t_k)$  and  $r(t_k)$  the corresponding actual volume and ribosome concentration at time  $t_k$ , respectively.  $(\lambda, \gamma)$  are unknown proportionality constants, and  $(\epsilon_k, \eta_k)$  uncorrelated sequence of measurement noise assumed normally distributed with mean 0.

As explained in Section 3.2.1, reconstructing  $\alpha(\cdot)$  requires information on  $\mu(\cdot)$ ,  $r(\cdot)$  and  $\dot{r}(\cdot)$ . From Eq. 3.5, we can obtain  $r(\cdot)$  and  $\dot{r}(\cdot)$  by smoothing interpolation and differentiation. Similarly,  $\mu(\cdot)$  can be derived from Eq. 3.4 as it is defined by

$$\dot{V}(t) = \mu(t) \cdot V(t).$$

However, smoothing interpolation is particularly sensible to the boundary conditions. Since each division in the time series generates a new boundary condition, smoothing interpolation of our data is expected to be little robust. For this reason, the growth rate of single cells is usually estimated by fitting an exponential function to the volume data between each cell [45, 161]. While this is suitable when bacteria are at steady state, it is not applicable during a growth transition, where we expect the growth rate to vary between successive divisions. Other techniques are less sensible to the above problem [178], and we decided to use Kalman smoothing for our purpose [118, 119, 184].

Kalman filtering [185] is a Bayesian algorithm using a series of noisy measurements to predict the state of a dynamical system. It has been extensively used in engineering applications (guidance, tracking, control, ...) requiring the real-time estimation of hidden variables in a dynamical system from present and past measurements. Kalman smoothing is an extension of Kalman filtering using information about past and present but also future measurements of the state of the system, and is widely applied to estimate unknown inputs in time series analysis [118, 119, 184]. The advantage of

Kalman filtering with respect to Kalman smoothing is that it uses all available information to estimate the hidden variables of a dynamical system, in applications where a real-time response is not required. For our problem, at each time step  $t_k$ , our hidden variables are the resource allocation variable  $\alpha(t_k)$  and the growth rate  $\mu(t_k)$ , while the available information consists of the  $N$  measurements  $\{F(t_0), F(t_1), \dots, F(t_{N-1})\}$  and  $\{L(t_0), L(t_1), \dots, L(t_{N-1})\}$  that were taken during the experiment.

The Kalman filtering problem can now be formulated as follows. The state of a growing bacterial cell is described by the dynamical system

$$\dot{r}(t) = \mu(t) \cdot \frac{\alpha(t)}{\beta} - \mu(t) \cdot r(t), \quad (3.6)$$

$$\dot{V}(t) = \mu(t) \cdot V(t), \quad (3.7)$$

with initial conditions  $r(0) = r_0$ ,  $V(0) = V_0$ , and the following measurement model:

$$L(t_k) = \lambda \cdot V(t_k) + \epsilon_k, \quad (3.8)$$

$$F(t_k) = \gamma \cdot r(t_k) + \eta_k, \quad (3.9)$$

where the variables are as defined as for Eqs 3.3, 3.4 and 3.5. While  $\beta$  can be obtained from the literature (Chapter 2),  $\lambda$  and  $\gamma$  are unknown parameters that need to be estimated along with  $\alpha(\cdot)$  and  $\mu(\cdot)$ . Since we are satisfied with a qualitative reconstruction of  $\alpha(\cdot)$ , we can simplify the system by defining  $r_\gamma = \gamma \cdot r$  and  $V_\lambda = \lambda \cdot V$ . The dynamical system is consequently rewritten as

$$\dot{r}_\gamma(t) = \mu(t) \cdot \frac{\gamma \cdot \alpha(t)}{\beta} - \mu(t) \cdot r_\gamma(t), \quad (3.10)$$

$$\dot{V}_\lambda(t) = \mu(t) \cdot V_\lambda(t), \quad (3.11)$$

with initial conditions  $r_\gamma(0) = \gamma r_0$ ,  $V_\lambda(0) = \lambda V_0$ , and the new measurement model:

$$F(t_k) = r_\gamma(t_k) + \eta_k, \quad (3.12)$$

$$L(t_k) = V_\lambda(t_k) + \epsilon_k. \quad (3.13)$$

The Kalman filtering problem thus becomes the reconstruction of  $\gamma\alpha(\cdot)/\beta$  and  $\mu(\cdot)$  from measurements  $\{F(t_0), \dots, F(t_{N-1})\}$  and  $\{L(t_0), \dots, L(t_{N-1})\}$ .

Note that the above problem is not linear, whereas Kalman filtering was initially introduced for linear systems [118, 185]. For this reason, we use a

nonlinear extension of Kalman smoothing called unscented Kalman smoothing [119, 186]. Details about its exact implementation are available in Material and Methods 3.4.5. While it is feasible to simultaneously reconstruct the signals  $\mu(\cdot)$  and  $\gamma\alpha(\cdot)/\beta$  using Kalman smoothing, it is not necessary in our case and might lead to unstable results. We therefore decided to proceed in two steps: first reconstruct  $\mu(\cdot)$  from the measurements  $L(\cdot)$ , then inject this result into the reconstruction of  $\gamma\alpha(\cdot)/\beta$  from the measurements  $F(\cdot)$ .

### 3.2.4 Estimation of growth rate and resource allocation profile

#### Growth rate estimation

As presented in the previous section, we want to reconstruct the growth rate  $\mu$  defined by Eq. 3.11, using measurements of the length  $L$  defined by Eq. 3.13. Note that Kalman smoothing is a procedure that returns the expected mean and covariance of the state of a dynamical system, given the measurements. Therefore, reconstructing  $\mu$  requires it to be explicitly included as a state variable of the dynamical system. This provides constraints on the variation of  $\mu$  that have the effect of a regularization. In particular, in the Kalman smoothing procedure, we model the variations of  $\mu$  as the outcome of a stochastic process [184]. The laws describing this process play the role of a prior on the expected regularity of  $\mu$  (the more noisy the process, the less constrained the variations of  $\mu$ ). In particular, we define  $\mu$  as the double integral of a Gaussian white noise  $w$ :

$$\dot{\mu} = v(t), \quad \dot{v} = w(t),$$

where  $v$  is an intermediate variable and  $w(t)$  is assumed normally distributed with mean 0 and standard deviation  $\theta$ :  $w(t) \sim \mathcal{N}(0, \theta)$ . The resulting system then becomes

$$\begin{aligned} \dot{V}_\lambda(t) &= \mu(t) \cdot V_\lambda(t), \\ \dot{\mu}(t) &= v(t), \\ \dot{v}(t) &= w(t), \end{aligned} \tag{3.14}$$

with the measurement model

$$L(t_k) = V_\lambda(t_k) + \epsilon_k, \tag{3.15}$$

and the initial conditions

$$V_\lambda(0) = V_{\lambda,0}, \quad \mu(0) = \mu_0, \quad v(0) = v_0,$$

that may themselves be Gaussian random variables with some given statistics. The advantages of this regularization method are that the reconstructed  $\mu$  is guaranteed to be second-order differentiable [184]. It is also equivalent to other methods like Tikhonov regularization [187–189], which minimizes least-square differences between predictions and observations along with penalizing the variations of the input.

Our prior is thus entirely contained in the parameters of the Kalman procedure:

- the mean of the initial state  $(V_{\lambda,0}, \mu_0, v_0)$ ,
- the covariance of the initial state  $(V_{\lambda,0}, \mu_0, v_0)$ ,
- the transition covariance for the derivatives of  $V_{\lambda}$ ,  $\mu$  and  $v$  (see below),
- the variance of the observation noise  $\epsilon_k$ .

The mean and covariance of the initial state  $(V_{\lambda,0}, \mu_0, v_0)$  represent the knowledge we have of the initial values of the variables. For instance, if we know exactly the initial value of the growth rate  $\mu$  (*via* independent measurements or literature data), we can use it as mean for  $\mu_0$  along with a very small variance. This will constraint the signal reconstruction by penalizing deviations from this value at  $t = 0$ . On the contrary, if we are very uncertain of the value for  $V_{\lambda,0}$  (because cells are not synchronized, and can be in any state when the data acquisition starts), we can use a large variance for this initial condition. In our framework, there is no transition covariance for  $V_{\lambda}$  and  $\mu$ , because they are not the result of a stochastic process (the equations that define their derivatives are fully deterministic). However, by definition  $v$  is the result of a stochastic process of mean 0 and transition variance  $\theta^2$ . This value is crucial and represents the intensity of the white noise  $w$  that serves as prior for the regularization of  $\mu$ . The smaller its value, the more the variations of  $v$  are penalized, hence the smoother the reconstructed  $\mu(\cdot)$ . Finally, the variance of the observation noise  $\epsilon_k$  is simply the expected measurement noise. Together, these parameters represents the *prior* information we have on the initial conditions, the smoothness of the reconstructed signal, and the precision of our measurements, in order to reconstruct the growth rate  $\mu(\cdot)$ . As we will see below, we widely use these properties to overcome the difficulties introduced by the discontinuities following cell divisions.

The observation of the length  $L$  of growing bacterial cells is inherently discontinuous, due to the division of cells at regular time-points (Fig. 3.2). The estimation problem can therefore not be solved for the experiment as a whole, but only for the time-intervals between two successive division events, when the cell length is expected to change continuously. We localize the division events by identifying the time points at which the empirical derivative of the length is below an arbitrary threshold, and correspondingly slice the total duration of the experiment into time-intervals with continuous changes

of cell length. The simplest way to proceed from there on would be to treat each portion independently. However, by doing this we would lose important information, since the growth rates of a mother and its daughter cells are expected to be similar. In other words, while not necessarily continuous, the growth rate of new-born cells will depend strongly on the growth rate of their mother just before division. This can be easily taken into account in the Kalman smoothing procedure by exploiting the prior information about the initial mean and covariance of the system variables. In particular, we set the mean of the initial growth rate of a daughter cell equal to the final estimated growth rate of the mother cell, while defining a reasonable variance of the growth rate to allow for uncertainty. The results of this Kalman smoothing procedure, along with the exact parameters used, are reported in Fig. 3.3 and Material and Methods 3.4.5.

On average, the results presented in Fig. 3.3 show a roughly constant growth rate on acetate, followed by a quick increase after the upshift. Contrary to what was visible in the plot with the fluorescence data, a steady state for the growth rate seems to have been reached before the end of the experiment. When considering the individual cells, the results are more difficult to interpret. A significant part of the cells stopped growing before the end of the experiment, probably due to the fact that all cells die from an unknown cause around the end of the experiment. As a consequence, the analysis has been limited to a time-interval after the upshift in which roughly 3/4 of the cells are still growing. Interestingly, 1/6 of the cells exhibit growth rate oscillations, with a 1 h-long pause around 200 min after the upshift, followed by a recovery of the growth rate until the end of the experiment (see S2 Fig and S3 Fig). In order to focus the analysis, we decided to classify the cells in three categories: dying cells (N=12) for which the growth rate drops to zero after the upshift, pausing cells (N=11) for which the growth rate drops to zero after the upshift and then recovers, and finally so-called normal cells (N=45) which do not exhibit any of the above behaviors (see S2 Fig and S3 Fig). In what follows, we focus on the 45 normal cells (Fig. 3.4) which show a globally coherent behavior in the time window considered here.

### Estimation of resource allocation profile

In order to estimate the time-varying allocation of resources after the upshift, we use a similar Bayesian regularization approach as for the estimation of the growth rate. We note  $u(t) = \gamma\alpha(t)/\beta$  the resource allocation input that we wish to reconstruct. With  $\hat{\mu}$  the estimation of  $\mu$  obtained in the

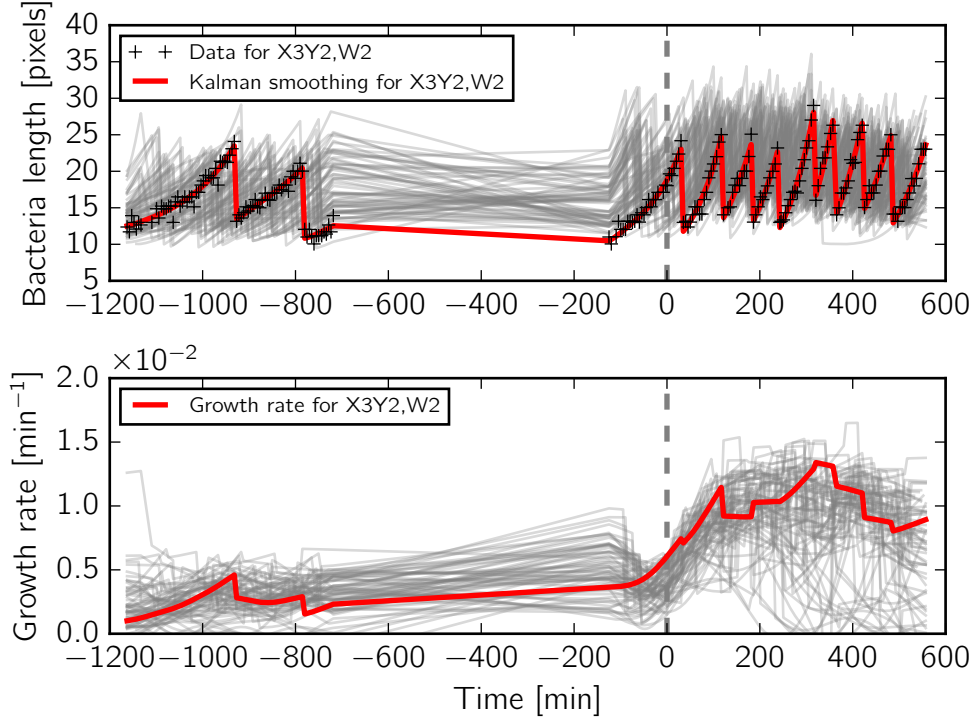


Figure 3.3 – **Growth-rate estimation using Kalman smoothing based on measurement of the length of bacteria growing in a microfluidic device.** Gray lines represent the estimation of the time-varying length (upper plot) and growth rate (lower plot) of 68 cells by the unscented Kalman smoothing procedure. The solid red lines highlight the result for one particular cell, located at the bottom of the well labelled "X3Y2, W2", while black crosses in the top graph are the data points for this cell. The vertical dashed lines represent the time of the upshift from growth on acetate to growth on glucose. As prior for the algorithm, we used an observation variance of  $9 \text{ pixels}^2$  for the length  $L$ . The transition variance  $\theta^2$  (*i.e.*, the smoothing factor for  $\mu$ ) is fixed at  $10^{-8} \text{ min}^{-6}$ . Inheritance between mother and daughter cells is taken into account by systematically choosing an initial mean growth rate equal to the last estimated value for  $\mu$  before cell division, and to half the last estimated value for  $V_\gamma$ , bearing in mind that *E. coli* cells divide symmetrically. At the start of the experiment, when no mother cell is available to provide initial estimates, the above values were fixed at 15 pixels for  $V_\gamma$  and  $0.004 \text{ min}^{-1}$  for  $\mu$ . The variances associated with these means are  $16 \text{ pixels}^2$  and  $10^{-4} \text{ min}^{-2}$ , respectively for  $V_\gamma$  and  $\mu$ . The initial mean of  $v$  is set equal to 0, with an initial variance of  $10^{-8} \text{ min}^{-6}$ . All the cross-covariances are set to 0 because the system variables are independent by construction.

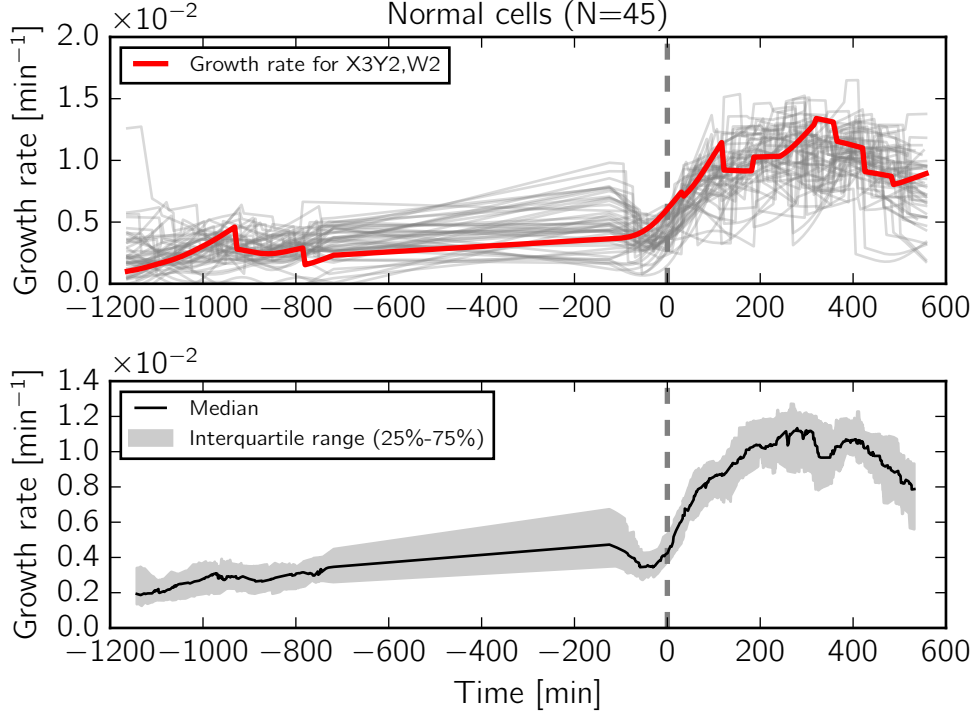


Figure 3.4 – **Growth-rate estimation using Kalman smoothing for the normal cells only.** The top graph is the same as the bottom graph of Fig. 3.3, except that the dying and pausing cells were removed. The bottom graph shows the 25% (lower gray curve), 50% (solid black curve) and 75% (upper gray curve) quartiles, computed at each time step. The gray area represents the interquartile range.

previous section, the full model for the reconstruction of  $u(\cdot)$  is given by

$$\begin{aligned} \dot{r}_\gamma(t) &= \hat{\mu}(t) \cdot u(t) - \hat{\mu}(t) \cdot r_\gamma(t), \\ \dot{u}(t) &= v(t), \\ \dot{v}(t) &= w(t), \end{aligned} \quad (3.16)$$

with the measurement model

$$F(t_k) = r_\gamma(t_k) + \eta_k, \quad (3.17)$$

and the initial conditions

$$r_\gamma(0) = r_{\gamma,0}, \quad u(0) = u_0, \quad v(0) = v_0,$$



that may themselves be Gaussian random variables with some given statistics. Contrary to the model used for the estimation of the growth rate, the system of Eq. 3.16 is linear. Furthermore, we do not have to deal with discontinuities in the measurements, since the fluorescence density  $F$  in the cells is continuous between mother and daughter cells (Fig. 3.2).

Given the analysis in Chapter 2, we expect  $u(\cdot)$  to exhibit bang-bang variations after the upshift. This is a discontinuous signal that could be complicated to reconstruct for certain choices of the parameters of the regularization method. In order to parametrize the Kalman smoothing algorithm, we generated synthetic data by simulating the model of Chapter 2. We estimated the noise characteristics of  $F$  using the RFU/pixel measurements from Fig. 3.2 (S7 Text) and chose parameters that allow the model to reproduce the observed growth rates and fluorescence densities. We simulated an upshift from acetate to glucose and tried to estimate  $\gamma\alpha(\cdot)/\beta$  from these synthetic data, selecting the *prior* in the Kalman smoothing procedure by trial and error. This approach may not have been optimal for our purpose and possible improvements are discussed in Section 3.3. The results along with the *prior* used are reported in Fig. 3.5. As expected, like any smoothing method, the algorithm has difficulty in reconstructing stiff variations in the state variables of Eq. 3.16. The switching profile of the resource allocation input is relatively well captured though, which indicates that the algorithm is in principle capable of reconstructing an on-off strategy (further improvements are discussed in Section 3.3). In what follows, we use the same *prior* of the Kalman smoothing procedure for the normal cells identified in the previous section, to test the occurrence of an on-off switching profile in our data.

The results of the reconstruction of the resource allocation profile  $u = \gamma\alpha(\cdot)/\beta$  are presented in Fig. 3.6. Within the interval between -150 and 0 min, resource allocation remains more or less stable, as expected for steady-state growth on acetate. On the contrary, the reconstructed resource allocation profile seems particularly unstable at the beginning and the end of the experiment. While oscillations do occur after the upshift, these need to be taken with much care. The problem of reconstructing an on-off strategy is more challenging than we initially thought, for the simple reason that the expected signal is similar to the kind of artifacts a poorly calibrated regularization method would generate. Data about the pre-upshift and post-upshift steady state are crucial for the calibration of the method and, as explained in Section 3.2.2, the experiment had to be interrupted before a steady state on glucose was attained. We extensively discuss this and other problems in Section 3.3 and give directions for future improvements.

Nevertheless, when not focusing on the individual cells but looking at the

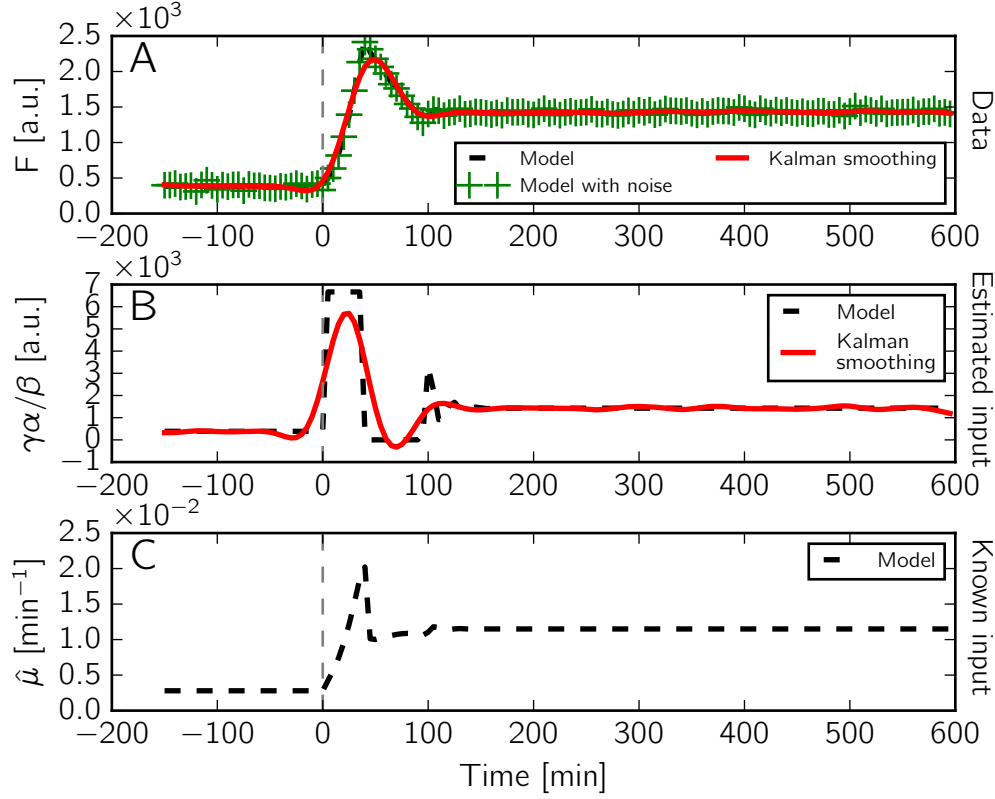


Figure 3.5 – **Performance of the Kalman smoothing procedure on synthetic data simulating an acetate-glucose upshift.** (A) Synthetic data simulating an upshift from acetate to glucose, with and without additive white noise, as well as the results obtained by Kalman smoothing. The synthetic data were generated by simulating the model presented in Eqs 2.3-2.4 with the on-off regulatory strategy (Eq. 2.19 and Fig. 2.7). The model parameters used for the simulation are  $e_{M,\text{Ace}} = 0.18 \text{ h}^{-1}$ ,  $e_{M,\text{Glu}} = 0.9 \text{ h}^{-1}$ ,  $k_R = 3.6 \text{ h}^{-1}$ ,  $\beta = 0.003 \text{ L g}^{-1}$ ,  $K_R = 1 \text{ g L}^{-1}$ . The predicted  $r(t)$  profile was multiplied by a factor  $\gamma = 0.02 \text{ RFU L g}^{-1}$  in order to obtain the corresponding fluorescence intensity profile  $F$  (dashed black curve). The noise level was estimated from the data (S7 Text) and added to  $F$ . The choice of the parameters of the Kalman smoothing procedure is discussed in the Material and Methods 3.4.5. (B) Estimation of the resource allocation profile  $\gamma\alpha/\beta$  based on the data in (A). Following Chapter 2,  $\alpha(t)$  displays a bang-bang-singular profile during the upshift (dashed black curve). While the Kalman smoother is not able to capture the discontinuous variations in  $\gamma\alpha/\beta$ , it qualitatively reproduces the input quite well (red solid curve). (C) The predicted growth rate during the upshift experiment. This information is used as an input in the smoothing procedure, since it is supposed to have been independently estimated from the measurements  $\{L(t_0), \dots, L(t_{N-1})\}$ .

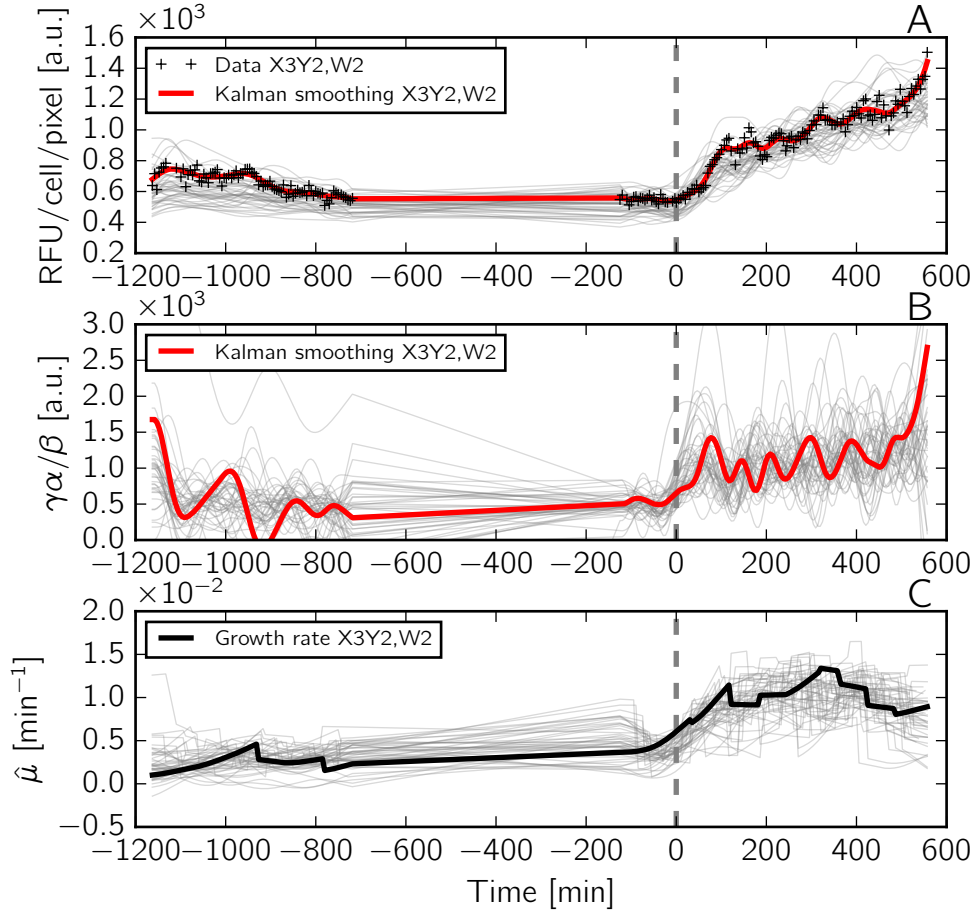


Figure 3.6 – **Estimation of the resource allocation profile using Kalman smoothing based on the fluorescence density measurements and the estimated growth rates from Fig. 3.3.** (A-B) Gray lines represent the estimation of the fluorescence density (RFU/pixel)  $F(\cdot)$  (in A) and the resource allocation profile  $\gamma\alpha/\beta$  (in B) by the Kalman smoothing procedure for the 45 normal cells. The solid red curves highlight the result for one particular cell, located at the bottom of the well labelled "X3Y2, W2", while black crosses in the top graph are the data points for this cell. The vertical dashed lines represent the time of the upshift from growth on acetate to growth on glucose. The prior values for the parameters of the smoothing algorithm are exactly the same as those used for Fig. 3.5 and are reported in the Material and Methods 3.4.5.

statistics of the entire data set, some interesting patterns emerge. Fig. 3.7 shows the median of the time-varying growth rate and resource allocation profile as well as the 25%-75% interquartile range. The use of these statistics gives a more robust view on the population level of the response of the cells to a nutrient upshift. When changing the carbon source from acetate to glucose, the growth rate increases to a value of around  $0.011 \text{ min}^{-1}$ , consistent with growth rates reported in the literature for the *E. coli* strain used here [76, 161], before decreasing when the first cells start to die (Fig. 3.7C). In addition, the data show one period of an oscillation in the first 3 h after the upshift, conserved in each of the 25%, 50% and 75% quartiles (Fig. 3.7B). The heatmap in Fig. 3.8 reveals that almost all of the normal cells show this oscillatory feature. Moreover, the first peak is seen to be even more pronounced on the level of the individual cells, reflecting the fact that in the different cells it occurs at different times after the upshift and is therefore dampened out at the population level. While the results of the microfluidic experiment presented here do not allow to confirm the occurrence of an on-off strategy for resource allocation after a nutrient upshift, the preliminary data are encouraging and prompt further investigation.

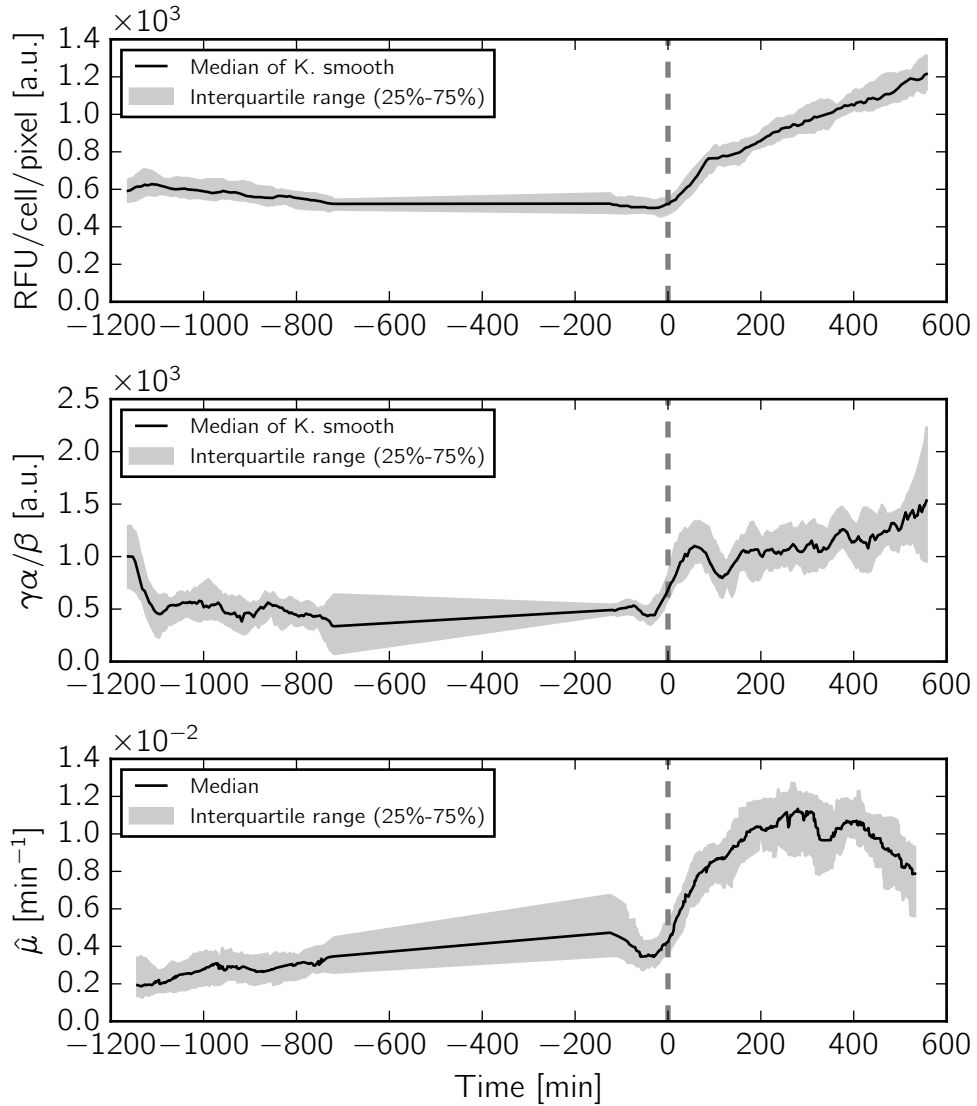


Figure 3.7 – **Robust statistics for the estimation results presented in Fig. 3.6.** Each graph shows the 25% (lower gray curve), 50% (solid black curve) and 75% (upper gray curve) quartiles, computed at each time step for the signals reconstructed in Fig. 3.6. The gray area represents the interquartile range. Interestingly, while most oscillations in the resource allocation profile  $\gamma\alpha(\cdot)/\beta$  cancel out at the population level, the first peak after the upshift is conserved.

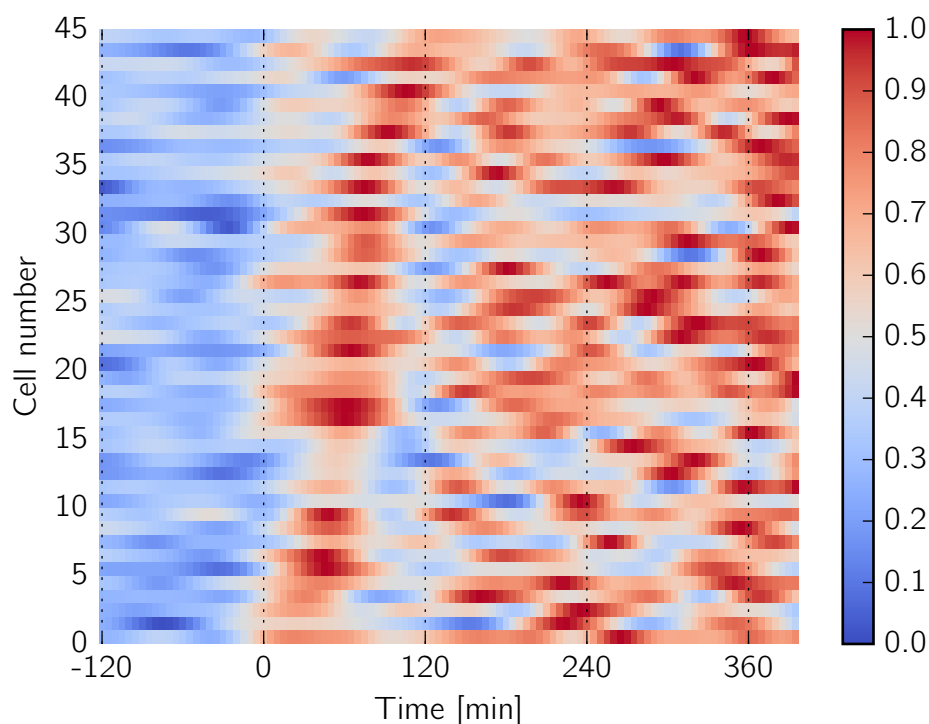


Figure 3.8 – **Global overview of all estimated resource allocation profiles presented in Fig. 3.6.** The individual  $\gamma\alpha(\cdot)/\beta$  curves have been normalized with respect to their maximum in the interval  $[0, 400]$  and then ordered with respect to their maximum in the interval  $[0, 120]$ .

### 3.3 Discussion

As we showed in Chapter 2, adopting a dynamical perspective might prove useful to unveil and understand the regulatory strategies employed by microorganisms. The criterion of biomass maximization has allowed to account for steady-state empirical growth laws. Interestingly, when the same criterion is applied to growth transitions, it predicts that microorganisms allocate their resources through bang-bang control of gene expression. For the control of ribosome synthesis, the implementation of this optimal control scheme results in an on-off strategy, where a bacterial cell is either producing only ribosomes or not producing them at all during the adaptation to a new growth medium. The ppGpp system, ubiquitous in bacteria and playing an important role in the control of ribosome synthesis in *E. coli*, satisfies the requirements posed by the on-off strategy. However, we currently lack exper-

imental data to show that it actually functions in this manner, among other things because of the difficulties to produce well-controlled growth transitions on the single-cell level in the laboratory. Can we set up appropriate experiments to measure the expression of ribosomes during a growth transitions? Will the bacterial cells exhibit the on-off strategy that was shown to be close to optimal in Chapter 2?

In this chapter, the above questions have been addressed by developing an experimental framework which allows the ribosome concentration to be quantified in individual *E. coli* cells in real time. Our main contribution consists in bringing together several experimental techniques developed in recent years. First, inspired by the work of Bakshi *et al.* [81], we constructed an *E. coli* strain with fluorescent ribosomes. In their paper, Bakshi *et al* essentially used this fluorescent marker to localize ribosomes in the cytoplasm through superresolution microscopy. While improving their design (Material and Methods 3.4.1), we used our strain for the purpose of quantifying ribosomal abundance in single cells. Second, we employed the mother machine developed by Wang *et al.* [45], originally developed to study long-term steady-state growth of *E. coli*, to observe individual *E. coli* cells growing in continuous culture by means of time-lapse fluorescence microscopy. Here, the mother machine was used to establish well-controlled growth transitions, *i.e.* from steady-state growth in minimal medium with acetate to steady-state growth in minimal medium with glucose. By changing the medium source of the mother machine, we were able to observe how *E. coli* cells adapt their growth rate and the expression of ribosomes. Overall, through the combination of fluorescent labeling of a core component of the gene expression machinery and the use of a microfluidic device, we have opened the way to experimentally study growth laws in a dynamical context.

We applied this experimental setup to the reconstruction of the resources allocation strategy employed by *E. coli* during a nutrient upshift, with the aim to observe the bang-bang profile predicted in Chapter 2. We developed a Kalman smoothing procedure adapted to our question and showed how it can be implemented to reconstruct the internal state of a cell from time-lapse microscopy images. Similar to hidden Markov chains, Kalman filtering and smoothing are powerful techniques for the reconstruction of unobserved signals from noisy measurements, and have had numerous applications in a variety of domains, though less in biology than in other fields. We showed in this chapter that the Bayesian framework of Kalman smoothing can be exploited for growth rate reconstruction from a probabilistic prior defining the relation between mother and daughter cells. We also developed another variant of Kalman smoothing for the reconstruction of the resource allocation variable  $\alpha(\cdot)$  that is at the heart of the models of Chapter 2. Our work

on synthetic data showed that, despite the challenge posed by the reconstruction of a discontinuous signal, the algorithm was capable of recovering an oscillatory pattern close to the expected on-off profile. Unfortunately, its application to the real data did not allow an unambiguous conclusion to be drawn: while oscillations do occur, and a first peak between 0 and 120 min after the upshift is visible in almost all of the cells, it turned out to be far from straightforward to distinguish a real oscillatory pattern from artifacts of the data analysis and signal reconstruction methods. Below we propose several improvements that will be addressed in future work.

Several experimental problems complicated the analysis of the data. Sufficiently long steady-state measurements before and after the transition of interest are needed to provide a reliable estimation of the intensity and the nature of the experimental noise. They are also critical to calibrate the parameters of the regularization method used for signal reconstruction. Despite what was initially planned, we did not completely reach a new steady state after the transition. Several causes are to blame. First, the random stall of the motorization along the Z-axis of the microscope, which does not allow overnight measurements and occurred during the slow growth on acetate in the presented experiment. Second, the massive death of bacteria starting after the transition on glucose. Future work will focus on solving these issues.

The death of bacteria is particularly worrying. During the construction of this strain, most people in the lab experienced contamination by an aggressive bacteriophage. Unfortunately, the strain used in this study was not spared, and our frozen stocks were later tested positive for phage-induced lysis. Interestingly, the lysis seems to depend strongly on the environmental conditions. To our disarray, the massive death observed in this experiment suggests that a strong environmental upshift could trigger the lytic cycle of the phage. While impeding any possible long-term measurement, the death of the bacteria is not our only worry. Phages are machines that are extremely well optimized to divert cell resources to their own end, especially the activity of the gene expression machinery. That could dramatically perturb the cell regulation of resource allocation, and so the conclusions of this study. For this reason, future work will start by the reconstruction of a clean phage-free strain from scratch.

It is also possible to optimize the image analysis techniques used in this study. We segmented the images using the most powerful and ubiquitous segmentation algorithm available: the human brain. In other words, we manually selected the pixels representing the poles of a single bacteria per well on the fluorescence images, and used this information to arbitrarily define a rectangle around the cell of interest. While this is far from satisfying, it may turn out to be difficult to improve upon this, as it seems that the first



rule of image analysis is that every application is more or less unique. Every algorithm has to make assumptions about the object it is trying to recognize. For imaging of microbial cells, it could be the curvature at the poles, the size of the pixels, the homogeneity of the background, the homogeneity of intracellular fluorescence, ... While some parameters can be tweaked, some assumptions are always hard-coded in the algorithm and reflect the philosophy used to address the identification problem. Since the preliminary analysis reported here was completed, we started a collaboration with the authors of FluoBacTracker [190] that focuses on adapting their software to our setup in a near future. It should provide a robust, automated way to identify and track the cells, enhancing the reproducibility of the analysis and increasing the number of cells available for computing population statistics.

The Kalman filtering algorithm that was later applied to these data can also be improved in many respects. As described in the Results section, the algorithm is an instance of Bayesian inference and several parameters describing the expected input define a probabilistic prior. In the work reported here, we chose these parameters through trial and error, or by calibrating the system on synthetic data whenever possible. A better technique would be to choose these parameters by *generalized cross-validation* (GCV) [191], a procedure maximizing the predictive power of the reconstructed signal by reducing overfitting. The use of GCV in our framework amounts to choosing the parameters of the Kalman smoother by optimizing predictions on a subset of the cells, then testing the performance on the remaining cells. While the initial conditions did not seem to play a huge role in the final shape of the signal, the choice of the smoothing factor  $\theta^2$  is critical and may strongly benefit from the proposed GCV extension.

The Kalman smoothing algorithm itself can also be further improved, notably by addressing the difficulty of estimating abrupt transitions. As can be seen in Fig. 3.5, the resource allocation profile  $\gamma\alpha(\cdot)/\beta$  reconstructed from the synthetic data increases before the nutrient shift and this artifact also occurs when using the measured fluorescent densities (Fig. 3.7), the change in resource allocation preceding the nutrient upshift by several dozens of minutes. While we know exactly when the change of medium occurs, this information is currently not used as a prior for the reconstruction. The Kalman smoother can be improved by implementing time-varying parameters for the regularization method, so as to strongly penalize variations of the reconstructed signal at steady state and to release this constraint following the change in medium.

The improvements of the approach proposed in this section, which would help in reaching a conclusion about the existence of the on-off strategy, are not very complicated to realize. Another improvement consists in taking

into account the autofluorescence of the bacteria, which we assumed to be negligible, mostly because its proper estimation in a dynamical, single-cell context is complicated (Material and Methods 3.4.4). A simple trick, however, could help in quantifying autofluorescence: mixing wild-type bacteria with bacteria having fluorescent ribosomes when loading the device. This would allow some channels to be occupied by wild-type cells, in addition to the channels in which the strain of interest is growing, and thus obtain a reliable estimation of the level of autofluorescence of the cells. Notice though that the proportion between the two types of bacteria need to be carefully adjusted in order to preserve enough cells for the estimation of the resource allocation profile.

## 3.4 Material and Methods

### 3.4.1 Bacterial strain construction

In order to quantify the concentration of ribosomes in the cell in real time, we constructed a strain containing a translational fusion of *gfp\_mut2* [91] to the C-terminus of *rpsB*, the gene coding for the ribosome subunit S2. The design was inspired by the work of Bakshi *et al.*, who used a similar construction to measure the spatial distribution of ribosomes in living *E. coli* cells [81]. However, contrary to Bakshi *et al.*, and in order not to create any interference with normal gene expression in the modified cells, we decided to keep the intergenic region between *rpsB* and the downstream gene, *tsf*, unchanged. Since the transcription factor *tsf* is under the control of the same promoter as *rpsB*, leaving a selection marker downstream of *rpsB* would interfere with the proper expression of *tsf*. Our strategy comprises two steps: (i) creation of the translational fusion by selecting an antibiotic resistance marker and (ii) removal of the selection marker. The strategy is outlined in Figure 3.9.

The DNA fragment containing the selection cassette was amplified using two long primers annealing, respectively, downstream of *rpsB* (starting just after the STOP codon), and to the end of *gfp\_mut2*. The selection cassette contains an antibiotic resistance gene, kanamycin (positive selection), and a gene coding for the CcdB toxin under the control of the *p<sub>BAD</sub>* promoter (negative selection in presence of arabinose). This cassette is referred below as *kan-p<sub>BAD</sub>-ccdB*.

Another DNA fragment containing *gfp\_mut2* [91] without the ATG start codon was amplified using long primers annealing, respectively, to the C-terminus of *rpsB* (just before the STOP codon), and the end of the *kan-p<sub>BAD</sub>-ccdB* cassette. The first primer also contained a 18-bp (base pair)

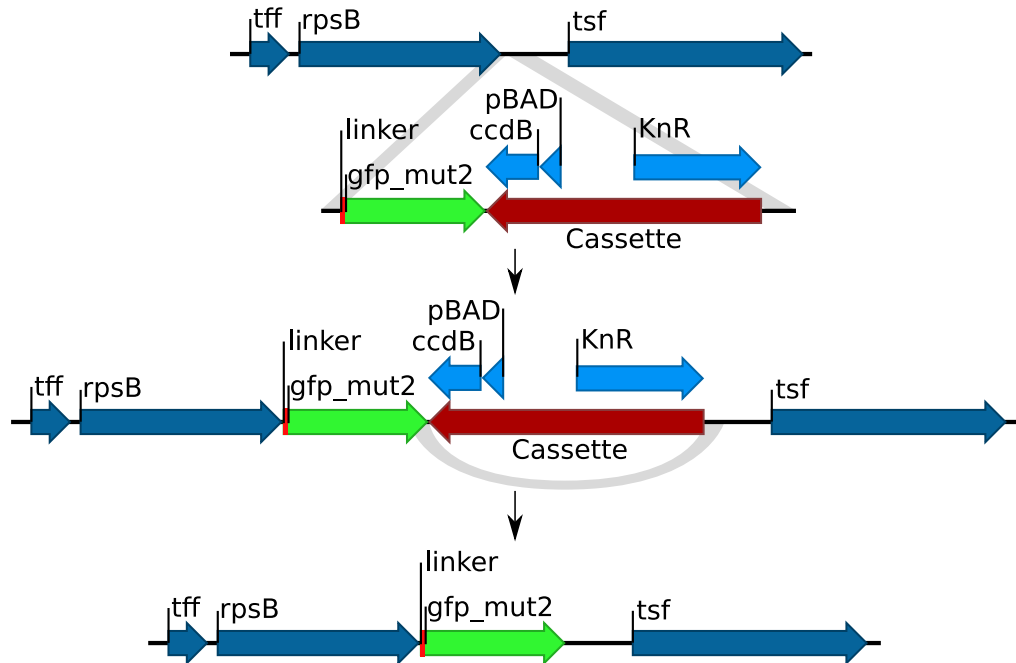


Figure 3.9 – **Construction of the *gfp*-tagged ribosome subunit.** The translational fusion of *gfp\_mut2* with *rpsB* was constructed on the chromosome of *E. coli* (top line) by homologous recombination of *gfp\_mut2* followed by a selection “cassette” (second line). A small linker was inserted between *rpsB* and *gfp\_mut2* in order to minimize interference of the fluorescent protein with ribosome functioning. The selection cassette consists of a positive selection marker, the gene coding for the resistance to the antibiotic kanamycin, and a negative selection marker, the gene coding for the toxin CcdB. The latter is transcribed from the *p<sub>BAD</sub>* promoter, which is only activated in the presence of arabinose in the culture medium. Homologous recombination is indicated by the grey shaded lines. The resulting recombination product (third line) contains the desired fusion protein followed by the selection cassette on the chromosome of the bacterium. A second homologous recombination, using an oligonucleotide complementary to the end of *gfp\_mut2* and the beginning the original region downstream of *rpsB* removes the selection cassette. The resulting strain (line four) carries the translational fusion of *gfp\_mut2* to *rpsB* without any other modification of the chromosome.

linker, inserting the same six amino-acids that Bakshi *et al.* have used for their construction [81].

The two fragments, the *gfp\_mut2* reporter including the linker and the *kan-p<sub>BAD</sub>-ccdB* cassette, were assembled by Gibson assembly [192] using a commercial Gibson Assembly mix (New England Biolabs). A final product of 2683 bp was obtained:

- \*(50 bp) the C-terminus of *rpsB* without the STOP codon
- (18 bp) a linker
- (714 bp) the *gfpmut2* sequence without the initial ATG
- (1851 bp) the *kan-p<sub>BAD</sub>-ccdB* cassette
- \*(50 bp) the region directly after *rpsB* in *E. coli*

Regions labeled with \* anneal to the *E. coli* chromosome. The complete sequence of this fragment, as well as the sequences of all the primers, are listed in [S5 Text](#).

This fragment was electroporated into the wildtype strain, BW25113, containing the pSIM5 plasmid for lambda-red recombination [193]. A kanamycin-resistant colony was selected on LA-glucose medium and verified to show green fluorescence (485 nm excitation, 535 nm emission) in a microplate reader (Tecan Infinite 200 pro). A 100-bp oligonucleotide, containing 50 bp of the end of *gfp\_mut2* and 50 bp of the region just downstream of *rpsB* was electroporated into this strain (sequence available in [S5 Text](#)). The homologous recombination removes the *kan-p<sub>BAD</sub>-ccdB* cassette. A colony was selected on an LA-arabinose plate. The clone was tested for kanamycin-sensitivity and green fluorescence. Finally, the strain was grown overnight at 42°C to get rid of the pSIM5 plasmid, containing a temperature-sensitive origin of replication. A chloramphenicol-sensitive colony was chosen and the region after *rpsB* was verified by sequencing (full sequence obtained available in [S5 Text](#)).

In parallel, the same protocol was used to construct *mCherry* and *cfp* variants of the same strain. However, only the *gfp\_mut2* and *mCherry* versions were successfully obtained. The full sequence of the final *rpsB-mCherry* strain is available in [S5 Text](#).

The *rpsB-gfp* and *rpsB-mCherry* strains were characterized on different media using a Tecan microplate reader ([S6 Text](#), in particular Fig. 3.10 for growth on glucose). They possess a wild-type growth rate and sufficient fluorescence to allow quantification. However, the *rpsB-mCherry* strain exhibited strange fluorescence dynamics, especially during growth transitions, which made it unsuitable for our study. We therefore concentrated our efforts on the *rpsB-gfp* strain (see [S6 Text](#) for more details).

### 3.4.2 Growth conditions

All experiments were carried out in sterilized media. Sterilization was performed by autoclaving for instruments and filtration for solutions (0.2  $\mu\text{m}$ ). During cloning, bacteria were grown in 20-mL Erlenmeyer flasks filled with LB or spread on Petri dishes with LA. In all the experiments, we used M9 minimal medium supplemented with trace elements, thiamine and a carbon source. The full recipe is reproduced below. Numbers in squared brackets are characteristics of the stock solution. All stock solutions were stored at room temperature, except for  $\text{FeSO}_4$  [30  $\text{g.L}^{-1}$ ] and Thiamine [10  $\text{g.L}^{-1}$ ], which were stored respectively at  $-20^\circ\text{C}$  and  $4^\circ\text{C}$ .

M9 medium (100 mL)		Traces elements (0.9 mL)	
$\text{CaCl}_2$ [1 $\text{mol.L}^{-1}$ ]	10 $\mu\text{l}$	$\text{H}_2\text{O}$ [18.2 $\text{M}\Omega.\text{cm}$ ]	200 $\mu\text{l}$
$\text{MgSO}_4$ [1 $\text{mol.L}^{-1}$ ]	200 $\mu\text{l}$	$\text{Na}_2\text{EDTA } 2\text{H}_2\text{O}$ [150 $\text{g.L}^{-1}$ ]	100 $\mu\text{l}$
5x Salts	20 mL	$\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ [45 $\text{g.L}^{-1}$ ]	100 $\mu\text{l}$
Traces elements	90 $\mu\text{l}$	$\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$ [3 $\text{g.L}^{-1}$ ]	100 $\mu\text{l}$
$\text{FeSO}_4$ [30 $\text{g.L}^{-1}$ ]	10 $\mu\text{l}$	$\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ [10 $\text{g.L}^{-1}$ ]	100 $\mu\text{l}$
Thiamine [10 $\text{g.L}^{-1}$ ]	50 $\mu\text{l}$	$\text{H}_3\text{BO}_3$ [10 $\text{g.L}^{-1}$ ]	100 $\mu\text{l}$
Carbon source	at will	$\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ [4 $\text{g.L}^{-1}$ ]	100 $\mu\text{l}$
$\text{H}_2\text{O}$ [18.2 $\text{M}\Omega.\text{cm}$ ]	to 100 mL	$\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ [3 $\text{g.L}^{-1}$ ]	100 $\mu\text{l}$

5x Salts (100 mL)	
$\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$	4.25 g
$\text{KH}_2\text{HPO}_4$	1.5 mg
$\text{NaCl}$	0.25 g
$\text{NH}_4\text{Cl}$	0.5 g
$\text{H}_2\text{O}$ [18.2 $\text{M}\Omega.\text{cm}$ ]	to 100 mL

A large quantity of M9 was prepared several days before the experiment, and stored at  $4^\circ\text{C}$ . It contained all the necessary components except  $\text{FeSO}_4$ , thiamine, and carbon sources, which were added just before inoculation of the pre-culture. Except where stated otherwise, the growth temperature was  $37^\circ\text{C}$ . M9 Acetate contains 0.2% acetate (in mass of  $\text{C}_2\text{H}_3\text{O}_2$  per mass of solution), and M9 Glucose contains 0.2% glucose (in mass of D-(+)-Glucose per mass of solution).

Four days before the measurements, the glycerol stock containing the *rpsB-gfp* strain was spread on a Petri dish of LA, and incubated at  $37^\circ\text{C}$ . On the following day, an isolated colony was used to inoculate a cotton-plugged, sterile flask containing 20 mL of M9 acetate (Fig. 3.1). The time of inoculation was calculated such that the culture obtains an OD of 0.3-0.4 OD at the beginning of the experiment. This optical density corresponds to a culture in mid-exponential growth phase.

At the day of the experiment, the pre-culture was concentrated by centrifugation and re-suspended in 5 mL of M9 acetate, supplemented with 50 mg.mL<sup>-1</sup> BSA (passivation buffer) for injection into the microfluidic device. Channels were filled by diffusion until most of them contained cells (~1 hour). Data acquisition started after a constant medium flow was successfully obtained (~1 hour, depending on the quality of the microfluidics device).

### 3.4.3 Microfluidic device

In order to observe the bacteria over a long time, we used the so-called mothermachine [45]. It consists of a series of wells (or channels), oriented at a 90° angle to a large, central channel through which growth medium is passed at a constant flow rate. The width of the wells, a little over a  $\mu\text{m}$ , constrains the bacteria to grow “in a line”. This design ensures that at least one cell per well (the one at the bottom of the channel) remains in the device during the entire experiment, while the others incrementally escape into the central channel as divisions occur. For the fabrication of the devices, we thoroughly followed the step described in the supporting information of [45]. We maintained a stock of chemically treated devices at room temperature (day 2 in workflow summary [45]). The day of the experiment, a single device was plasma cleaned, bond to a glass coverslip, and injected with bacteria (see section above).

The device was connected using 0.023" inner diameter polyethylene tubes to a waste and a sterile bottle containing 200 mL of growth medium, which was enough for several days of acquisition. A microfluidic pump (Elvesys) containing an output flow sensor module was plugged to the medium bottle and a pressure up to 2 bars was applied to the bottle. The output flow was set to 50  $\mu\text{L}\cdot\text{min}^{-1}$ .

For imaging, the device was placed on a motorized inverted microscope (Zeiss Axiovert 200M) with a phase contrast objective lens (Zeiss Plan-Neofluar, Ph3 100x/1.3), placed in a thermostated box at 37°C. In this setup, fluorescence illumination is provided by a mercury lamp (Osram, 1xHBO 103X/2) and visualization is performed with narrow-bandpass excitation and emission filters (Chroma, #49002 ET-GFP and Chroma, #49005 TR/DsRED ET). The exposure time is externally controlled by mechanical shutters (Uniblitz-VS35). Images were acquired with a 16-bit gray level CCD camera cooled to -80°C (Roper Scientific, Princeton Instruments PHOTOMAX 512) controlled by a custom-made software using Visual Basic and the Type libraries of the Winview software (Princeton Instruments). Every 5 minutes, autofocus was numerically performed by maximizing the contrast

of a region of interest, and a series of acquisitions were made. A total of 6 fields, each containing 15 wells was observed during the entire experiment.

### 3.4.4 Cell segmentation

The raw data from the camera are in the form of 512x512-pixel 16-bit SPE images (proprietary format produced by the Princeton Instruments camera). They were converted into 16-bit Tif images using the *SPE* plugin of ImageJ [194]. The rest of the analysis was performed using Python 3.5.2. In particular, we used OpenCV 3.1.0-dev, Scikit-Image 0.12.3, Numpy 1.11.2 to manipulate the images.

In order to correct for the drift of the device, we calculated the offset between consecutive images of a given field using cross-correlation in the Fourier space [195] (see Scikit-Image documentation at [196]). All images for a field were thereby aligned to the first acquisition by a simple translation.

In order to simplify the image treatment, we extracted individual wells from the images by a combination of manual pixel selection and automatic segmentation. In particular, we selected the entrance of the two wells at the border of the image, and used this information along with the regularity of the microfluidic device to compute a mask that allowed to extract each well into an image of size 100x21 pixels. These sub-images were labeled  $\{W_0, W_1, \dots, W_{14}\}$  depending on the position of the well in the original image, from left to right.

In this study, we performed a preliminary analysis of each well, focusing on the cell at the bottom of the channel. For each image, we manually selected the poles of the cell of interest. The distance between the two selected pixels was directly used as the length  $L$  of the bacterium in pixels. The two selected positions were also used to compute a rectangular 6-pixel-wide mask around the cell (see Fig. 3.2). We then computed the RFU/pixel in this cell mask by summing each pixel and dividing by the mask size.

The camera noise and background were evaluated by taking a picture with a closed shutter at the end of the experiment. Pixels in this picture were found to be Gaussian distributed with a mean of 1101.0 and a standard deviation of 10.788. A correction for the camera background was thus applied by removing 1101 from the computed RFU/pixel.

The autofluorescence of the bacteria and the background fluorescence of the medium were supposed to be negligible and were thus not corrected. As a control, the same device was used to image bacteria in stationary phase that do not produce any fluorescent protein. They were indistinguishable from the camera noise. Of course, using this control we can not affirm that the autofluorescence is also negligible when the bacteria are actively growing.



Furthermore, the autofluorescence could change in different media of interest, and may even be different in steady state and during growth transitions. An independent estimate of the autofluorescence is therefore difficult to obtain. In the section 3.3, we discuss possible improvements of the experimental setup that would allow to dynamically co-estimate the autofluorescence with the ribosome abundance in a single experiment.

### 3.4.5 Kalman smoothing

The data of the length and the RFU/pixel of each bacteria were analyzed using Python 3.5.2. In particular, we manipulated the data using Pandas 0.19.1, Pykalman 0.9.5 and the Curves submodule of Wellfare 0.1.1.

Historical details about the Kalman smoothing procedure are reported in section 3.2.3 of the main text. We used the Additive Unscented Kalman Filter implementation of the Pykalman python module. This class is reported to be more stable and computationally efficient with non-linear problems containing additive noise.

As described in section 3.2.3, the reconstruction of the growth rate was done on continuous portions of the length, *i.e.* between cell divisions. The full problem, as described in Eq. 3.14, is reproduced below for clarity:

$$\begin{aligned}\dot{V}_\lambda(t) &= \mu(t) \cdot V_\lambda(t), \\ \dot{\mu}(t) &= v(t), \\ \dot{v}(t) &= w(t),\end{aligned}$$

with the measurement model:

$$L(t_k) = V_\lambda(t_k) + \epsilon_k,$$

and the initial conditions

$$V_\lambda(0) = V_{\lambda,0}, \quad \mu(0) = \mu_0, \quad v(0) = v_0.$$

The parameters used as priors for the reconstruction of  $\mu$  are described below. We used an observation variance of 9 pixels<sup>2</sup> for the length  $L$ . The transition variance  $\theta^2$  (*a.k.a.* the smoothing factor for  $\mu$ ) is fixed at  $10^{-8} \text{ min}^{-6}$  for the entire time series. Inheritance between mother and daughter cells is taken into account by systematically choosing an initial mean for  $\mu_0$  equal to the last estimated value for  $\mu$ , and for  $V_{\gamma,0}$  to half the last estimated value for  $V_\gamma$  (as expected for a completely symmetrical division). At the beginning of the experiment, when no mother cells are available, these values were fixed to 15 pixels for  $V_\gamma$  and  $0.004 \text{ min}^{-1}$  for  $\mu$ . The variances associated



with these means are 16 pixels<sup>2</sup> and  $10^{-4}$  min<sup>-2</sup>, respectively, for  $V_{\gamma,0}$  and  $\mu_0$ . The initial mean of  $v$  is always taken as null, with an initial variance of  $10^{-8}$  min<sup>-6</sup>. All the cross-covariances are set to 0 because the system variables are independent by construction.

For the reconstruction of gene activity, we had to cope with the large gap in data acquisition between -720 and -150 min. Reconstruction was then done independently on the continuous sections before and after this gap. The full problem for the estimation of  $u = \gamma\alpha/\beta$ , as described in Eq. 3.16, is reproduced here for clarity:

$$\begin{aligned} \dot{r}_\gamma(t) &= \hat{\mu}(t) \cdot u(t) - \hat{\mu}(t) \cdot r_\gamma(t), \\ \dot{u}(t) &= v(t), \\ \dot{v}(t) &= w(t), \end{aligned}$$

with the measurement model

$$F(t_k) = r_\gamma(t_k) + \eta_k,$$

and the initial conditions

$$r_\gamma(0) = r_{\gamma,0}, \quad u(0) = u_0, \quad v(0) = v_0,$$

where  $\hat{\mu}$  is the estimate of  $\mu$  above. In both case, we used as prior an observation variance of 800.4 RFU<sup>2</sup> for  $F$ . The transition variance  $\theta^2$  (*a.k.a.* the smoothing factor for  $\gamma\alpha/\beta$ ) is fixed at  $10^2$  RFU<sup>2</sup>.min<sup>-4</sup>. The initial state means used are 600 RFU for  $r_{\gamma,0}$  and 1000 RFU for  $u_0$ . The variances associated with these means are purposely large and fixed at  $10^6$  RFU<sup>2</sup> for  $r_{\gamma,0}$  and  $u_0$ . The initial mean for  $v_0$  is always taken as null, with an initial variance of  $10^{-8}$  RFU.min<sup>-2</sup>, imposing a null second derivative for the reconstructed signal  $\gamma\alpha/\beta$ . Here again, all the cross-covariances are set to 0.


All the parameters cited in this section were chosen through trial and error, and are therefore certainly optimizable. Possible improvements are discussed in Section 3.3.

## 3.5 Supporting Information for Chapter 3

### 3.5.1 S5 Text – DNA sequences used for the strain construction

Exhaustive list of the primers used

...for the *gfp\_mut2* amplification.


 GTTCTCAGGATCTGGCTTCCCAGGCGGAAGAAAGCT  
TCGTAGAAGCTGAGCAGGAAAGGCGACAGGAGAGTAAAG  
GAGAAGAAGCTTTTCACTG (Length: 93)


*The 50 first bp anneals with the C-terminus of rpsB (just before the STOP codon). The 18 next bp code for the linker (see Material and Methods 3.4.1).*

 TGATGTTCTGGGGAATATAATTATTTGTATAGTTCA  
TCCATGCC (Length: 44)

*The 20 last bp anneals with the end of the kan-p<sub>BAD</sub>-ccdB cassette.*

...for the *mCherry* amplification

 GTTCTCAGGATCTGGCTTCCCAGGCGGAAGAAAGCT  
TCGTAGAAGCTGAGCAGGAAAGGCGACAGGAGACTAGCAA  
AAGATCCAAGGG (Length: 88)

 TGATGTTCTGGGGAATATAATTATTTGTACAGCTCA  
TCCATG (Length: 42)

*Same design as for *gfp\_mut2*, except it amplifies *mCherry*.*

...for the cassette amplification

 TGGATGAACTATACAAATAATTATATTTCCCAGAAC  
TCAGG (Length: 42, used to assemble with *gfp*)

*The 20 first bp anneals with the end of *gfp\_mut2*.*

 TGGATGAGCTGTACAAATAATTATATTTCCCAGAAC  
TCAG (Length: 41, used to assemble with *mCherry*)



*The 20 first bp anneals with the end of *mCherry*.*

 GAGCTTGCCGCCTTTCTGCAACTCGAACTATTTTGGG  
GGAGTTATCAAGCTTAGAAGAACTCGTCAAGAAGG

(Length: 72, used for both)

*The 50 last bp anneals downstream *rpsB* (just before the STOP codon).*



**...for the whole insert amplification**

 GTTCTCAGGATCTGGCTTCCCAGG (Length: 24)  
 GAGCTTGCCGCCTTTCTGCA (Length: 20)

**...for the cassette elimination**

*rpsB-gfp* TTGTAACAGCTGCTGGGATTACACATGGCATGGAT  
 GAACTATACAAATAAGCTTGATAACTCCCCAAAATAGTT  
 CGAGTTGCAGAAAGGCGGCAAGCTC (Length: 100)  
*rpsB-mCherry* GCGCGGAGGGTCGTCATTCTACCGGTGGCAT  
 GGATGAGCTGTACAAATAAGCTTGATAACTCCCCAAAAT  
 AGTTCGAGTTGCAGAAAGGCGGCAAGCTC (Length: 100)

**...for the sequencing of the final strain**

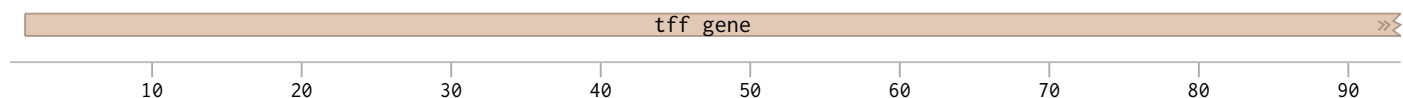
 CGTCTGAAAGACCTGGAAAC (Length: 20)  
 AAACGTGTACTACCTGGTCTATAAGG (Length: 26)

**Full annotated sequences of the inserts**

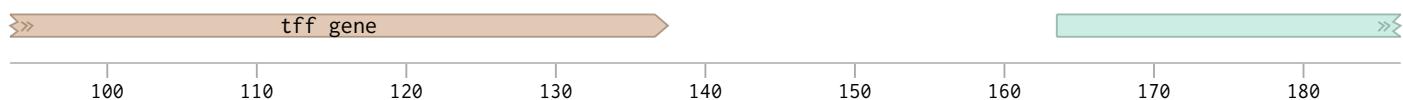
The following pages contain the full annotated sequences of the region downstream of *rpsB* after the insertion of the *gfp\_mut2-cassette* and the *mCherry-cassette*, respectively (third line in Fig. 3.9). The positions where the primers anneal to the template are indicated using the same color code as in the description of the primers above. The primers reported in the publication of Bakshi *et al.* [81] are indicated for information.

## gfp\_mut2-cassette after rpsB (E. coli) (4574 bp)

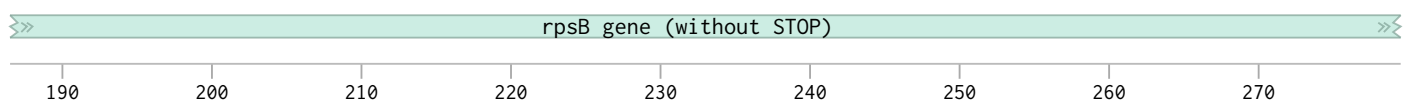
CCGGA CTTCGATCCATTTTCGTATACACAGACTGGACGGAAGCGACAATCTCACTTTGTGTAACAACACACACGTATCGGCACATATTCGGG  
GGCCTGAAGGCTAGGTAAAGCATATGTGTCTGACCTGCCTTCGCTGTTAGAGTGAAACACATTGTTGTGTGTGCATAGCCGTGTATAAGGCC



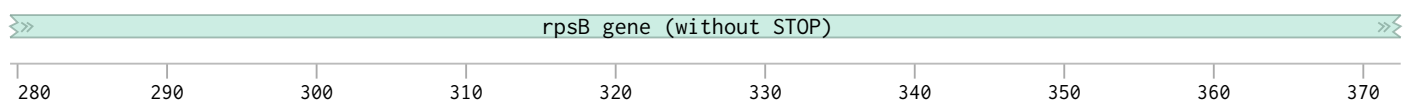
GTGCCCTTTGGGGTCGGTAATATGGGATACGTGGAGGCATAACCCCACTTTTATATAGAGGTTTTAATCATGGCAACTGTTTCCATGCGCGA  
CACGGGAAACCCAGCCATTATACCCTATGCACCTCCGTATTGGGGTTGAAAATATATCTCCAAAATTAGTACCGTTGACAAAGGTACGCGCT



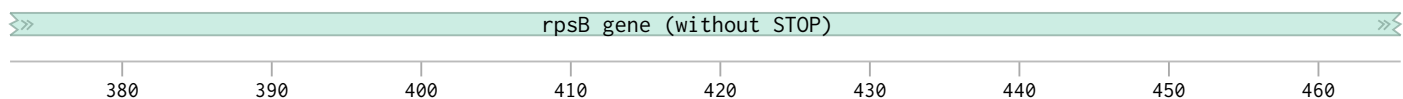
CATGCTCAAGGCTGGTGTTCACCTTCGGTCACCAGACCCGTTACTGGAACCCGAAAATGAAGCCGTTTCATCTTCGGTGC GCGTAACAAAGTTCA  
GTACGAGTTCCGACCACAAGTGAAGCCAGTGGTCTGGGCAATGACCTTGGGCTTTTACTTCGGCAAGTAGAAGCCACGCGCATTGTTTCAAGT



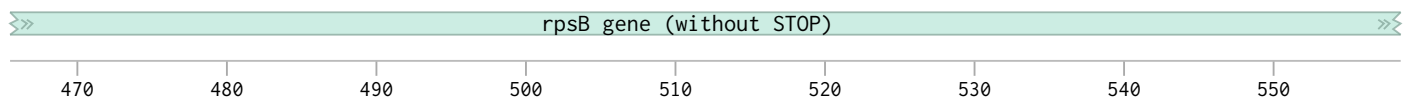
CATCATCAACCTTGAGAAAAGTACCGATGTTCAACGAAGCTCTGGCTGAACTGAACAAGATTGCTTCTCGCAAAGGTAAAATCCTTTTCGT  
GTAGTAGTTGGAAGCTTTTTGACATGGCTACAAGTTGCTTCGAGACCGACTTGACTTGTCTAACGAAGAGCGTTTCCATTTTAGGAAAAGCA



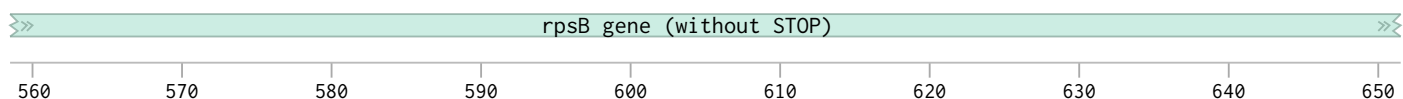
TGGTACTAAACGCGCTGCAAGCGAAGCGGTGAAAGACGCTGCTCTGAGCTGCGACCAGTTCTTCGTGAACCATCGCTGGCTGGGCGGTATGCT  
ACCATGATTTGCGCGACGTTTCGCTTCGCCACTTTCTGCGACGAGACTCGACGCTGGTCAAGAAGCACTTGGTAGCGACCGACCCGCCATACGA



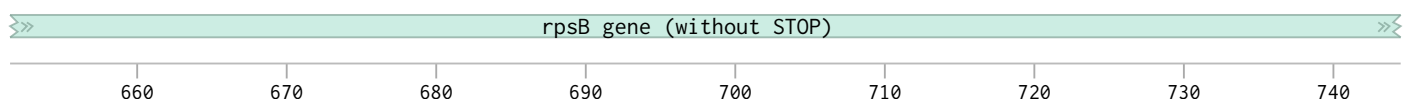
GACTAACTGAAAACCGTTCGTCACTCCATCAAACGTCTGAAAGACCTGAAAAGTCACTCTCAGGACGGTACTTTTCGACAAGCTGACCAAGAA  
CTGATTGACCTTTTGGCAAGCAGTCAGGTAGTTTGCAGACTTTCTGGACCTTTGAGTCAGAGTCTGCCATGAAAGCTGTTTCGACTGTTTCTT



AGAAGCGCTGATGCGCACTCGTGAGCTGGAGAACTGGAAAACAGCCTGGGCGGTATCAAAGACATGGGCGGTCTGCCGACGCTCTGTTTGT  
TCTTCGCGACTACGCGTGAAGCACTCGACCTCTTTGACCTTTTGTGCGACCCGCCATAGTTTCTGTACCCGCCAGACGGCCTGCGAGACAAACA



AATCGATGCTGACCACGAACACATTGCTATCAAAGAAGCAAACAACCTGGGTATTCCGGTATTTGCTATCGTTGATACCAACTCTGATCCGGA  
TTAGCTACGACTGGTGTGTTGTAACGATAGTTTCTTCGTTTGTGGACCCATAAGGCCATAAACGATAGCAACTATGGTTGAGACTAGGCCCT



CGGTGTTGACTTCGTTATCCCGGGTAACGACGACGCAATCCGTGCTGTGACCCTGTACCTGGGCGCTGTTGCTGCAACCGTACGTGAAGGCCG  
GCCACAACCTGAAGCAATAGGGCCATTGCTGCTGCGTTAGGCACGACACTGGGACATGGACCCGCGACAACGACGTTGGCATGCACTTCCGGC

»» rpsB gene (without STOP) »»

750 760 770 780 790 800 810 820 830

TTCTCAGGATCTGGCTTCCAGGCCGAAGAAAGCTTCGTAGAAGCTGAGCAGGAAAGCGACAGGAGAGTAAAGGAGAAGAACTTTTCACTGG  
AAGAGTCCTAGACCGAAGGGTCCGCCTTCTTTGAAAGCATCTTCGACTCGTCTTTCCGCTGTCTCTCATTTCCTTCTTTGAAAAGTGACC

»» RpsBfluo\_gfpmut2-left »»

Primer Bakshi yfp-kan (L)  
»» rpsB gene (without STOP) linker gfp\_mut2 (without ATG) »»

840 850 860 870 880 890 900 910 920 930

AGTTGTCCCAATTCTTGTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCACTGGAGAGGGTGAAGGTGATGCAACATACGGAAA  
TCAACAGGGTTAAGAACAACCTAATCTACCACTACAATTACCCGTGTTTAAAGACAGTCACCTCTCCCACTTCCACTACGTTGTATGCCTTT

»» gfp\_mut2 (without ATG) »»

940 950 960 970 980 990 1,000 1,010 1,020

ACTTACCCTTAAATTTATTTGCACTACTGGAAAACCTACTGTTCCATGGCCAACACTTGTCACTACTTTTCGCGTATGGTCTTCAATGCTTTGC  
TGAATGGGAATTTAATAAACGTGATGACCTTTTGTGACAAAGGTACCGGTTGTGAACAGTGATGAAAGCGCATACCAGAAGTTACGAAACG

»» gfp\_mut2 (without ATG) »»

1,030 1,040 1,050 1,060 1,070 1,080 1,090 1,100 1,110

GAGATACCCAGATCATATGAAACAGCATGACTTTTTCAAGAGTGCCATGCCCCAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGA  
CTCTATGGGTCTAGTATACTTTGTCGTAAGTAAAGTTCTCACGGTACGGGCTTCCAATACATGTCCTTTCTTGATATAAAAAGTTTCTACT

»» gfp\_mut2 (without ATG) »»

1,120 1,130 1,140 1,150 1,160 1,170 1,180 1,190 1,200

CGGGAACACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAGGTATTGATTTTTAAAGAAGATGG  
GCCCTTGATGTTCTGTGCACGACTTCAGTTCAAACCTCCACTATGGGAACAATTATCTTAGCTCAATTTCCATAACTAAAATTTCTTCTACC

»» gfp\_mut2 (without ATG) »»

1,210 1,220 1,230 1,240 1,250 1,260 1,270 1,280 1,290 1,300

AAACATTCTTGACACAAATTGGAATACAACCTATAACTCACACAATGTATACATCATGCCAGACAAACAAAAGAATGGAATCAAAGTTAACTT  
TTTGTAAAGAACCTGTGTTTAACTTATGTTGATATTGAGTGTGTACATATGTAGTACCGTCTGTTTCTTACCTTAGTTTCAATTGAA

»» gfp\_mut2 (without ATG) »»

1,310 1,320 1,330 1,340 1,350 1,360 1,370 1,380 1,390

CAA AATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGCCCTGTCCTTTTACC  
GTTTTAATCTGTGTTGTAACCTTCTACCTTCGCAAGTTGATCGTCTGGTAATAGTTGTTTTATGAGGTTAACCGCTACCGGGACAGGAAAATGG

>> gfp\_mut2 (without ATG) >>

1,400 1,410 1,420 1,430 1,440 1,450 1,460 1,470 1,480

AGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAAACAGCTGC  
TCTGTTGTAATGGACAGGTGTGTAGACGGGAAAGCTTCTAGGGTTGCTTTTCTCTCTGGTGTACCAGGAAGAACTCAAACATTGTGCGACG

>> gfp\_mut2 (without ATG) >>

1,490 1,500 1,510 1,520 1,530 1,540 1,550 1,560 1,570 1,580

RpsBfluo\_cassette-left (gfpmut2 and cfp)

TGGGATTACACATGGCATGGATGAACTATACAAATAATTATATTCCCCAGAACATCAGGTTAATGGCGTTTTTGTATGTCATTTTCGCGGTGGC  
ACCCTAATGTGTACCGTACCTACTTGATATGTTTATTAATAAAGGGTCTTGTAGTCCAATTACCGCAAAAACACTACAGTAAAAGCGCCACCG

RpsBfluo\_cfp/gfpmut2-right

>> gfp\_mut2 (without ATG) <<< ccdB  
<<< Cassette <<<

1,590 1,600 1,610 1,620 1,630 1,640 1,650 1,660 1,670

TGAGATCAGCCACTTCTCCCCGATAACGGAGACCGGCACACTGGCCATATCGGTGGTCATCATGCGCCAGCTTTCATCCCCGATATGCACCA  
ACTCTAGTCGGTGAAGAAGGGGCTATTGCCTCTGGCCGTGTGACCGGTATAGCCACCAGTAGTACGCGGTCGAAAGTAGGGGCTATACGTGGT

<<< ccdB <<<  
<<< Cassette <<<

1,680 1,690 1,700 1,710 1,720 1,730 1,740 1,750 1,760

CCGGGTAAAGTTCACGGGAGACTTTATCTGACAGCAGACGTGCACTGGCCAGGGGGATCACCATCCGTGCGCCGGGCGTGTCAATAATATCAC  
GGCCATTTCAAGTGCCTCTGAAATAGACTGTCTGTCTGCACGTGACCGGTCCCCCTAGTGGTAGGCAGCGGGCCCGCACAGTTATTATAGT

<<< ccdB <<<  
<<< Cassette <<<

1,770 1,780 1,790 1,800 1,810 1,820 1,830 1,840 1,850 1,860

TCTGTACATCCACAAACAGACGATAACGGCTCTCTCTTTTATAGGTGTAACCTTAAACTGCATCGTTTCACTCCATCCAAAAAACGGGTAT  
AGACATGTAGGTGTTTGTCTGCTATTGCCGAGAGAGAAAATATCCACATTTGGAATTTGACGTAGCAAAGTGAGGTAGGTTTTTTTGGCCATA

<<< ccdB <<<  
<<< Cassette <<<

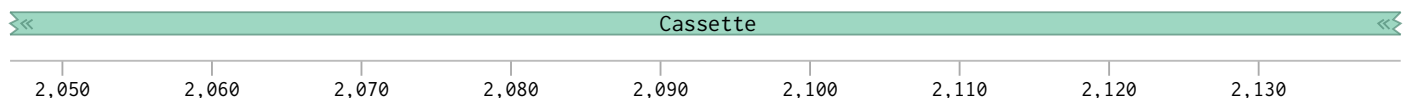
1,870 1,880 1,890 1,900 1,910 1,920 1,930 1,940 1,950

GGAGAAACAGTAGAGAGTTGCGATAAAAAGCGTCAGGTAGGATCCGCTAATCTTATGGATAAAAATGCTATGGCATAGCAAAGTGTGACGCC  
CCTCTTTGTCATCTCTCAACGCTATTTTTTCGACGTCCATCCTAGGCGATTAGAATACCTATTTTTACGATACCGTATCGTTTCACACTGCGGC

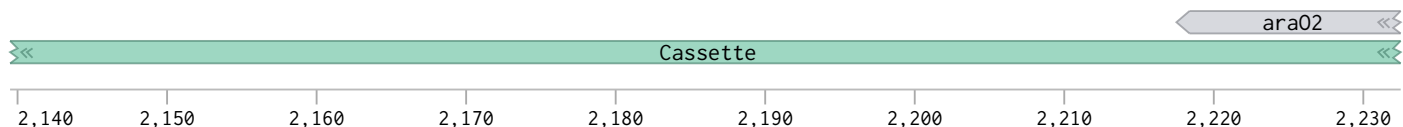
<<< araI2 araI1  
<<< pBAD <<<  
<<< Cassette <<<

1,960 1,970 1,980 1,990 2,000 2,010 2,020 2,030 2,040

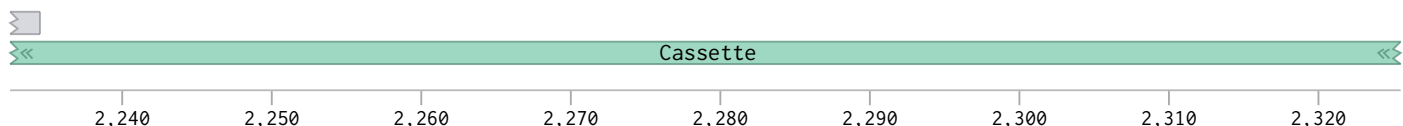
TGCAAATAATCAATGTGGACTTTTCTGCCGTGATTATAGACTTTTGTACGCGTTTTTGTGCATGGCTTTGGTCCCGCTTTGTTACAGAATG  
ACGTTTATTAGTTACACCTGAAAAGACGGCACTAATATCTGTGAAAACAATGCGCAAAAACAGTACCGAAACCAGGGCGAAACAATGTCTTAC



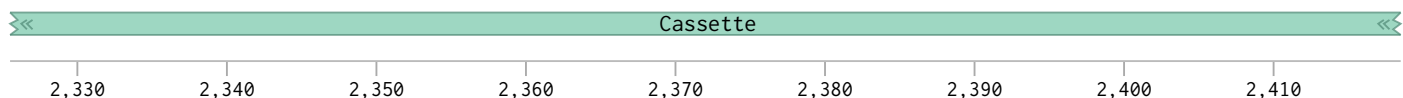
CTTTAATAAGCGGGTTACCGTTGGGTTAGCGAGAAGAGCCAGTAAAAGACGCAGTGACGGCAATGTCTGATGCAATATGGACAATTGGTT  
GAAAATTATTGCCCCAATGGCCAACCCAATCGCTTCTCGGTCATTTTCTGCGTACTGCCGTTACAGACTACGTTATACCTGTTAACCAA



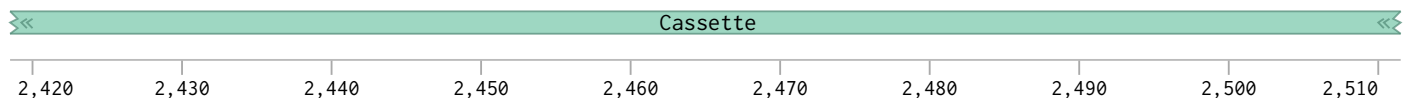
TCTTCTCTGAATGGTGGGAGTATGAAAAGTATGGCTGAAGCGCAAATGATCCCCTGCTGCCGGGATACTCGTTTAAACGCCCTTTGAAGCTC  
AGAAGAGACTTACCACCCTCATACTTTTCATACCGACTTCGCGTTTTACTAGGGGACGACGGCCCTATGAGCAAATTGCGGGAAAACCTTCGAG



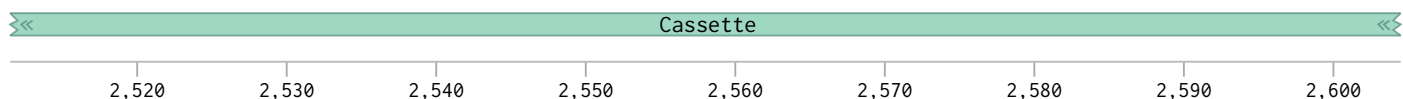
ACGCTGCCGAAGCACTCAGGGCGCAAGGGCTGCTAAAGGAAGCGGAACACGTAGAAAGCCAGTCCGCAGAAACGGTGCTGACCCCGGATGAA  
TGCGACGGCGTTTCGTGAGTCCCGCGTTCCCGACGATTTCTTCGCCTTGTGCATCTTTCGGTCAGGCGTCTTTCGCCACGACTGGGGCCTACTT



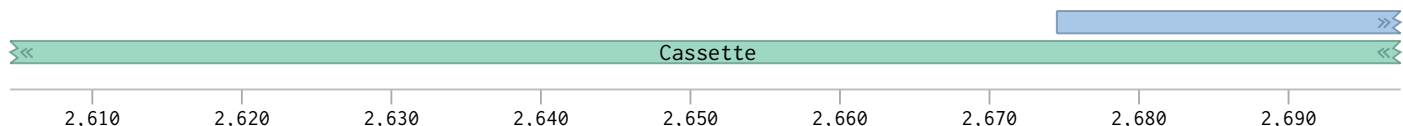
TGTCAGCTACTGGGCTATCTGGACAAGGGAAAACGCAAGCGCAAAGAGAAAGCAGGTAGCTTGCAGTGGGCTTACATGGCGATAGCTAGACTG  
ACAGTCGATGACCCGATAGACCTGTTCCCTTTTTCGTTTCGCGTTTTCTTTTCGTCATCGAACGTCACCCGAATGTACCGCTATCGATCTGAC



GGCGTTTTATGGACAGCAAGCGAACCGGAATTGCCAGCTGGGGCGCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAACTGGATGGCTTT  
CCGCCAAAATACCTGTGCTTCGCTTGGCCTTAACGGTGCACCCCGCGGGAGACCATTCCAACCTTCGGGACGTTTCATTTGACCTACCGAAA



CTTGCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCA  
GAACGGCGGTTCTAGACTACCGCTCCCCTAGTTCTAGACTAGTTCTCTGTCCTACTCCTAGCAAAGCGTACTAACTTGTTCCTACCTAACGT



CGCAGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCT  
GCGTCCAAGAGGCCGGCAACCCACCTCTCCGATAAGCCGATACTGACCCGTGTTGTCTGTTAGCCGACGAGACTACGGCGGCACAAGGCCGA

KnR (kanamycin resistance) >>>  
Cassette <<<

2,700 2,710 2,720 2,730 2,740 2,750 2,760 2,770 2,780 2,790

GTCAGCGCAGGGGCGCCCGTTCTTTTTGTCAAGACCGACCTGTCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGGGCTATCGTGGCT  
CAGTCGGTCCCGCGGGCCAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTCTGCTCCGTCGCGCCGATAGCACCGA

KnR (kanamycin resistance) >>>  
Cassette <<<

2,800 2,810 2,820 2,830 2,840 2,850 2,860 2,870 2,880

GGCCACGACGGGCGTTCCTTGCAGCTGTGCTCGACGTTGCTACTGAAGCGGGAAGGGACTGGCTGCTATTGGCGAAGTGCCGGGGCAGGA  
CCGGTGTGCCCCGAAGAACGCGTCGACACGAGCTGCAACAGTGACTTCGCCCTCCCTGACCGACGATAACCCGTTACGGCCCCGTCTCT

KnR (kanamycin resistance) >>>  
Cassette <<<

2,890 2,900 2,910 2,920 2,930 2,940 2,950 2,960 2,970

TCTCCTGCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCC  
AGAGGACAGTAGAGTGGAACGAGGACGGCTCTTTCATAGGTAGTACCGACTACGTTACGCCGCCGACGTATGCGAACTAGGCCGATGGACGG

KnR (kanamycin resistance) >>>  
Cassette <<<

2,980 2,990 3,000 3,010 3,020 3,030 3,040 3,050 3,060

ATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAAGAGCATCA  
TAAGCTGGTGGTTCGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACCTTCGGCCAGAACAGCTAGTCCTACTAGACCTGCTTCTCGTAGT

KnR (kanamycin resistance) >>>  
Cassette <<<

3,070 3,080 3,090 3,100 3,110 3,120 3,130 3,140 3,150 3,160

GGGGCTCGCGCCAGCCGAACGTTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCC  
CCCCGAGCGGGTTCGGCTTGACAAGCGGTCCGAGTCCGCGGTACGGGCTGCCGCTCCTAGAGCAGCACTGGGTACCGCTACGGACGAACGG

KnR (kanamycin resistance) >>>  
Cassette <<<

3,170 3,180 3,190 3,200 3,210 3,220 3,230 3,240 3,250

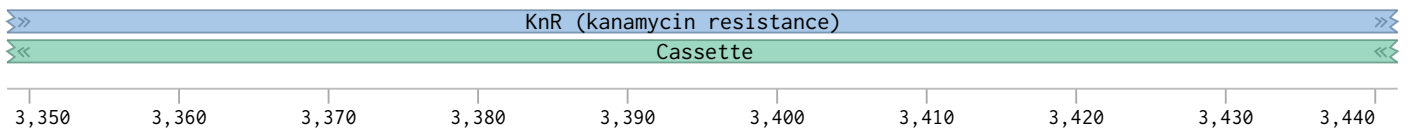
GAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCCTATCAGGACATAGCGTTGGCTAC  
CTTATAGTACCACCTTTTACCGGCGAAAAGACCTAAGTAGCTGACACCGGCCACCCACACCGCCTGGCGATAGTCTGTATCGCAACCGATG

KnR (kanamycin resistance) >>>  
Cassette <<<

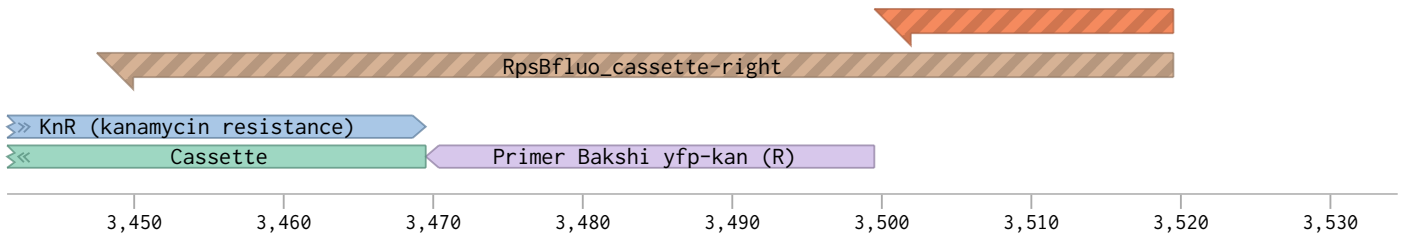
3,260 3,270 3,280 3,290 3,300 3,310 3,320 3,330 3,340



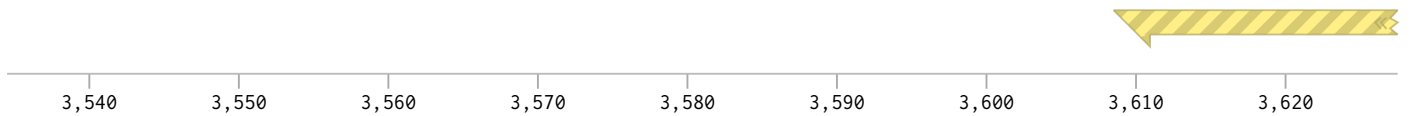
CCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCGTGCTTTACGGTATCGCCGCTCCCGATTTCGAGCGCATCGCCTT  
GGCACTATAACGACTTCTCGAACC GCGCTTACCCGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAA



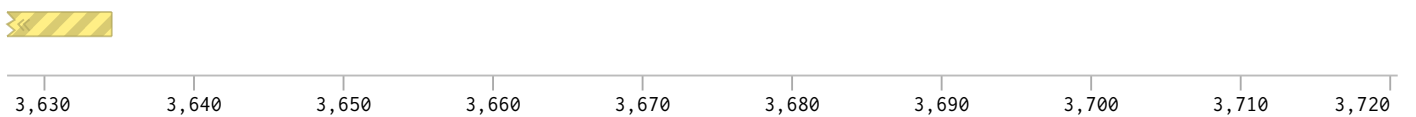
CTATCGCCTTCTTGACGAGTTCTTCTAAGCTTGATAACTCCCCAAAATAGTTCGAGTTGCAGAAAGGCGGCAAGCTCGAGAATTCCCGGGAG  
GATAGCGGAAGAAGCTGCTCAAGAAGATTGAACTATTGAGGGGGTTTTATCAAGCTCAACGTCTTTCGCCGTTTCGAGCTCTTAAGGGCCCTC



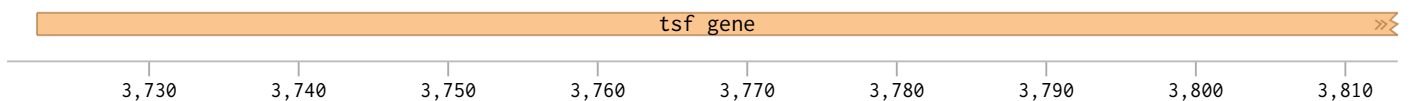
CTTACATCAGTAAGTGACCGGGATGAGCGAGCGAAGATAACGCATCTGCGGCGGAAAATGAAGGGGGAGAGCCCTTATAGACCAGGTAGTA  
GAATGTAGTCATTCAGTGGCCCTACTCGCTCGCTTCTATTGCGTAGACGCCGCGCTTATACTTCCCCCTCTCGGAATATCTGGTCCATCAT



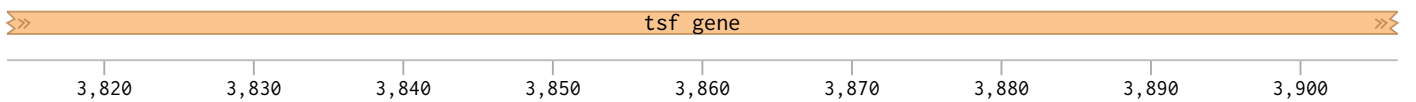
CACGTTTGGTTAGGGGGCTGCATATGGCCCCCTTTTCACTTTTATATCTGTGCGGTTAATGCCGGGCAGATCACATCTCCGAGGATTTTA  
GTGCAAACCAATCCCCGGACGTATACCGGGGGAAAAAGTAAAATATAGACACGCCAAATTACGGCCCGTCTAGTGTAGAGGCTCCTAAAAA



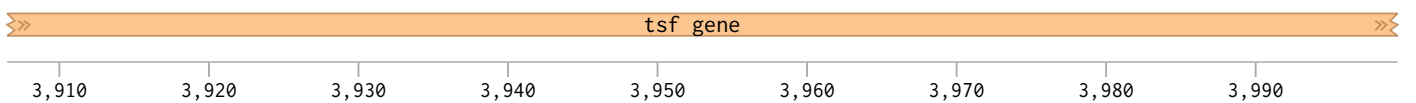
GAATGGCTGAAATTACCGCATCCCTGGTAAAAGAGCTGCGTGAGCGTACTGGCGCAGGCATGATGGATTGCAAAAAAGCACTGACTGAAGCTA  
CTTACCGACTTTAATGGCGTAGGGACCATTTTCTCGACGCACTCGCATGACCGCGTCCGTAACGTTTTTTTCGTGACTGACTTCGAT



ACGGCGACATCGAGCTGGCAATCGAAAACATGCGTAAGTCCGGTGCTATTAAGCAGCGAAAAAAGCAGGCAACGTTGCTGCTGACGGCGTGA  
TGCCGCTGTAGCTCGACCGTTAGCTTTTGTACGCATTACAGCCACGATAATTTTCGTCGCTTTTTTCGTCCGTTGCAACGACGACTGCCGCACT



TCAAACCAAAATCGACGGCAACTACGGCATCATTCTGGAAGTTAACTGCCAGACTGACTTCGTTGCAAAGACGCTGGTTTCCAGGCGTTTCG  
AGTTTTGGTTTTAGCTGCCGTTGATGCCGTAGTAAGACCTCAATTGACGGTCTGACTGAAGCAACGTTTTCTGCGACCAAAGGTCCGCAAGC



CAGACAAAGTTCTGGACGCAGCTGTTGCTGGCAAATCACTGACGTTGAAGTTCTGAAAGCACAGTTCGAAGAAGAACGTGTTGCGCTGGTAG  
GTCTGTTTCAAGACCTGCGTCGACAACGACCGTTTTAGTGACTGCAACTTCAAGACTTTCGTGTCAAGCTTCTTCTGCACAACGCGACCATC

tsf gene

4,000 4,010 4,020 4,030 4,040 4,050 4,060 4,070 4,080 4,090

CGAAAATTGGTGAAAACATCAACATTCGCCGCGTTGCTGCGCTGGAAGGCGACGTTCTGGGTTCTTATCAGCACGGTGCGCGTATCGGCGTTC  
GCTTTTAACCACTTTTGTAGTTGTAAGCGGCGCAACGACGCGACCTTCCGCTGCAAGACCCAAGAATAGTCGTGCCACGCGCATAGCCGCAAG

tsf gene

4,100 4,110 4,120 4,130 4,140 4,150 4,160 4,170 4,180

TGTTGCTGCTAAAGGCGCTGACGAAGAGCTGGTTAAACACATCGCTATGCACGTTGCTGCAAGCAAGCCAGAATTCATCAAACCGGAAGACG  
ACCAACGACGATTTCCGCGACTGCTTCTCGACCAATTTGTGTAGCGATACGTGCAACGACGTTTCGTTCCGTCTTAAGTAGTTTGGCCTTCTGC

tsf gene

4,190 4,200 4,210 4,220 4,230 4,240 4,250 4,260 4,270

TATCCGCTGAAGTGGTAGAAAAAGAATACCAGGTACAGCTGGATATCGCGATGCAGTCTGGTAAGCCGAAAGAAATCGCAGAGAAAATGGTTG  
ATAGGCGACTTCACCATCTTTTTCTTATGGTCCATGTGACCTATAGCGCTACGTGACACCATTTCGGCTTTCTTTAGCGTCTCTTTTACCAAC

tsf gene

4,280 4,290 4,300 4,310 4,320 4,330 4,340 4,350 4,360 4,370

AAGGCCGCATGAAGAAATTCACCGCGAAGTTTCTCTGACCGGTCAGCCGTTTCGTTATGGAACCAAGCAAACTGTTGGTCAGCTGCTGAAAG  
TTCCGCGTACTTCTTTAAGTGGCCGCTTCAAAGAGACTGGCCAGTCGGCAAGCAATACCTTGGTTCGTTTTGACAACCAGTCGACGACTTTC

tsf gene

4,380 4,390 4,400 4,410 4,420 4,430 4,440 4,450 4,460

AGCATAACGCTGAAGTGACTGGCTTCATCCGCTTCGAAGTGGGTGAAGGCATCGAGAAAGTTGAGACTGACTTTGCAGCAGAAGTTGCTGCGA  
TCGTATTGCGACTTCACTGACCGAAGTAGGCGAAGCTTCACCACTTCCGTAGCTCTTTCAACTCTGACTGAAACGTCGTCTTCAACGACGCT

tsf gene

4,470 4,480 4,490 4,500 4,510 4,520 4,530 4,540 4,550

TGTCCAAGCAGTCTTAA

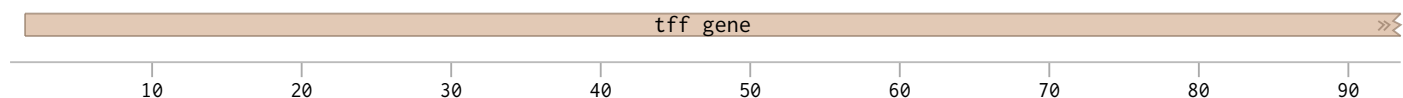
ACAGGTTTCGTCAGAATT

tsf gene

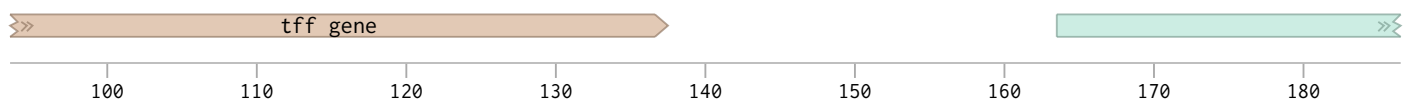
4,560 4,570

## mCherry-cassette after rpsB (E. coli) (4577 bp)

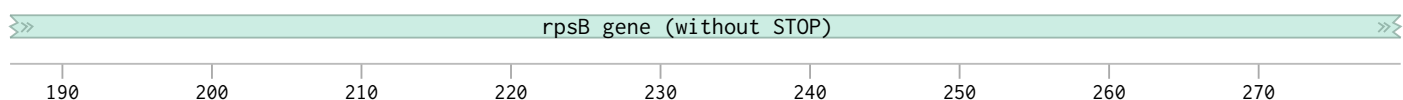
CCGGA CTTCGATCCATTTCTGATACACAGACTGGACGGAAGCGACAATCTCACTTTGTGTAACAACACACACGTATCGGCACATATTCCGGG  
GGCCTGAAGGCTAGGTAAAGCATATGTGTCTGACCTGCCTTCGCTGTTAGAGTGAAACACATTGTTGTGTGTGCATAGCCGTGTATAAGGCC



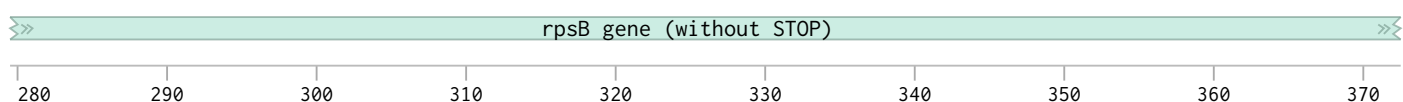
GTGCCCTTTGGGGTCGGTAATATGGGATACGTGGAGGCATAACCCCACTTTTATATAGAGGTTTTAATCATGGCAACTGTTTCCATGCGCGA  
CACGGGAAACCCAGCCATTATACCTATGCACCTCCGTATTGGGGTTGAAAATATATCTCCAAAATTAGTACCGTTGACAAAGGTACGCGCT



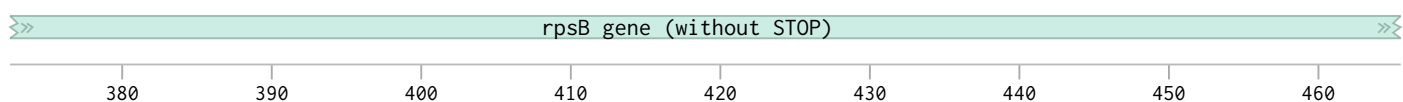
CATGCTCAAGGCTGGTGTTCACCTTCGGTCACCAGACCCGTTACTGGAACCCGAAAATGAAGCCGTTTCATCTTCGGTGC GCGTAACAAAGTTCA  
GTACGAGTTCCGACCACAAGTGAAGCCAGTGGTCTGGGCAATGACCTTGGGCTTTTACTTCGGCAAGTAGAAGCCACGCGCATTGTTTCAAGT



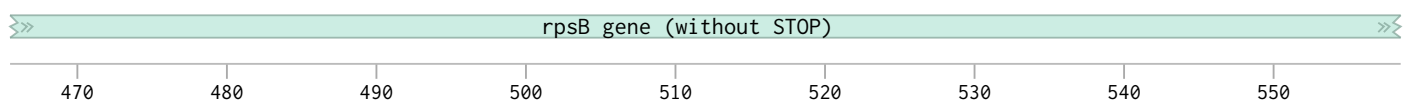
CATCATCAACCTTGAGAAAAGTACCGATGTTCAACGAAGCTCTGGCTGAACTGAACAAGATTGCTTCTCGCAAAGGTAAAATCCTTTTCGT  
GTAGTAGTTGGAAGCTTTTTGACATGGCTACAAGTTGCTTCGAGACCGACTTGACTTGTCTAACGAAGAGCGTTTCCATTTTAGGAAAAGCA



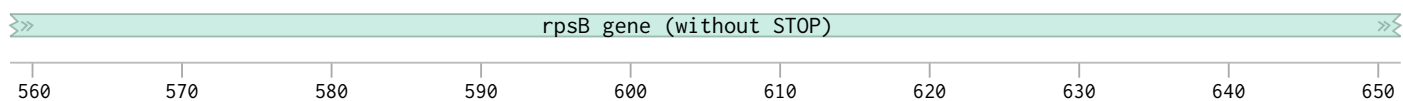
TGGTACTAAACGCGCTGCAAGCGAAGCGGTGAAAGACGCTGCTCTGAGCTGCGACCAGTTCTTCGTGAACCATCGCTGGCTGGGCGGTATGCT  
ACCATGATTTGCGCGACGTTTCGCTTCGCCACTTTCTGCGACGAGACTCGACGCTGGTCAAGAAGCACTGGTAGCGACCGACCCGCCATACGA



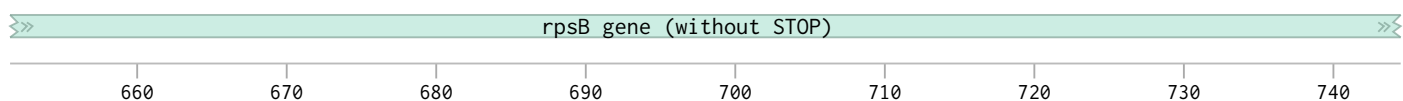
GACTAACTGGAAAACCGTTCGTCAGTCCATCAAACGTCTGAAAGACCTGGAAACTCAGTCTCAGGACGGTACTTTGACAAGCTGACCAAGAA  
CTGATTGACCTTTTGGCAAGCAGTCAGGTAGTTTGCAGACTTTCTGGACCTTTGAGTCAGAGTCTGCCATGAAAGCTGTTGACTGTTTCTT



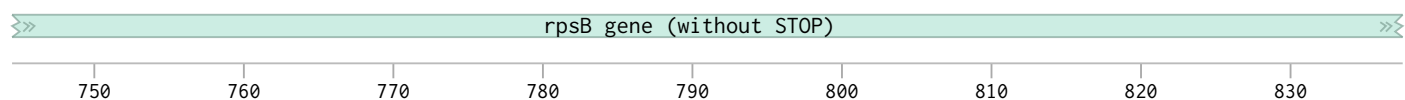
AGAAGCGCTGATGCGCACTCGTGAGCTGGAGAACTGGAAAACAGCCTGGGCGGTATCAAAGACATGGGCGGTCTGCCGGACGCTCTGTTTGT  
TCTTCGCGACTACGCGTGAGCACTCGACCTCTTTGACCTTTTGTGCGACCCGCCATAGTTTCTGTACCCGCCAGACGGCCTGCGAGACAAACA



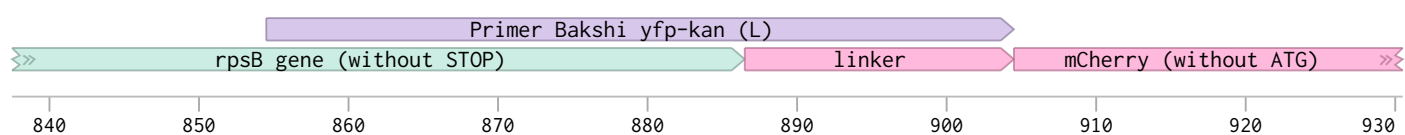
AATCGATGCTGACCACGAACACATTGCTATCAAAGAAGCAAACAACCTGGGTATTCCGGTATTTGCTATCGTTGATACCAACTCTGATCCGGA  
TTAGCTACGACTGGTGTGTTGTAACGATAGTTTCTTCGTTTGTGGACCCATAAGGCCATAAACGATAGCAACTATGGTTGAGACTAGGCCCT



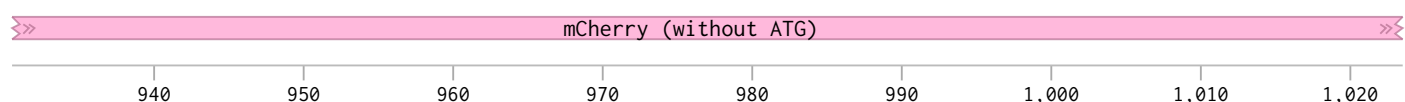
CGGTGTTGACTTCGTTATCCCGGGTAACGACGACGCAATCCGTGCTGTGACCCTGTACCTGGGCGCTGTTGCTGCAACCGTACGTGAAGGCCG  
GCCACAACCTGAAGCAATAGGGCCATTGCTGCTGCGTTAGGCACGACACTGGGACATGGACCCGCGACAACGACGTTGGCATGCACTTCCGGC



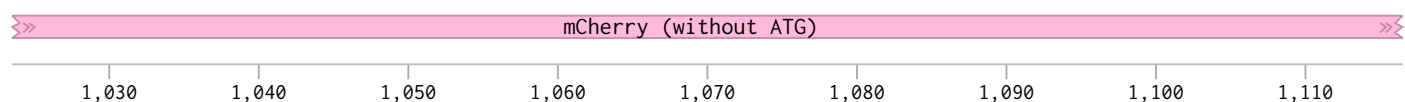
TTCTCAGGATCTGGCTTCCAGGCGGAAGAAAGCTTCGTAGAAGCTGAGCAGGAAAGGCGACAGGAGACTAGCAAAGATCCAAGGGCGAGGA  
AAGAGTCCTAGACCGAAGGGTCCGCCTTCTTTGAAGCATCTTCGACTCGTCTTCCGCTGTCTCTGATCGTTTTCTAGGTTCCCGCTCCT



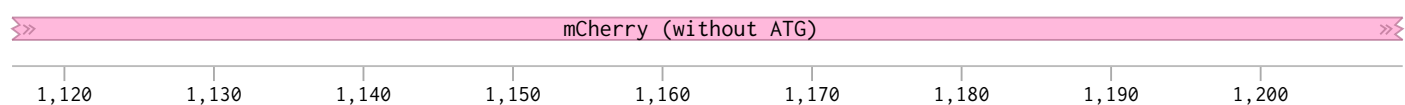
GGATAACATGGCTATCATTAAAGAGTTCATGCGCTTCAAAGTTCACATGGAGGGTTCTGTAAACGGTCACGAGTTCGAGATCGAAGGCGAAGG  
CCTATTGTACCGATAGTAATTTCTCAAGTACGCGAAGTTTCAAGTGTACCTCCAAGACAATTGCCAGTGCTCAAGCTCTAGCTTCCGCTTCC



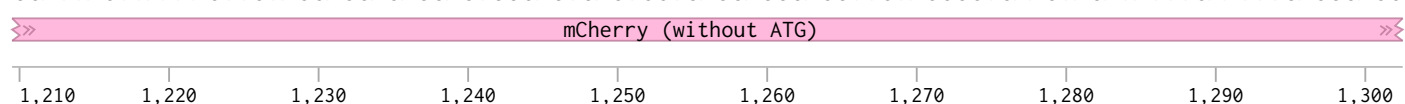
CGAGGGCCGTCCGTATGAAGGCACCCAGACCGCAAACCTGAAAGTACTAAAGGCGGCCCGCTGCCTTTTGCCTGGGACATCCTGAGCCCGCA  
GCTCCCGCAGGCATACTTCCGTGGGTCTGGCGGTTTACTTTCACTGATTTCCGCCGGCGACGAAAACGCACCCTGTAGGACTCGGGCGT



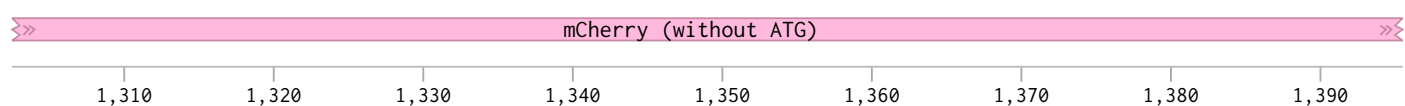
ATTTATGTACGGTTCTAAAGCGTATGTTAAACACCCAGCGGATATCCCGACTATCTGAAGCTGTCTTTTCCGGAAGGTTTCAAGTGGGAACG  
TAAATACATGCCAAGATTTTCGCATACAATTTGTGGGTGCGCTATAGGGCCTGATAGACTTCGACAGAAAAGGCCTTCAAAGTTTACCCTTGC



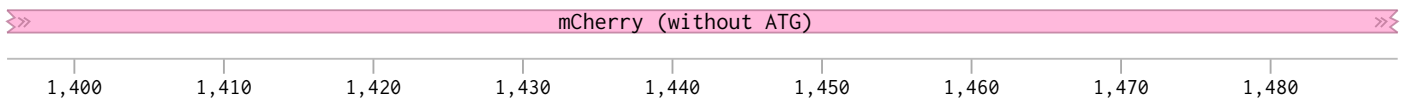
CGTAATGAATTTTGAAGATGGTGGTGTGCTGACCGTCACTCAGGACTCCTCCCTGCAAGATGGCGAGTTCATCTATAAAGTTAAACTGCGTGG  
GCATTACTTAAAATTCTACCACCACAGCACTGGCAGTGAGTCTGAGGAGGGACGTTCTACCGCTCAAGTAGATATTTCAATTTGACGCACC



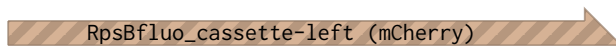
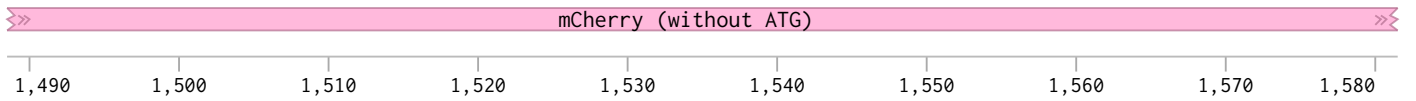
TACTAATTTTCCATCTGATGGCCCGTGATGCAGAAAAGACGATGGGTTGGGAGGCGTCTAGCGAACGCATGTATCCGGAAGATGGTGCCT  
ATGATTAAGGTAGACTACCGGGCCACTACGTCTTTTTCTGCTACCAACCCTCCGAGATCGTTGCGTACATAGGCCTTCTACCACGCGA



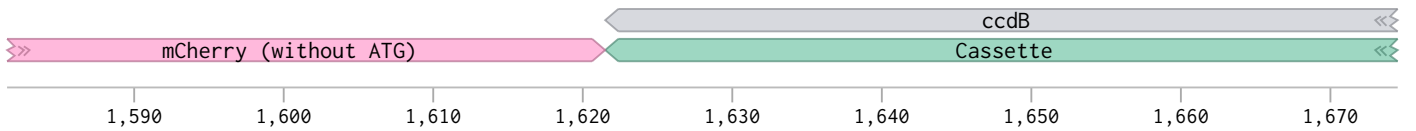
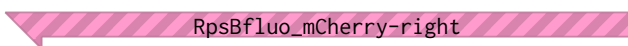
GAAAGGCGAAATTAACAGCGCCTGAAACTGAAAGATGGCGGCCATTATGACGCTGAAGTAAAACACGTACAAAGCCAAGAAACCTGTGCA  
CTTCCGCTTTAATTTGTCGCGGACTTTGACTTTCTACCGCGTAATACTGCGACTTCACTTTTGGTGCATGTTTCGGTTCCTTTGGACACGT



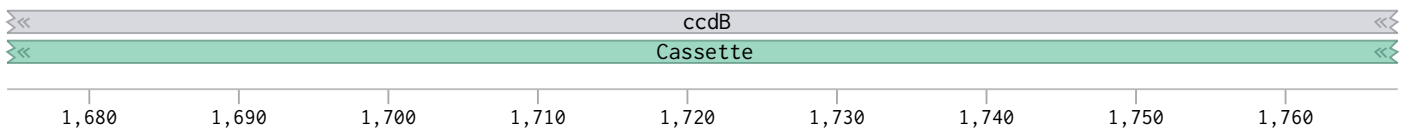
GCTGCCTGGCGGTACAATGTGAATATTAACCTGGACATCACCTCTCATAATGAAGATTATACGATCGTAGAGCAATATGAGCGCGGGAGGG  
CGACGGACCGCGCATGTTACACTTATAATTTGACCTGTAGTGGAGAGTACTTCTAATATGCTAGCATCTCGTTATACTCGCGCGCTCCC



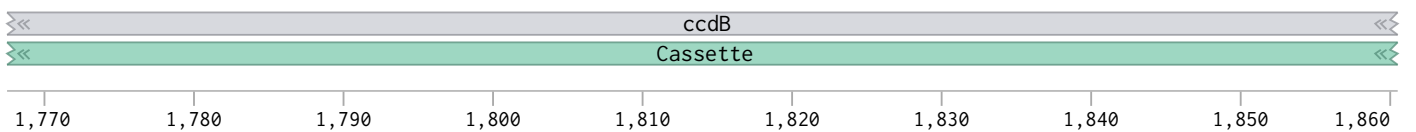
TCGTCACTTACCGGTGGCATGGATGAGCTGTACAAATAATTATATTTCCCAGAACATCAGGTTAATGGCGTTTTTGGTGCATTTTCGCGGT  
AGCAGTAAGATGGCCACCGTACCTACTCGACATGTTTATTAATATAAGGGTCTTGTAGTCCAATTACCGCAAAAACACAGTAAAAGCGCCA



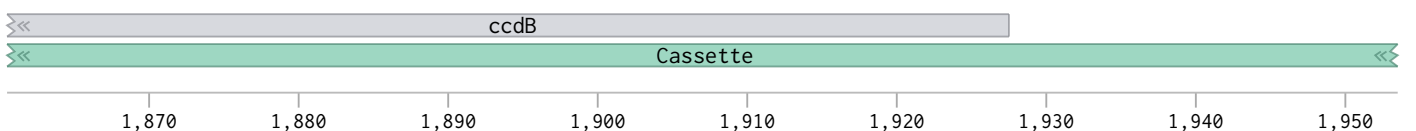
GGCTGAGATCAGCCACTTCTCCCGATAACGGAGACCGGCACACTGGCCATATCGGTGGTCATCATGCGCCAGCTTTCATCCCCGATATGCA  
CCGACTCTAGTCGGTGAAGAAGGGGCTATTGCCTCTGGCCGTGTACCGGTATAGCCACCAGTAGTACGCGGTGAAAGTAGGGGCTATACGT



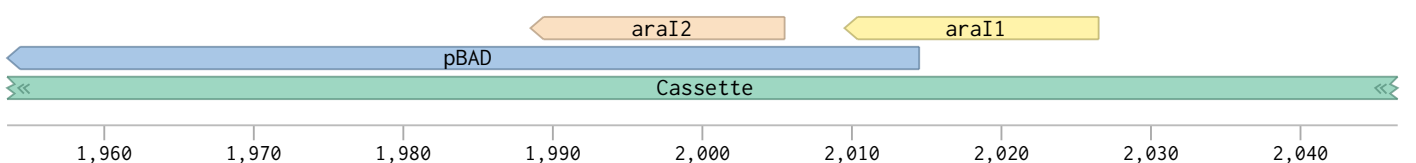
CCACCGGTAAAGTTCACGGGAGACTTTATCTGACAGCAGACGTGCACTGGCCAGGGGGATCACCATCCGTCGCCCGGGCGTGAATAATAT  
GGTGGCCATTTCAAGTGCCCTCTGAAATAGACTGTCGTCTGCACGTGACCGGTCCCCTAGTGGTAGGCAGCGGCCCGCACAGTTATTATA



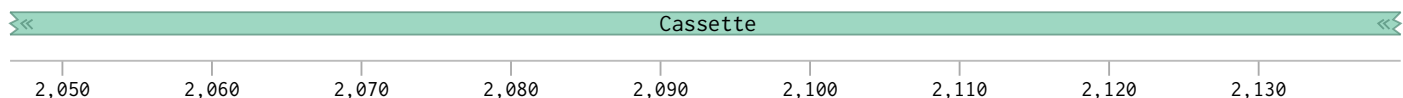
CACTCTGTACATCCACAAACAGACGATAACGGCTCTCTTTTATAGGTGTAACCTTAAACTGCATCGTTTCACTCCATCCAAAAAACGGG  
GTGAGACATGTAGGTGTTTGTCTGCTATTGCCGAGAGAGAAAATATCCACATTTGGAATTTGACGTAGCAAAGTGAGGTAGGTTTTTTGCC



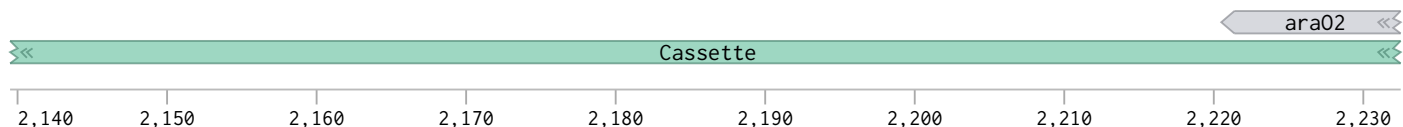
TATGGAGAAACAGTAGAGAGTTGCGATAAAAAGCGTCAGGTAGGATCCGCTAATCTTATGGATAAAAATGCTATGGCATAGCAAAGTGTGACC  
ATACCTCTTTGTCATCTCTCAACGCTATTTTTTCGAGTCCATCCTAGCGGATTAGAATACCTATTTTTACGATACCGTATCGTTTCACACTGC



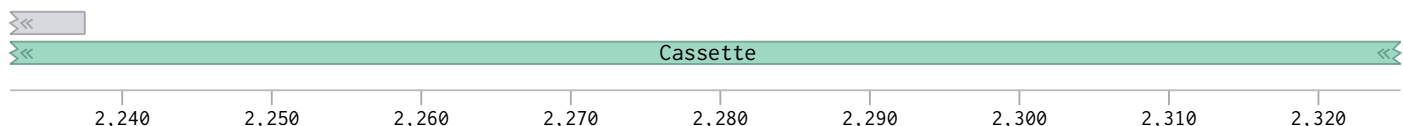
CCGTGCAAATAATCAATGTGGACTTTTCTGCCGTGATTATAGACACTTTTGTACGCGTTTTTGTTCATGGCTTTGGTCCCGCTTTGTTACAGA  
GGCACGTTTATTAGTTACACCTGAAAAGACGGCACTAATATCTGTGAAAACAATGCGCAAAAACAGTACCGAAACCAGGGCGAAACAATGTCT



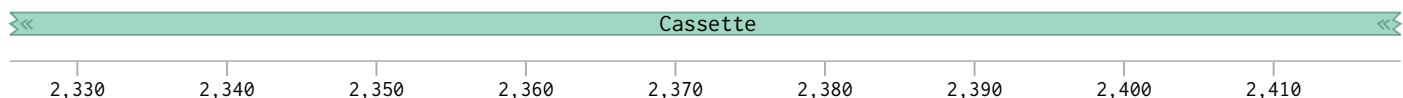
ATGCTTTTAATAAGCGGGTTACCGGTTGGGTTAGCGAGAAGAGCCAGTAAAAGACGCGAGTGACGGCAATGTCTGATGCAATATGGACAATTG  
TACGAAAATTATTCGCCCAATGGCCAACCCAATCGCTTCTCGGTCATTTTCTGCGTCACTGCCGTTACAGACTACGTTATACCTGTTAAC



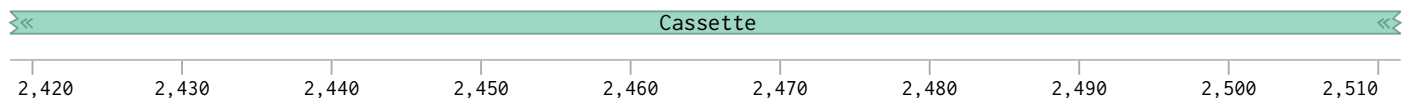
GTTTCTTCTCTGAATGGTGGGAGTATGAAAAGTATGGCTGAAGCGCAAATGATCCCCTGCTGCCGGGATACTCGTTTAAACGCCCTTTTGAAG  
CAAAGAAGAGACTTACCACCCTCATACTTTTCATACCGACTTCGCGTTTTACTAGGGGACGACGGCCCTATGAGCAAATTGCGGGAAACTTC



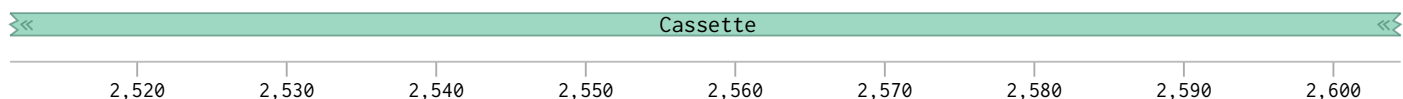
CTCACGCTGCCGAAGCACTCAGGGCGCAAGGGCTGCTAAAGGAAGCGGAACACGTAGAAAAGCCAGTCCGCAGAAACGGTGCTGACCCCGGAT  
GAGTGCACGGCGTTCGTGAGTCCCGCGTTCACGACGATTTCTTCGCTTGTGCATCTTTTCGGTCAGGCGTCTTTGCCACGACTGGGGCCTA



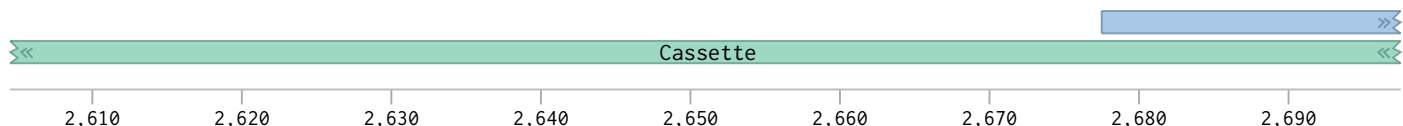
GAATGTCAGCTACTGGGCTATCTGGACAAGGGAAAACGCAAGCGCAAAGAGAAAGCAGGTAGCTTGAGTGGGCTTACATGGCGATAGCTAGA  
CTTACAGTCGATGACCCGATAGACCTGTTCCCTTTTGCCTTCGCGTTTCTTTTCGTCCATCGAACGTCACCCGAATGTACCGCTATCGATCT



CTGGGCGTTTTATGGACAGCAAGCGAACCGGAATTGCCAGCTGGGGCGCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAACTGGATGGC  
GACCCGCCAAAATACCTGTCGTTTCGCTTGGCCTTAACGGTCGACCCCGCGGGAGACCAATCCAACCCTTCGGGACGTTTCATTTGACCTACCG



TTTCTTGCCCCAAGGATCTGATGGCGCAGGGATCAAGATCTGATCAAGAGACAGGATGAGGATCGTTTTCGCATGATTGAACAAGATGGATT  
AAAGAACGGCGGTTCTAGACTACCGCGTCCCCTAGTTCTAGACTAGTTCTCTGTCCTACTCCTAGCAAAGCGTACTAACTTGTCTACCTAA



GCACGCAGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCG  
CGTGCCTCAAGAGGCCGGCGAACCCACCTCTCCGATAAGCCGATACTGACCCGTGTTGTCTGTTAGCCGACGAGACTACGGCGGCACAAGGC

KnR (kanamycin resistance) >>>  
Cassette <<<

2,700 2,710 2,720 2,730 2,740 2,750 2,760 2,770 2,780 2,790

GCTGTCAGCGCAGGGGCGCCCGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGGGCTATCGTG  
CGACAGTCGCGTCCCCGCGGGCAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTCCTGCTCCGTCGCGCCGATAGCAC

KnR (kanamycin resistance) >>>  
Cassette <<<

2,800 2,810 2,820 2,830 2,840 2,850 2,860 2,870 2,880

GCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCA  
CGACCGGTGCTGCCCCAAGGAACGCGTCGACACGAGCTGCAACAGTGACTTCGCCCTTCCTGACCGACGATAACCCGCTTCACGGCCCCGT

KnR (kanamycin resistance) >>>  
Cassette <<<

2,890 2,900 2,910 2,920 2,930 2,940 2,950 2,960 2,970

GGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTG  
CCTAGAGGACAGTAGAGTGAACGAGGACGGCTCTTTCATAGGTAGTACCGACTACGTTACGCCCGGACGTATGCGAACTAGGCCGATGGAC

KnR (kanamycin resistance) >>>  
Cassette <<<

2,980 2,990 3,000 3,010 3,020 3,030 3,040 3,050 3,060

CCCATTGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTA CTGGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAAGAGCA  
GGTAAGCTGGTGGTTGCTTTGTAGCGTAGCTCGCTCGTGATGAGCCTACCTTCGGCCAGAACAGCTAGTCCTACTAGACCTGCTTCTCGT

KnR (kanamycin resistance) >>>  
Cassette <<<

3,070 3,080 3,090 3,100 3,110 3,120 3,130 3,140 3,150 3,160

TCAGGGGCTCGCGCCAGCCGAAGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTT  
AGTCCCCGAGCGCGGTGGCTTGACAAGCGGTCCGAGTTCGCGCGTACGGGCTGCCGCTCCTAGAGCAGCACTGGGTACCGCTACGGACGAA

KnR (kanamycin resistance) >>>  
Cassette <<<

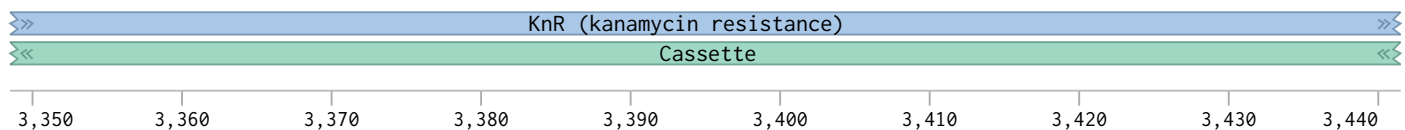
3,170 3,180 3,190 3,200 3,210 3,220 3,230 3,240 3,250

GCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGC  
CGGCTTATAGTACCACCTTTTACCGGCGAAAAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCTGGCGATAGTCCTGTATCGCAACCG

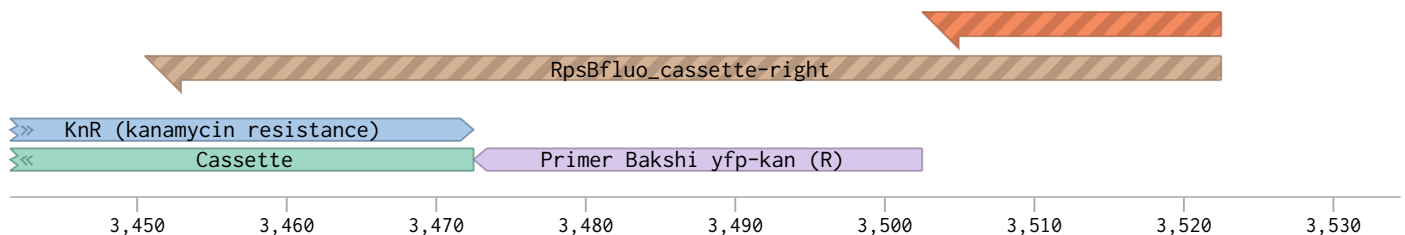
KnR (kanamycin resistance) >>>  
Cassette <<<

3,260 3,270 3,280 3,290 3,300 3,310 3,320 3,330 3,340

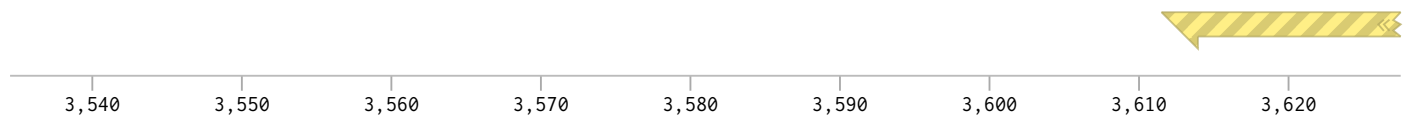
TACCCGTGATATTGCTGAAGAGCTTGCGGGCGAATGGGCTGACCGCTTCTCGTGCTTACGGTATCGCCGCTCCGATTTCGAGCGCATCGC  
 ATGGGCACTATAACGACTTCTCGAACCGCCGTTACCCGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAGCG



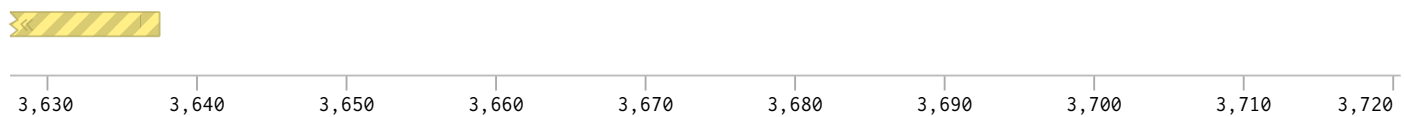
CTTCTATCGCCTTCTTGACGAGTTCTTCTAAGCTTGATAACTCCCCAAAATAGTTCGAGTTGCAGAAAGGCGGCAAGCTCGAGAATCCCCGG  
 GAAGATAGCGGAAGAAGTCTCAAGAAGATTGCAACTATTGAGGGGGTTTTATCAAGCTCAACGTCTTTCGCGCTTCGAGCTCTTAAGGGCC



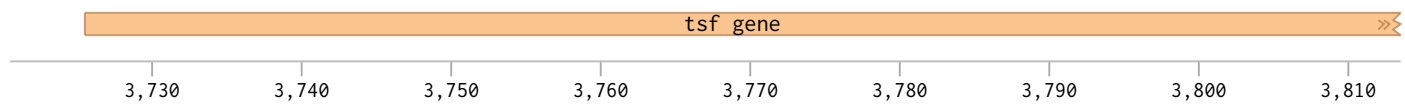
GAGCTTACATCAGTAAGTGACCGGGATGAGCGAGCGAAGATAACGCATCTGCGGCGGAAATATGAAGGGGGAGAGCCCTTATAGACCAGGTA  
 CTCGAATGTAGTCATTCCTACTGGCCCTACTCGCTCGCTTCTATTGCGTAGACGCCGCGCTTTATACTTCCCCCTCTCGGGAATATCTGGTCCAT



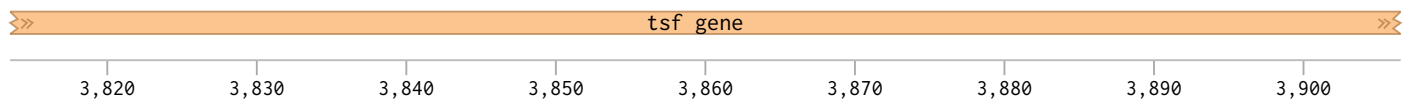
GTACACGTTTTGGTTAGGGGGCCTGCATATGGCCCCCTTTTTCACTTTTATATCTGTGCGGTTTAATGCCGGGCAGATCACATCTCCGAGGATT  
 CATGTGCAAACCAATCCCCCGACGTATACCGGGGAAAAAGTAAAATATAGACACGCCAAATTACGGCCCCGTCTAGTGTAGAGGCTCCTAA



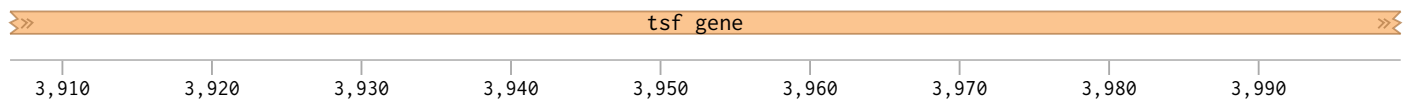
TTAGAATGGCTGAAATTACCGCATCCCTGGTAAAAGAGCTGCGTGAGCGTACTGGCGCAGGCATGATGGATTGCAAAAAAGCACTGACTGAAG  
 AATCTTACCGACTTTAATGGCGTAGGGACCATTTTCTCGACGCACTCGCATGACCGCGTCCGTAACCTTTTTCGTGACTGACTTC



CTAACGGCGACATCGAGCTGGCAATCGAAAACATGCGTAAGTCCGGTGCTATTAAGCAGCGAAAAAGCAGGCAACGTTGCTGCTGACGGCG  
 GATTGCCGCTGTAGCTCGACCGTTAGCTTTTGTACGCATTACAGCCACGATAATTCGTCGCTTTTTTCGTCCGTTGCAACGACGACTGCCGC



TGATCAAAACAAAATCGACGGCAACTACGGCATCATTCTGGAAGTAACTGCCAGACTGACTTCGTTGCAAAGACGCTGGTTTCCAGGCGT  
 ACTAGTTTTGGTTTTAGCTGCCGTTGATGCCGTAGTAAGACCTTCAATTGACGGTCTGACTGAAGCAACGTTTTCTGCGACCAAAGGTCCGCA





TCGCAGACAAAGTTCTGGACGCAGCTGTTGCTGGCAAATCACTGACGTTGAAGTTCTGAAAGCACAGTTCGAAGAAGAACGTGTTGCGCTGG  
AGCGTCTGTTTCAAGACCTGCGTCGACAACGACCGTTTTAGTGACTGCAACTTCAAGACTTTCGTGTCAAGCTTCTTCTGCACAACGCGACC

tsf gene

4,000 4,010 4,020 4,030 4,040 4,050 4,060 4,070 4,080 4,090

TAGCGAAAATTGGTAAAACATCAACATTCGCCGCGTTGCTGCGCTGGAAGGCGACGTTCTGGGTTCTTATCAGCACGGTGCGCGTATCGGCCG  
ATCGCTTTTAACCACTTTTGTAGTTGTAAGCGGCGCAACGACGCGACCTTCCGCTGCAAGACCCAAGAATAGTCGTGCCACGCGCATAGCCGC

tsf gene

4,100 4,110 4,120 4,130 4,140 4,150 4,160 4,170 4,180

TTCTGGTTGCTGCTAAAGGCGCTGACGAAGAGCTGGTTAAACACATCGCTATGCACGTTGCTGCAAGCAAGCCAGAATTCATCAAACCGGAAG  
AAGACCAACGACGATTTCCGCGACTGCTTCTCGACCAATTTGTGTAGCGATACGTGCAACGACGTTTCGTTTCGGTCTTAAGTAGTTTGGCCTTC

tsf gene

4,190 4,200 4,210 4,220 4,230 4,240 4,250 4,260 4,270

ACGTATCCGCTGAAGTGGTAGAAAAAGAATACCAGGTACAGCTGGATATCGCGATGCAGTCTGGTAAGCCGAAAGAAATCGCAGAGAAAATGG  
TGCATAGCGGACTTCACCATCTTTTTCTTATGGTCCATGTCGACCTATAGCGCTACGTCAGACCATTTCGGCTTCTTTAGCGTCTCTTTTACC

tsf gene

4,280 4,290 4,300 4,310 4,320 4,330 4,340 4,350 4,360 4,370

TTGAAGCCGCATGAAGAAATTCACCGGCGAAGTTTCTCTGACCGGTCAGCCGTTTCGTTATGGAACCAAGCAAACTGTTGGTCAGCTGCTGA  
AATTCCGGCGTACTTCTTTAAGTGGCCGCTTCAAGAGACTGGCCAGTCGGAAGCAATACCTTGGTTTCGTTTTGACAACCAGTCGACGACT

tsf gene

4,380 4,390 4,400 4,410 4,420 4,430 4,440 4,450 4,460

AAGAGCATAACGCTGAAGTGACTGGCTTCATCCGCTTCGAAGTGGGTGAAGGCATCGAGAAAGTTGAGACTGACTTTGCAGCAGAAGTTGCTG  
TTCTCGTATTGCGACTTCACTGACCGAAGTAGGCGAAGCTTCAACCACTTCCGTAGCTCTTTCAACTCTGACTGAAACGTCGTCTTCAACGAC

tsf gene

4,470 4,480 4,490 4,500 4,510 4,520 4,530 4,540 4,550

CGATGTCCAAGCAGTCTTAA

GCTACAGGTTGTCAGAATT

tsf gene

4,560 4,570

### Sequencing of the final strains



After the strains were constructed using the protocol described in Material and Methods 3.4.1, the downstream region of the *rpsB* gene was amplified and sequenced. Alignments of the results confirmed the expected sequences, except for a few single-base mutations. The raw results of the sequencing are reproduced below.

#### Final sequences of the *rpsB-gfp* strain (downstream region of *rpsB*)

AGAAAGAAGCGCTGTATGCGCACTCGTGAGCTGGAGAAA  
 CTGGAACACAGCCTGGGCGGTATCAAAGACATGGGCGGTCTGC  
 CGGACGCTCTGTTTGTAAATCGATGCTGACCACGAACACATTGC  
 TATCAAAGAAGCAAACAACCTGGGTATTCCGGTATTTGCTATC  
 GTTGATACCAACTCTGATCCGGACGGTGTGACTTCGTTATCC  
 CGGGTAACGACGACGCAATCCGTGCTGTGACCCTGTACCTGGG  
 CGCTGTTGCTGCAACCGTACGTGAAGGCCGTTCTCAGGATCTG  
 GCTTCCCAGGCGGAAGAAAGCTTCGTAGAAGCTGAGCAGGAAA  
 GGCGACAGGAGCGTAAAGGAGAAGAAGCTTTTCACTGGAGTTGT  
 TCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAA  
 TTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATA

CGCCGCAGATGCGTTATCTTCGCTCGCTCATCCCGGTCA  
 CTTACTGATGTAAGCTCCCGGGAATTCTCGAGCTTGCCGCCTT  
 TCTGCAACTCGAACTATTTTGGGGGAGTTATCAAGCTTATTTG  
 TATAGTTCATCCATGCCATGTGTAATCCCAGCAGCTGTTACAAA  
 CTCAAGAAGGACCATGTGGTCTCTCTTTTCGTTGGGATCTTTC  
 GAAAGGGCAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAA  
 GGACAGGGCCATCGCCAATTGGAGTATTTTGTGATAATGGTC  
 TGCTAGTTGAACGCTTCCATCTTCAATGTTGTGTCTAATTTTGA  
 AGTTAACTTTGATTCCATTCTTTTGTGTTGTCTGCCATGATGTAT  
 ACATTGTGTGAGTTATAGTTGTATTCCAATTTGTGTCCAAGAA  
 TGTTTCCATCTTCTTTAAAATCAATACCTTTTAACTCGATTCTA  
 TTAACAAGGGTATCACCTTCAAACCTTGACTTCAGCACGTGTCTT  
 GTAGTTCCCGTCATCTTTGAAAAATATAGTTCTTTCCCTGTACAT  
 AAACCTTCGGGCATGGCACTCTTGAAAAAGTCATGCTGTTTCA  
 TATGATCTGGGTATCTCGCAAAGCATTGAAGACCATACGCGAA  
 AAGTAGTGACAAGTGTTGGCCATGGAACAGGTAGTTTTCCAGT  
 AGTGCAAATAAATTTAAGGGTAAAGTTTTCCGTATGTTGCATC  
 ACCTTCACCCTCTCCACTGACAGAAAAATTTGTGCCCATTTAAC  
 ATCACCATCTAATTCAACAAGAATTGGAAACAACCTCCAGTGAA  
 AGT

**Final sequences of the *rpsB-mCherry* strain (downstream region of *rpsB*)**


 TTTCGACAGCTGACCAAGAAGAAGCGCTGATGCGCACTC  
 GTGAGCTGGAGAACTGGAAAACAGCCTGGGCGGTATCAAAGA  
 CATGGGCGGTCTGCCGGACGCTCTGTTTGTAATCGATGCTGAC  
 CACGAACACATTGCTATCAAAGAAGCAAACAACCTGGGTATTC  
 CGGTATTTGCTATCGTTGATACCAACTCTGATCCGGACGGTGT  
 TGACTIONCGTTATCCCGGGTAACGACGACGCAATCCGTGCTGTG  
 ACCCTGTACCTGGGCGCTGTTGCTGCAACCGTACGTGAAGGCC  
 GTTCTCAGGATCTGGCTTCCCAGGCGGAAGAAAGCTTCGTAGA  
 AGCTGAGCAGGAAAGGCGACAGGAGACTAGCAAAAGATCCAAG  
 GCGGAGGAGGATAACATGGCTATCATTAAAGAGTTCATGCGCT  
 TCAAAGTTCACATGGAGGGTTCTGTAAACGGTCACGAGTTCGA  
 GATCGAAGGCGAAGGCGAGGGCCGTCCGTATGAAGGCACCCAG  
 ACCGCCAAACTGAAAGTGACTAAAGGCGGCCCGCTGCCTTTTG  
 CGTGGGACATCCTGAGCCCGCAATTTATGTACGGTTCTAAAGC  
 GTATGTTAAACACCCAGCGGATATCCCGGACTATCTGAAGCTG  
 TCTTTTCCGGAAGGTTTCAAGTGGGAACGCGTAATGAATTTTG  
 AAGATGGTGGTGTCTGACCGTCACTCAGGACTCCTCCCTGCA  
 AGATGGCGAGTTCATCTATAAAGTTAAACTGCGTGGTACTAAT  
 TTTCCATCTGATGGCCCGGTGATGCAGAAAAAGACGATGGGTT  
 GGGAGGCGTCTAGCGAACGCATGTATCCGGAAGATGGTGCGCT  
 GAAAGGCGAAATTAACAGCGCCTGAAACTGAAAGATGGCGG  

 TTCGCGCCGCAGATGCGTTATCTTCGCTCGCTCATCCCGG  
 TCACTTACTGATGTAAGCTCCCGGGAATTCTCGAGCTTGCCGC  
 CTTTCTGCAACTCGAACTATTTTGGGGGAGTTATCAAGCTTAT  
 TTGTACAGCTCATCCATGCCACCGGTAGAATGACGACCCTCCG  
 CGCGCTCATATTGCTCTACGATCGTATAATCTTCATTATGAGAG  
 GTGATGTCCAGTTTAATATTCACATTGTACGCGCCAGGCAGCT  
 GCACAGGTTTCTTGGCTTTGTACGTGGTTTTCACTTCAGCGTC  
 ATAATGGCCGCCATCTTTCAGTTTTCAGGCGCTGTTTAATTTGCG  
 CTTTCAGCGCACCATCTTCCGGATACATGCGTTTCGCTAGACGCC  
 TCCCAACCCATCGTCTTTTTCTGCATCACCGGGCCATCAGATGG  
 AAAATTAGTACCACGCAGTTTAACTTTATAGATGAACTCGCCA  
 TCTTGCAGGGAGGAGTCCGTGAGTGACGGTCACGACACCACCAT  
 CTTCAAATTCATTACGCGTTCCTCACTTCAAACCTTCCGGAAAA  
 GACAGCTTCAGATAGTCCGGGATATCCGCTGGGTGTTTAACAT  
 ACGCTTTAGAACCGTACATAAATTGCGGGCTCAGGATGTCCCA  
 CGCAAAGGCAGCGGGCCGCCTTTAGTCACTTTCAGTTTGGCG  
 GTCTGGGTGCCTTCATACGGACGGCCCTCGCCTTCGCCTTCGA  
 TCTCGAACTCGTGACCGTTAACAGAACCCTCCATGTGAACTTTG  
 AAGCGCATGAACTCTTTAATGATAGCCATGTTATCCTCCTCGCC  
 CTTGGAT

### 3.5.2 S6 Text – Strain validation in batch growing conditions

The *rpsB-gfp* and *rpsB-mCherry* strains were characterized on different media using a Tecan microplate reader. Strains were grown in M9 minimal medium at 37°C in 96-well microplates. The absorbance and fluorescence were measured approximately every minute for up to 24 h. In a single experiment, this allowed to generate up to 96 growth curves like the one schematized in Fig. 1.3.

In Fig. 3.10, we show growth curves obtained on M9 supplemented with 0.2% glucose. As for the construction of Bakshi *et al.* [81], our strains possess a wild-type growth rate (Fig. 3.10A), indicating that the *gfp*-tagging or *mCherry*-tagging of the ribosomal S2 subunit does not impede the functioning of the ribosome. It is known that *E. coli* cultures exhibit a high autofluorescence overlapping the GFP spectrum [197], but the fluorescence of the *rpsB-gfp* strain is at least 4-time higher than the *wild-type* (Fig. 3.10B). Contrary to green autofluorescence, red autofluorescence of *E. coli* cultures is very low (Fig. 3.10B) so the corrected signal level for *mCherry* is even stronger. Overall, both strains possess enough fluorescence to allow quantification at the population level, all the more so at the cell level because most of the autofluorescence is concentrated in the medium outside the cell [197].

In Fig. 3.10C, we display the ratio between the (corrected) fluorescence of the strain and the absorbance ("corrected" meaning that we subtracted the autofluorescence of the wild-type strain). As a first approximation, this ratio can be used as a proxy for the fluorescence concentration in the cells, therefore for the ribosome concentration. The instability at the beginning of the experiment (before 400 min) is characteristic of the difficulty of obtaining robust estimations when the absorbance of the culture is low [178]. The RFU/Abs ratio is stable in the interval [400,600] min for both *rpsB-gfp* and *rpsB-mCherry* strains, indicative of exponential steady-state growth. Note that this short time interval (a couple of generations) of readability before approaching the stationary phase is one of the main problems preventing the use of batch conditions for obtaining robust data about growth transitions. Indeed, most of our attempts to perform robust growth transitions in this region were unsuccessful.

Results for *rpsB-gfp* and *rpsB-mCherry* start to differ when glucose is exhausted and the cells enter stationary phase. As would be expected, the fluorescence of the *rpsB-gfp* strain starts to slowly decrease as GFP proteins are degraded or photobleached. On the contrary, the fluorescence of the *rpsB-mCherry* strain quickly triples when entering stationary phase, at a rate that is much higher than the physical limit imposed by the protein

synthesis rate [33].

In order to explain these strange dynamics, we set-up an experiment to evaluate the degradation and maturation rates of GFP and mCherry in our strains. The procedure was to instantly stop all protein synthesis in mid-exponential phase using an antibiotic, then to observe the evolution of the fluorescence in a condition with no synthesis of new reporter proteins. The estimation was performed using the following model:

$$\frac{dP_m}{dt} = K \cdot P - (\mu + D) \cdot P_m, \quad (3.18)$$

$$\frac{dP}{dt} = f(t) - (\mu + D + K)P, \quad (3.19)$$

where  $P_m$  is the concentration of mature (fluorescent) proteins,  $P$  is the concentration of immature (non-fluorescent) proteins,  $K$  is the maturation rate,  $D$  is the degradation rate (assumed identical for mature and immature proteins),  $f(t)$  the protein synthesis rate, and  $\mu$  the growth rate. By stopping all protein synthesis in the bacteria, we obtain  $f(t) = \mu = 0$ , hence the model becomes:

$$\frac{dP_m}{dt} = K \cdot P - D \cdot P_m, \quad (3.20)$$

$$\frac{dP}{dt} = -(D + K)P. \quad (3.21)$$

This system of differential equations can be analytically solved to yield:

$$P_m(t) = (P_m(0) + P(0) (1 - e^{-Kt})) e^{-Dt}$$

with  $P_m(0)$  and  $P(0)$  the initial concentrations of mature proteins and immature proteins, respectively. By dividing by  $P_m(0)$  and taking the log of both sides, we can rewrite it as

$$\log \frac{P_m(t)}{P_m(0)} = -Dt + \log \left( 1 + \frac{P(0)}{P_m(0)} (1 - e^{-Kt}) \right). \quad (3.22)$$

Since growth is stopped, the bacterial volume is constant and  $\frac{P_m(t)}{P_m(0)}$  can now be easily measured by directly taking the fluorescence normalized by its initial value, while the maturation rate  $K$ , the degradation rate  $D$ , and the initial fraction of immature proteins  $\frac{P(0)}{P_m(0)}$  are free parameters that can be fitted on fluorescence time series. The curves used for the estimation are represented in Fig. 3.11 and the results are reported in Tab. 3.1. As expected, both reporter proteins are stable and exhibit a long half-life in exponential phase (>24 h). The maturation rate of GFP was too fast for obtaining a numerical

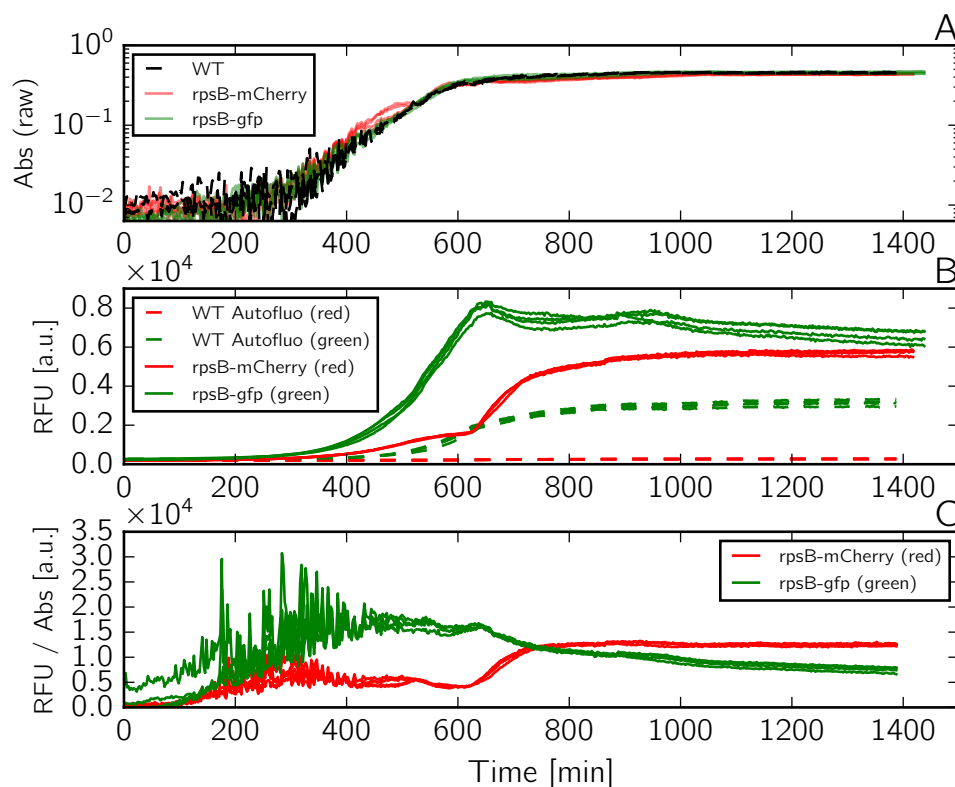


Figure 3.10 – **Growth curves in M9 0.2% glucose for the *rpsB-gfp* and *rpsB-mCherry* strains.** Growth and measurements were performed at 37°C on a 96-well microplate in a Tecan infinite 200 pro. We monitored 4 wells for each of the WT, *rpsB-gfp*, and *rpsB-mCherry* strains. Curves were time shifted to correct for variability in the inoculation process (-50 min for WT, and -20 min for *rpsB-mCherry*). (A) Absorbance at 600 nm, corrected for background by subtracting the absorbance of the M9 medium. Curve superposition indicates similar growth rates between the WT and the modified strains. (B) Fluorescence measured in each well in Relative Fluorescence Units (RFU). Green fluorescence (485 nm excitation, 535 nm emission) is measured for the WT (dashed lines) and the *rpsB-gfp* (solid lines) strains. Red fluorescence (560 nm exc., 635 nm em.) is measured for the WT (dashed lines) and the *rpsB-mCherry* (solid lines) strains. Fluorescence levels of the modified strains are far above the autofluorescences measured on the WT strain. (C) Ratio of fluorescence over absorbance (proxy for the fluorescence concentration in the cells). Autofluorescence background was corrected by removing the fluorescence measured on the WT. We observed a strange increase in fluorescence concentration for the *rpsB-mCherry* strain after entry into stationary phase, something that does not occur on the *rpsB-gfp* strain. At each measurement cycle, the following procedure was applied: shaking (Orbital 6mm) for 30 s, shaking (Linear 6mm) for 30 s, waiting for 5 s.

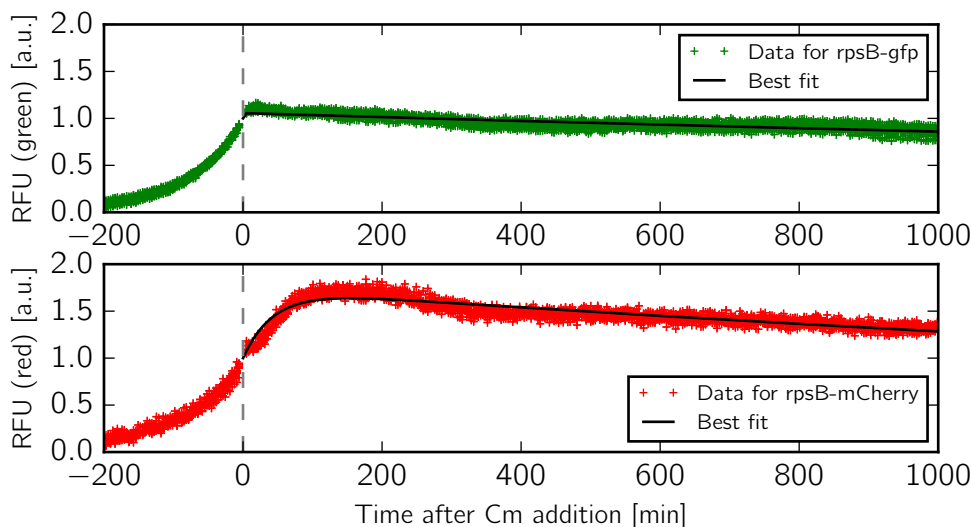


Figure 3.11 – **Estimation of the maturation and degradation rates of the reporter proteins in the *rpsB-gfp* and *rpsB-mCherry* strains.** The strains were grown in the same condition as in Fig. 3.10, except that a high concentration of Chloramphenicol was added in mid-exponential phase ( $t = 0$ , vertical dashed grey line). The fluorescence was normalized by the value at  $t = 0$  and the degradation and maturation rates were estimated by using the model in Eq. 3.22. Parameters of the best fit (solid black line) are reported in Tab. 3.1. Data are the result of the aggregation of 9 and 8 independent growth curves for *rpsB-gfp* and *rpsB-mCherry*, respectively.

value. However, the maturation rate of mCherry is rather slow, with a half-maturation time on the order of 30 min (time needed to mature 50% of a pool of mCherry). Even though degradation and maturation rate can be slightly different in stationary phase, the estimated value are not sufficient to explain the rapid increase in fluorescence observed in Fig. 3.10.

The fluorescence of GFP and mCherry is known to be affected by intracellular physiological changes like pH or  $pO_2$  [198]. In particular, mCherry has been reported to be extremely sensible to the presence of oxygen for its maturation [198]. However, we reproduced the following results in conditions where the oxygen is not limiting. In particular, the abrupt increase of mCherry fluorescence is conserved when the stationary phase is attained at low bacterial density, in a larger volume (20-mL flask), and on several type of carbon sources (acetate, xylose, glycerol, maltose). We have so far no explanation for these strange dynamics, and we therefore concentrated our

	<b>GFP mut2</b>	<b>mCherry</b>
$P(0)/P_m(0)$	$0.05635 \pm 0.001202$	$0.7349 \pm 0.0027362$
$D$ [ <b>min<sup>-1</sup></b> ]	$0.0002064 \pm 2.1 \cdot 10^{-6}$	$0.0003027 \pm 2.7 \cdot 10^{-6}$
<b>half-life [min]</b>	3325 - 3392	2269 - 2309
$K$ [ <b>min<sup>-1</sup></b> ]	$\emptyset$	$0.02254 \pm 0.0003$
<b>half-maturation [min]</b>	$\emptyset$	30.27 - 31.16

Table 3.1 – Fitted parameters for the degradation and maturation of GFP mut2 and mCherry, according to the model in Eq. 3.22 and data in Fig. 3.11

efforts on the *rpsB-gfp* strain in the microfluidic experiments.



### 3.5.3 S7 Text – Noise estimation in the microscopy experiment

We estimated the noise in the RFU/pixel/cell measurements by using the data points just before the upshift (Fig. 3.12). Since the cells have been growing for 20 h on acetate in the microfluidic device, bacteria are assumed to be in balanced growth in this region of  $\sim 2$  hours (roughly 23-24 points depending on the well of interest). For this reason, ribosome concentration, and therefore the fluorescence concentration, are expected to be constant. Fig. 3.12 shows the distribution of points in this region, which can be approximated by a Gaussian distribution, for a particular cell.

Interestingly, as can be seen in Fig. 3.13, the mean and standard deviation of the distribution vary from cell to cell in a slightly correlated manner (Pearson  $R^2$ : 0.3213, p-value:  $4.923 \cdot 10^{-5}$ ). We did not investigate if this heterogeneity between cells is the result of true biological variations or a bias due to the microfluidic device. By looking at the points of Fig. 3.13, there are however no strong correlations of the mean and standard deviation with the acquisition field (XY), or the position of the well on the image (W).

Noise characteristics were computed by normalizing and aggregating data for the 45 normal cells presented in Fig. 3.13. Two scenarios were considered. First, the noise is additive, which means that we have the measurement model presented in Section 3.2.4:

$$F(t_k) = \gamma r(t_k) + \eta_k, \quad (3.23)$$

where  $F(t_k)$  is the fluorescence measured at the time step  $t_k$ ,  $r(t_k)$  is the true ribosomal concentration at this time step,  $\gamma$  is an unknown factor, and  $\eta_k$  is the measurement noise. In this scenario, for a time independent ribosome concentration  $r$  (*i.e.* at steady state), the ribosome concentration is equal to its mean on the interval of interest (noted  $m_k(\cdot)$ ). Given that

$$m_k(F) = m_k(\gamma r) + m_k(\eta_k) = \gamma m_k(r), \quad (3.24)$$

the noise residues are given by

$$\eta_k = F(t_k) - m_k(F). \quad (3.25)$$

In other words, we can normalize the data points of each cell by removing their mean in order to aggregate the noise residues and thus obtain an estimation of the noise level (Fig. 3.14, top panel).

Alternatively, the noise could be multiplicative. The noise model is then different of the one presented in Section 3.2.4, and can be re-written:

$$F(t_k) = \gamma r(t_k) \cdot (1 + \lambda_k), \quad (3.26)$$

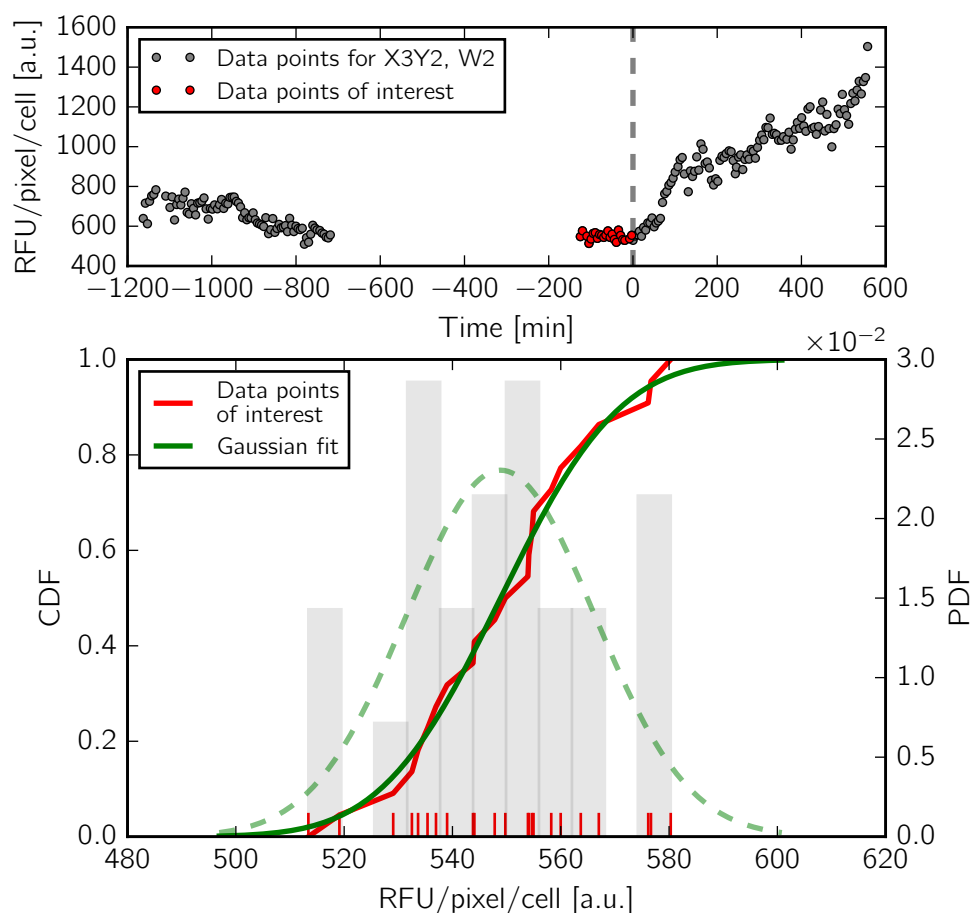


Figure 3.12 – **Points of interest for the noise estimation.**

(*Top graph*) RFU/pixel/cell data points for the cell at the bottom of the well labeled X3Y2,W2. Noise estimation was performed on the points just before the upshift (red points), where the bacteria are assumed to grow at steady state. (*Bottom graph*) Cumulative density function (CDF, in red) and probability density function (PDF, in gray) of the points highlighted in red on the top graph. They are visually compared with CDF (solid green line) and PDF (dashed green line) of a Gaussian fit with mean 548.88 and standard deviation 17.316.

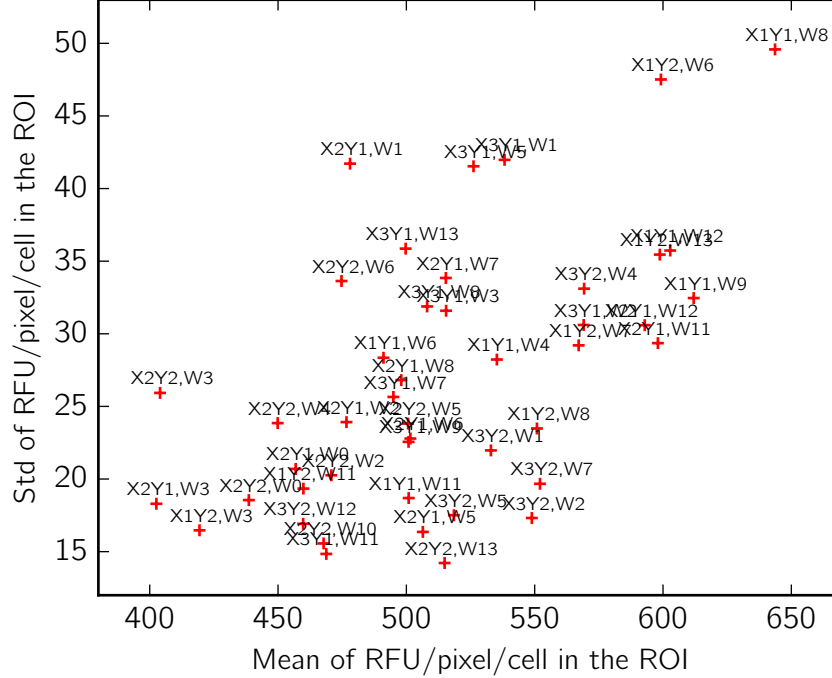


Figure 3.13 – **Distribution of the means and standard deviations in the region of interest for the normal cells.** The empirical mean and standard deviation were evaluated in the region before the upshift (red points on Fig. 3.12) for each of the 45 normal cells. They appear to be slightly correlated (Pearson  $R^2$ : 0.3213, p-value:  $4.923 \cdot 10^{-5}$ ) which could indicate a multiplicative instead of an additive noise.

were  $\gamma r(t_k) \cdot \lambda_k$  is the measurement noise, proportional to the value measured. In that scenario, with the same notations as above, the noise  $\lambda_k$  is given by

$$\lambda_k = \frac{F(t_k)}{m_k(F)} - 1, \quad (3.27)$$

which means we can normalize data points from different cells by dividing by the mean and removing 1 (Fig. 3.14, bottom panel).

Both models were considered and are represented in Fig. 3.14. Despite the correlation identified in Fig. 3.13, the means of the residues in both models are very close to zero, and the distributions are well approximated by a Gaussian distribution. For this reason, we decided in the main analysis to stick with an additive noise, the main reason being that the implementation used for the Kalman smoothing procedure are faster and more stable with

an additive noise model. We assumed an additive white Gaussian noise with mean 0 and standard deviation of 28.29 for the simulation of the synthetic data and the signal reconstruction through the Kalman smoothing procedure. As discussed in Section 3.3, the only way to definitively choose a noise model over the other would have been to obtain two different steady-state growths for each cell. Unfortunately, the presented time series were not long enough to obtain a new steady state on glucose, something that will be corrected in future experiments.

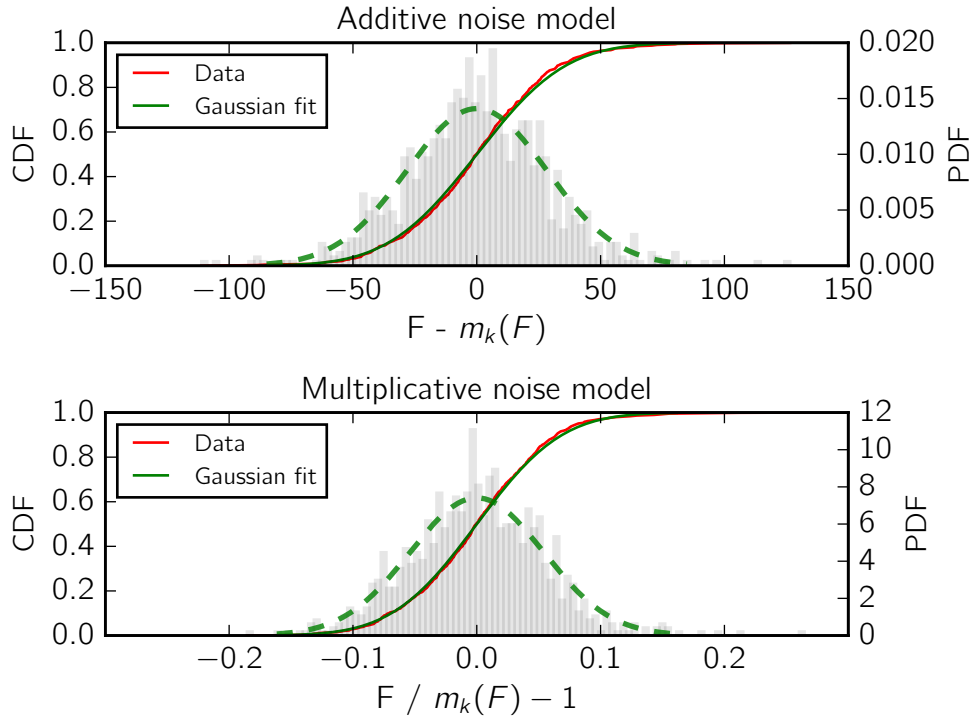


Figure 3.14 – **Distribution of the noise residues after normalization and aggregation for the 45 normal cells.** Two noise models were considered for the normalization: either the temporal mean  $m_k(\cdot)$  was removed from the measurements independently for each cell (additive noise model in Eq. 3.25), or 1 was removed from the ratio of the measurements with the temporal mean  $m_k(\cdot)$  (multiplicative noise model in Eq. 3.27). For the additive noise model, the fitted Gaussian distribution has a mean of  $1.256 \cdot 10^{-15}$  and a standard deviation of 28.29. For the multiplicative noise model, the fitted Gaussian distribution has a mean of  $2.666 \cdot 10^{-18}$  and a standard deviation of 0.05383.

### 3.5.4 S8 Text – Examples of complete analysis for 5 cells

In order to obtain a more visual idea of the results of the analysis, we display below the complete signal reconstructions for 5 cells. In Figs 3.15-3.19, the image on the left is the last image analyzed for the corresponding well. The cell either at the top or the bottom (depending on the orientation in the device) was manually segmented by selecting two pixels at the poles on the fluorescence images (red cross). A 6-pixel-wide rectangular mask was computed for each image, resulting in the masked image on the right. In each of the graphs, the black and green points represent data points, while the red solid lines are the results of the signal reconstruction *via* the Kalman smoothing procedure. The parameters of the Kalman smoothing are described in Material and Methods 3.4.5 and Figs 3.3 and 3.6. Specific comments on each cell are given in the caption of Figs 3.15-3.19.

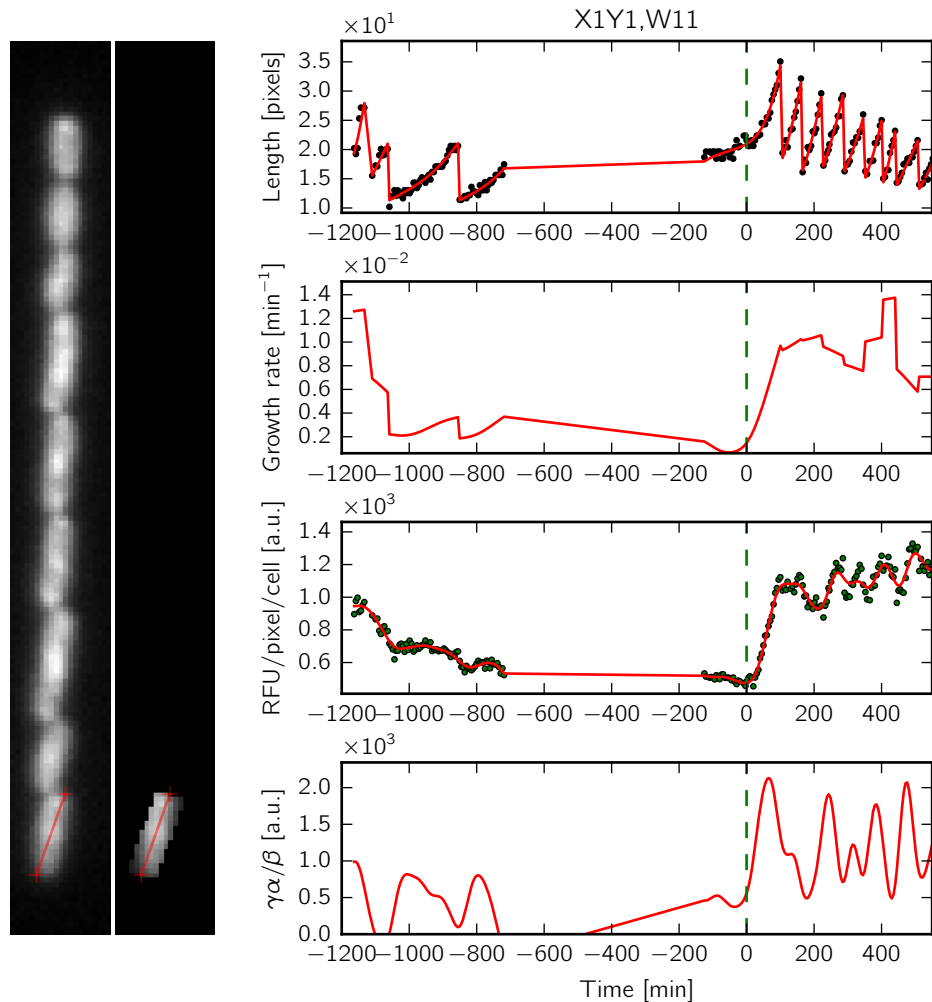


Figure 3.15 – Complete analysis for the cell located at the bottom of the well labelled "X1Y1, W11". (Figure description available in the introduction of [S8 Text](#).)

Growth rate and resource allocation are particularly unstable at the beginning of the experiment, but seem to stabilize before the upshift. Oscillations in the RFU/pixel/cell signal are clearly visible and result in oscillations in the resource allocation signal reconstruction.

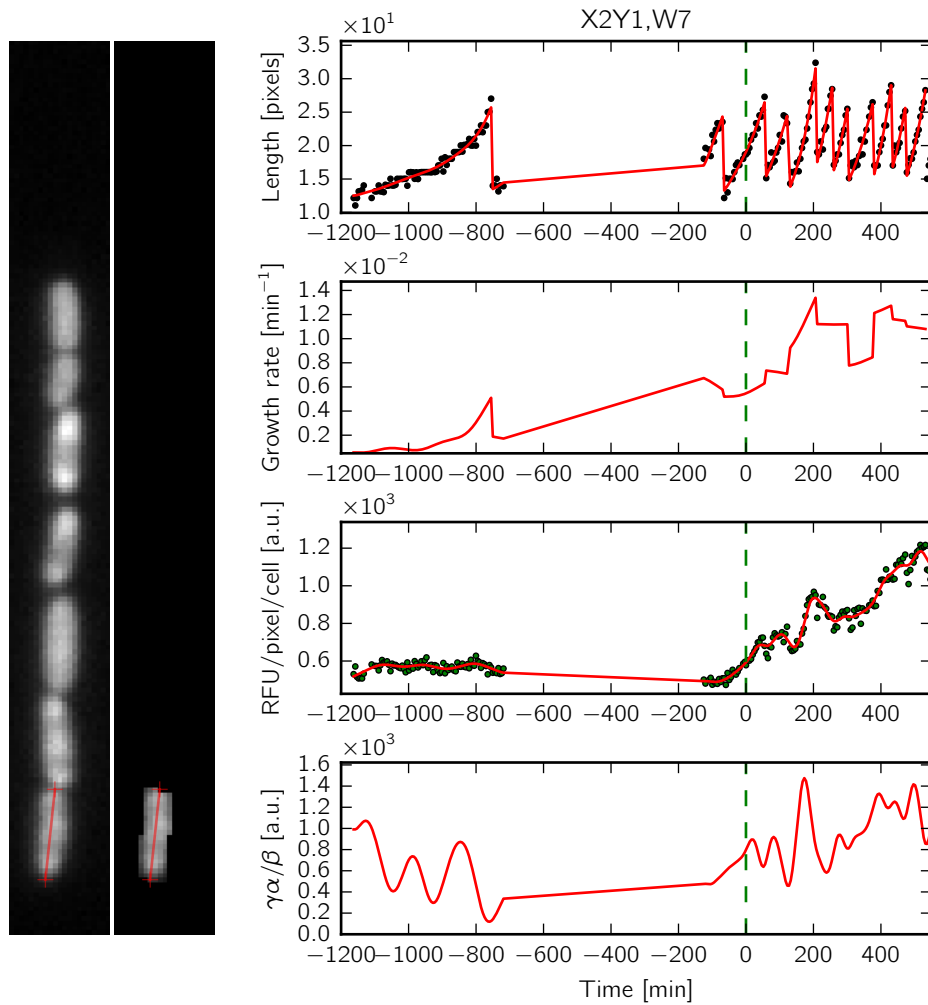


Figure 3.16 – Complete analysis for the cell located at the bottom of the well labelled "X2Y1, W7". (Figure description available in the introduction of [S8 Text](#).)

Here, the resource allocation is unstable at the beginning of the experiment despite the apparent regularity of the RFU/pixel/cell. This could indicate a smoothing factor that is too low for this particular cell.



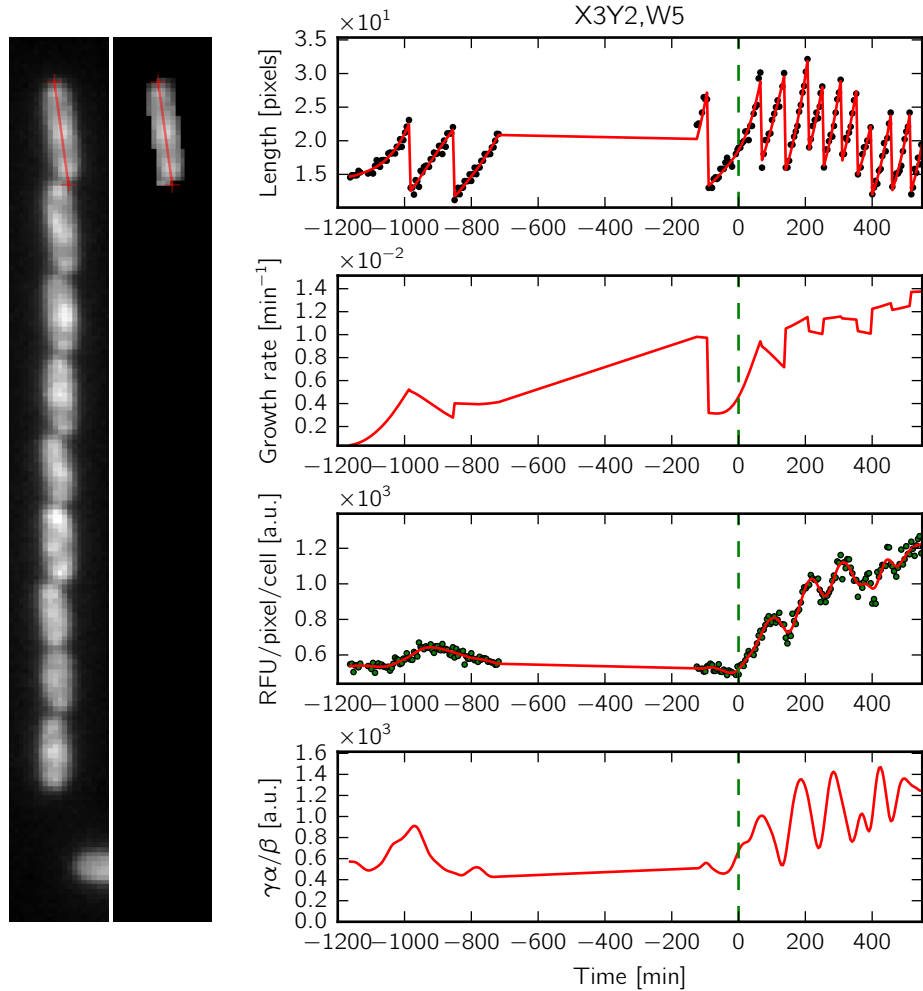


Figure 3.17 – Complete analysis for the cell located at the bottom of the well labelled "X3Y2, W5". (Figure description available in the introduction of [S8 Text](#).)

We see for this cell that the reconstruction of the growth rate on the acetate medium (before 0) is affected by the huge gap in the acquisition, the increase before -150 min clearly being an artifact. However, because the Kalman smoothing procedure we used allows for flexibility between the mother and the daughter cells, we quickly recover a more realistic growth rate before the upshift. Despite an instability at the beginning of the experiment, the resource allocation reconstruction exhibits a remarkable stability during growth on the acetate medium, followed by oscillations after the upshift on glucose (also visible directly on the RFU/pixel/cell data).

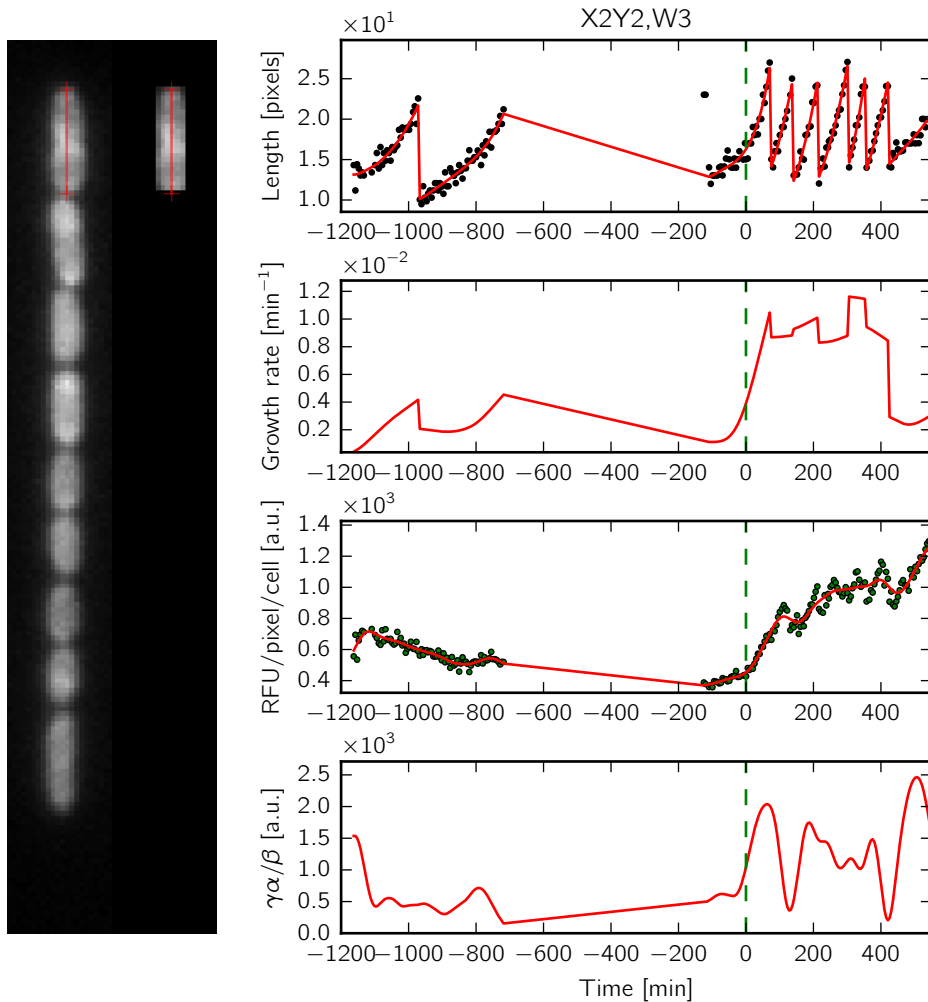


Figure 3.18 – **Complete analysis for the cell located at the bottom of the well labelled "X2Y2, W3".** (Figure description available in the introduction of [S8 Text.](#))

The reconstruction on this cell is a good example of what was expected: the growth rate is stable on acetate, than quickly increases after the upshift on glucose. The resource allocation is stable on acetate (indicating a steady state) than starts to oscillate after the upshift on glucose, even though the smoothing factor seems a little too high for this particular cell (some oscillations are poorly predicted after the upshift, in particular in the interval [200,400] min).

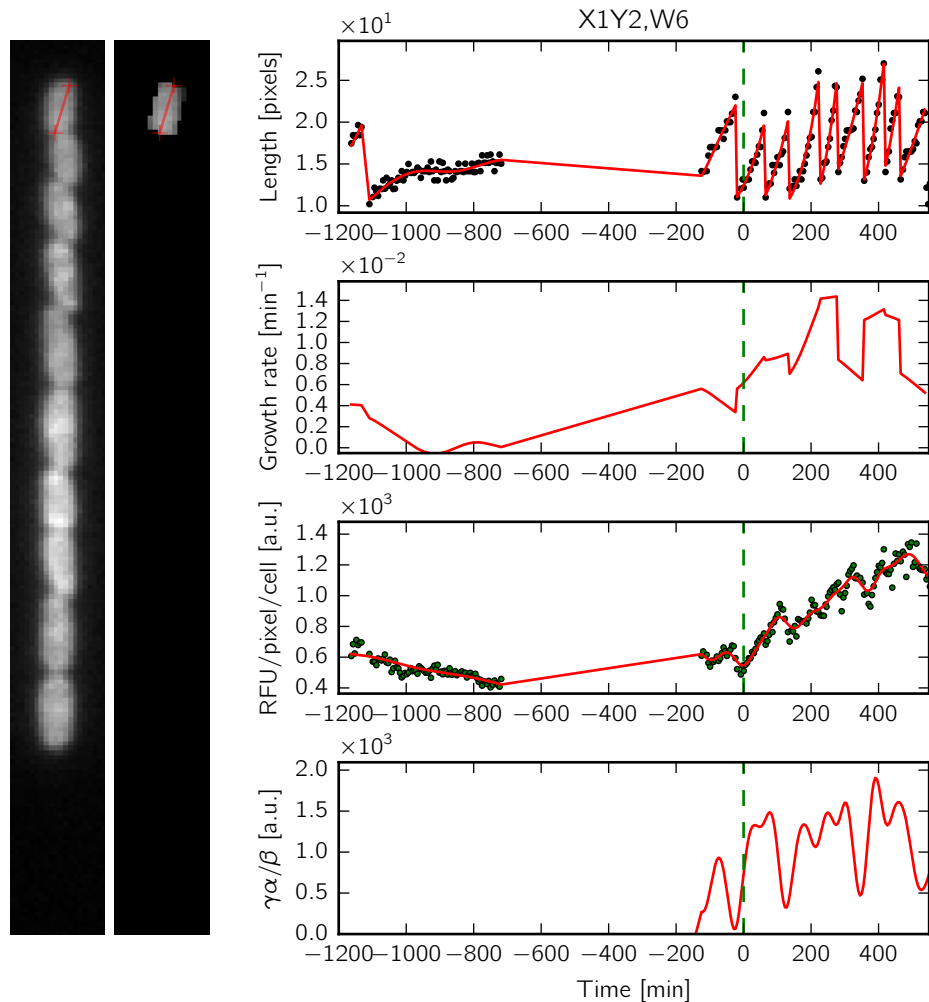


Figure 3.19 – Complete analysis for the cell located at the bottom of the well labelled "X1Y2, W6". (Figure description available in the introduction of [S8 Text](#).)

The results on this cell are particularly questionable: we do not seem to obtain a good steady state on acetate since the reconstructed resource allocation appears either negative or oscillating before the upshift. It is thus impossible to conclude if the oscillations observed after the upshift are real or due to overfitting. This cell highlights why robust steady states before and after the transition are important for a reliable signal reconstruction.

### 3.5.5 S2 Figure – Cell categories identified in the analysis

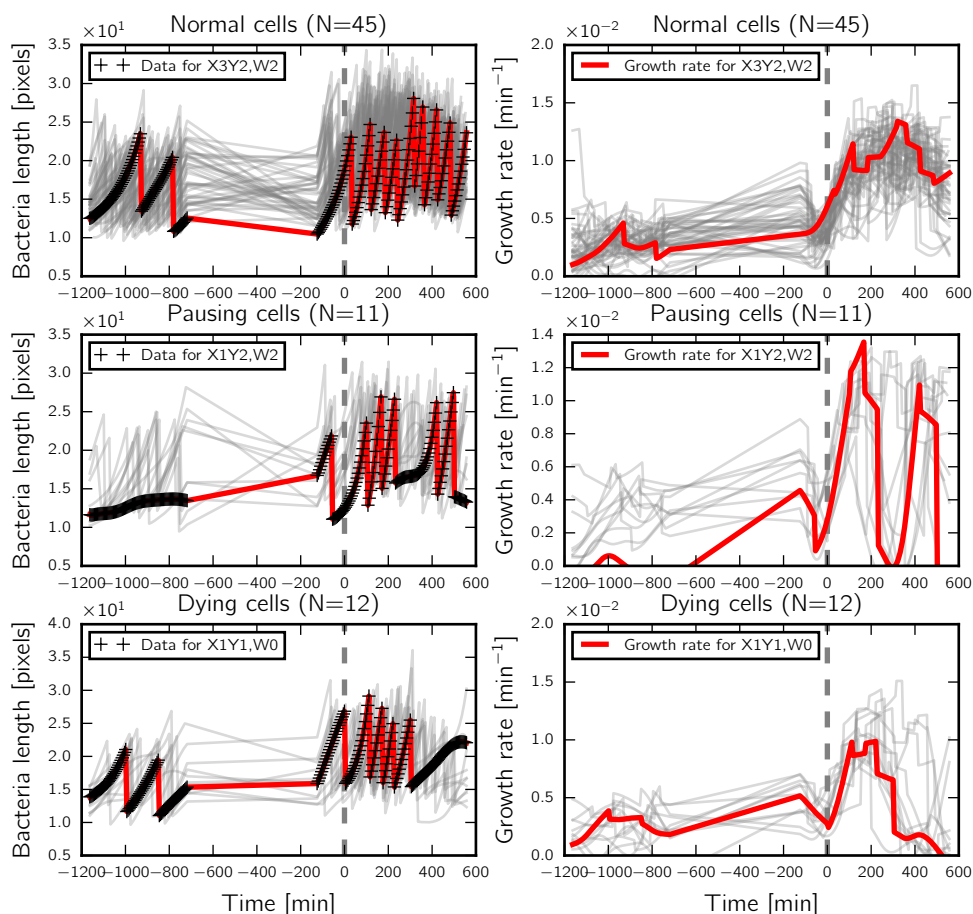


Figure 3.20 – Cell categories identified in the microscopy analysis. As stated in Section 3.2.4, the reconstruction of the growth rate for the 68 available cells motivated their classification in three categories: dying cells stop growing (lowest graphs), pausing cells present growth rate oscillations (middle graph), and normal cells that do not present any of the above (highest graph).

## 3.5.6 S3 Figure – Robust statistics for the cell categories

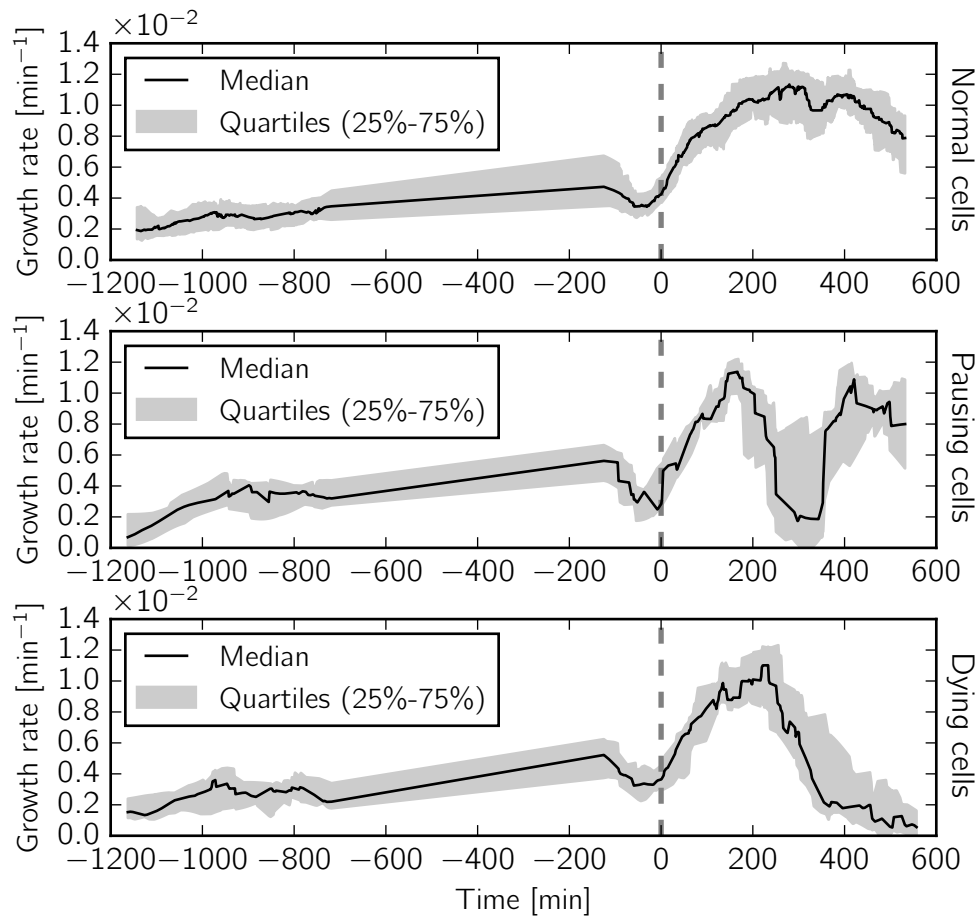


Figure 3.21 – **Robust statistics for the cell categories identified in the microscopy analysis.** Each graph shows the 25% (lower gray curve), 50% (solid black curve) and 75% (upper gray curve) quartiles, computed at each time step for the growth rates presented in S2 Fig. The gray area represents the interquartile range.

# Chapter 4

## Discussion

*"Apples fall onto the Earth because natural selection eliminated apples falling towards the sky."* – @tomroud [199], original source unknown.

---

### Résumé du Chapitre 4: Discussion

Contrairement à sa cousine la physique, la biologie manque encore de théories quantitatives à fort pouvoir prédictif. Même si de grosses avancées ont eu lieu en biologie évolutive, d'autres domaines sont loin d'être aussi développés d'un point de vue mathématique. C'est notamment le cas de la croissance, qui malgré la place fondamentale qu'elle occupe dans ce qui définit un être-vivant, repose sur des mécanismes qui semblent varier énormément d'un organisme à l'autre, et pour lesquels des lois fondamentales sont dures à identifier.

Chez les microorganismes, nous disposons néanmoins de lois de croissance. Elles montrent de manière empirique que malgré la grande diversité des mécanismes moléculaires qui assurent le contrôle de la croissance, la composition des microorganismes obéit à des règles universelles lorsqu'ils se multiplient à l'état stationnaire dans différents environnements. En particulier, l'abondance de leurs ribosomes s'ajuste linéairement avec la richesse du milieu, d'une manière qui maximise leur taux de croissance. Mais comme décrit au cours du Chapitre 1, ces lois ont été essentiellement établies en croissance stationnaire, un état très rarement rencontré par ces organismes dans leur milieu naturel. Pour quelles raisons les microorganismes seraient-ils optimisés pour un état qu'ils n'ont que très peu rencontré au cours de leur évolution ?

Les lois de croissance à l'état stationnaire ne seraient-elles pas une application particulière de lois plus générales qui s'appliquent en environnements variables ?

Notre but dans ce manuscrit a été d'établir un cadre de travail à la fois théorique et expérimental dans lequel une perspective dynamique peut être adoptée sur les lois de croissance. Au cours du Chapitre 2, nous avons utilisé un modèle d'auto-réplicateur pour évaluer la redistribution optimale des ressources lors d'un changement environnemental. En élargissant les principes d'optimalité établis dans les lois de croissance stationnaire, nous avons montré qu'une distribution des ressources de type tout-ou-rien (*bang-bang* en anglais) optimise la biomasse générée tant que la composition cellulaire n'est pas à l'équilibre. Cela nous a servi de point de référence pour comparer entre elles différentes stratégies de régulation. Ainsi, nous avons montré que les stratégies mesurant l'état interne de la cellule s'avèrent plus efficaces que celles tirant leur information de l'environnement, contrairement à l'état stationnaire de croissance où toutes ces stratégies sont strictement équivalentes. De plus, une stratégie légèrement plus complexe mais proche du système ppGpp chez *Escherichia coli*, est théoriquement capable de s'approcher de la distribution tout-ou-rien identifiée comme optimale.

La simplicité du modèle, même si elle peut rendre certains lecteurs sceptiques, s'est avérée cruciale dans notre étude. En effet, cela nous a permis de clairement spécifier les hypothèses mises en jeu. La plus critique s'avère être celle qui considère la production de biomasse comme étant le facteur de *fitness* principal que l'on soit en environnement stable ou dynamique. Bien que largement répandue, cette hypothèse n'est finalement basée que sur des arguments indirects, ou bien établis dans des conditions de laboratoire, et sur des souches de laboratoire (Chapitre 1). De fait, en conditions naturelles et à l'échelle des temps évolutifs, il n'est pas certain que ce critère soit vraiment universel. Une étape importante a donc été de vérifier expérimentalement les prédictions que nous avons pu tirer de cette hypothèse.

C'est ce qui a été abordé au cours du Chapitre 3 de ce manuscrit. Nous avons mis en place un cadre d'étude expérimental dans lequel la distribution des ressources peut être observée lors d'une transition de croissance. Tout d'abord, une souche d'*Escherichia coli* a été modifiée en attachant une protéine fluorescente (GFP) à l'une des sous-unités du ribosome. Cette souche a ensuite été observée en utilisant une technique avancée de microfluidique permettant de réaliser des transitions robustes et contrôlées du milieu de culture. Enfin, les mesures obtenues ont pu être exploitées *via* le développement d'une technique de reconstruction du signal utilisant le lissage de Kalman. Plusieurs améliorations du dispositif seront cependant nécessaires pour conclure définitivement sur l'existence d'une production tout-ou-rien des ribosomes après

une transition. Deux scénarios sont cependant envisageables.

Dans le premier cas, la nature oscillatoire de la synthèse des ribosomes pourrait être confirmée. Cela suggérerait que les mécanismes de régulation de la distribution des ressources maximisent la biomasse produite lors d'une transition, et donc que les microorganismes sont effectivement adaptés à des environnements changeants. Pour confirmer ces résultats, de nouvelles données devront nécessairement être acquises, notamment en mesurant des transitions similaires dans une variété plus large de conditions environnementales (changement de source de carbone, ajout d'acides aminés dans le milieu pour contourner le métabolisme, utilisation d'une restriction en azote au lieu du carbone, ...). Le dispositif expérimental pourrait également être modifié pour contrôler un plus grand nombre de paramètres. Par exemple, en attachant un autre rapporteur fluorescent à une enzyme clé du métabolisme, nous pourrions vérifier si les synthèses de la machinerie d'expression génique et de la machinerie métabolique sont effectivement en anti-phase lors des oscillations.

Dans le second cas, le comportement observé serait différent d'une production tout-ou-rien. Par construction, cela signifierait qu'une hypothèse du modèle n'est pas respectée. Par exemple, cela pourrait signifier que l'optimisation de la biomasse produite lors d'une transition n'est pas un objectif pour la cellule. Une autre possibilité serait que les coûts pour la cellule d'une telle régulation dépassent les éventuels bénéfices qu'elle peut en tirer. En effet, en ne comparant les schémas de régulation que par leurs bénéfices, nous avons fait l'hypothèse implicite que leurs coûts pour la cellule sont négligeables, sinon au moins comparables. Il est possible que même si la production de biomasse s'avère être un critère déterminant, les limites physiques imposées par le coût des régulations empêchent la cellule de faire mieux que le comportement observé. La prise en compte de ces coûts dans le modèle devra donc être explorée avant toute conclusion.

Dans tous les cas, cette exploration des coûts des schémas de régulation pourrait s'avérer cruciale dans l'établissement d'un lien fondamental entre la complexité des schémas de régulation d'une espèce microbienne, et la dynamique de son environnement. En effet, comme montré dans le Chapitre 2, la cellule ne tire avantage d'un schéma de régulation complexe que lorsque son environnement varie. On peut élargir ce résultat en supposant que le bénéfice apporté par des régulations toujours plus complexes finit par saturer. En revanche, le coût de ces régulations devrait continuer à croître, dans la mesure où des régulations plus élaborées impliquent pour la cellule la synthèse de davantage de systèmes de mesure, et donc détournent une part toujours plus grande des ressources cellulaires. On peut donc conjecturer que ces deux tendances vont finir par se croiser au niveau d'un goulot d'étranglement évolutif, au delà duquel les coûts excèdent les bénéfices. On peut de plus s'attendre



à ce que ce goulot d'étranglement survienne plus ou moins loin en fonction de l'intensité et la fréquence des variations environnementales auxquelles les microorganismes sont soumis. En d'autres termes, là où les lois de croissance actuelles décrivent uniquement des variations physiologiques avec la qualité du milieu, nous pourrions ajouter une autre dimension à ces lois qui représenterait la variabilité temporelle de la disponibilité des nutriments dans l'environnement naturel.

---

## Beginning of Chapter 4

Physics has yet to be reduced to a formula that will fit on a piece of clothing [200]. This is even more true for Biology, but the extensive work on the mechanisms of evolution are helping to close this gap (see for instance the Price equation, which summarize in a short and elegant formula the mechanisms of evolution and natural selection [201]). But evolution by itself is not sufficient to define life. Even the least restrictive definition we have – the one we use to look for life in the universe – defines a living system as "a self-sustaining chemical system capable of Darwinian evolution" [202, 203]. Self-sustainment appears to be as fundamental as evolution, and relies on the transformation of matter and energy from the environment into organic matter, in other words, on growth. But the mathematical formulation of questions regarding the self-sustaining character of living systems has not advanced as much as the mathematical analysis of evolution.

Nevertheless, fundamental growth laws have been established for microorganisms. These growth laws show that, regardless of the molecular mechanisms underlying growth control, microorganisms tend to follow the same empirical regularities when grown at steady state in different environmental conditions [24, 31–33]. In particular, they adjust their internal molecular composition after a change in the environment in such a way as to maximize their growth rate (see in particular [24, 32]). Growth laws are a strong support for the theory of a modular organization of microorganisms [32, 204–206]. They represent a big step forward towards gaining a general understanding of the physiology of microorganisms. They have been established at steady state, however, a state in which most microorganisms spend very little time [4–10]. Why would microorganisms be optimal for a situation they rarely encounter? Would known growth laws be specific cases of more general laws allowing microorganisms to respond to changes in the environment?

Our aim in this manuscript has been to establish a theoretical and experimental framework extending growth laws to a dynamical context. We

focused on the growth law describing how ribosome abundance adapts to a change in environment so as to maximize steady-state growth in different media. Applying the same criterion of growth rate maximization, what is the optimal way to allocate resources during a growth transition between two different environments? In Chapter 2, we used a simple self-replicator model of resource allocation, and showed that the steady state is invariant over a number of regulatory schemes: the growth rate can be maximized by measuring either the environment or the internal state of the cell. In both cases, the expression of genes encoding the metabolic and gene expression machineries has to be set to a specific value. By contrast, biomass maximization during a growth transition requires information about the internal state rather than the environment. Moreover, expression of the metabolic and gene expression machineries is not set to one specific value, but varies with time in an on-off manner. What is optimal at steady state and during a growth transition is thus not the same, but this does not mean that a single regulatory system cannot meet both demands.

The model developed here is an instance of proof-of-concept model [49]. It does not necessarily aim at quantitatively predicting or controlling the behavior of microorganisms. By using a simple and abstract representation, it provides a convenient and tractable way of evaluating the implications of dynamical optimality, in particular its divergence from steady-state optimality. Is a dynamical perspective on growth laws necessary? Do regulatory mechanisms need to differ in a dynamical context? Those are the questions that were investigated by means of this model. To our surprise, it additionally provided a new way of looking at the regulatory systems controlling ribosomal abundance in many bacteria.

Some readers may be skeptical about the interest of using such a simple model for addressing the above questions. Given the profusion of knowledge and data available on biological systems, one might be tempted to include in the model every known detail about the molecular implementation of the system of interest. One of the main drawbacks of such mechanistic models is that we quickly lose track of the big picture and the underlying assumptions. When including many molecular details, the models also tend to become untractable, prohibiting the use of mathematical tools like optimal control theory that are quickly overwhelmed by the sheer number of variables and parameters in the model. Besides, simple, abstract models make it easy for the modeler to state the assumptions that are made, and more importantly to identify which ones are logistical, exploratory, or critical (see [49], in particular Box 1). Logistical assumptions do not affect the conclusions of the model, and are only necessary for tractability. A good example in our case is the Michaelis-Menten kinetics that were considered in Chapter 2.

Exploratory assumptions, however, might be important to vary and test. For instance, an exploratory assumption of our model was to consider that the proposed strategies had to be optimal at steady state as well. This was used as a necessary step to reduce the space of possible solutions, and make the comparison between different strategies time-independent. Relaxing this assumption would require the exploration of other environmental changes beyond a simple nutrient upshift, but would probably raise interesting considerations [207–210].

But the assumptions that are most in need of discussion are the critical assumptions, *i.e.* the ones that are invalidated if the model predictions are not met in nature. In our case, a critical assumption is that biomass production is the main fitness factor in stable or changing environments. Throughout the manuscript, we did not really challenge this hypothesis, but only provided arguments about why this assumption is reasonable in some experimental conditions. The arguments given, however, are mostly based on results obtained with laboratory strains in laboratory conditions. While it is clear that microorganisms can be cultivated and are even naturally found in conditions where maximizing the growth rate ensures their persistence [21–24], we have no direct proof that biomass production is the fitness factor that allowed them to naturally persist on evolutionary time-scales. In addition, one must be careful when explaining the behavior of living systems solely as a consequence of natural selection. Such systems are embedded in the physical world, and sometimes universal laws just result from the physical limitations that apply to the system (*e.g.*, Monod’s law described in Chapter 1).

The second part of this manuscript was dedicated to setting up an experimental framework for studying resource allocation during growth transitions. Much emphasis has been given to the experimental issues encountered when studying growth transitions and to come up with possible solutions. Fluorescent reporter proteins were used to cope with the need for *in-vivo* measurements at high temporal resolution. Advances in microfluidics were exploited to perform steady-state-to-steady-state transitions following a change in medium, in order to buffer the effect of the pre-culture history [39–41]. Finally, the signal of interest was reconstructed through the use of Kalman smoothing, a powerful signal processing algorithm that has, as we showed, many features that make it suitable for the analysis of microscopy time-series data [118, 119, 185]. Overall, the set-up still needs many improvements, though most of them are experimental issues that should be easily fixable.

Chapter 3 must not be taken as a "test of the model" presented in Chapter 2, but rather as a test of its assumptions (see Box 2 in [49]). The model is correct in that bang-bang resource allocation follows mathematically from the assumption that microorganisms maximize their biomass production at

steady state and during growth transitions. The model predictions can be used, however, as a tool to test whether this critical assumption is valid. In Chapter 3, we thus apply the experimental set-up to observe the resource allocation profile following a nutrient upshift in *E. coli*. While we were not able to arrive at an unambiguous conclusion about the correctness of the model prediction in this manuscript, we can draw up two possible scenarios in the wake of future improvements of the experimental set-up.

In one scenario, the bang-bang or at least oscillatory nature of resource allocation during a growth transition is confirmed. This would suggest that the regulatory mechanisms of resource allocation do maximize biomass production during growth transitions, hence that microorganisms are adapted for changing environments. In order to further establish these results, we would need additional data for a broader range of environmental conditions. The most straightforward extension would be to test other upshift and downshift schemes (different carbon sources, amino-acids supplementation, nitrogen instead of carbon limitation, ...). The experimental set-up could also be slightly modified to test a broader range of variables. For instance, the strain could be modified so as to express an additional fluorescent protein that would report the abundance of a key enzyme of metabolism. Along with the ribosomal reporter, this would allow us to test the prediction that the metabolic and gene expression machineries are expressed in anti-phase.

In the other scenario, we do not observe any trace of a bang-bang resource allocation scheme. In that case, the crucial question is: "Which critical assumptions are not valid?" We cannot discard the possibility that an assumption that was originally identified as exploratory or logistical might in fact be critical for the predictions of the model. Although unlikely *a priori*, this might be the case for the expressions used for the macroreactions in the model (Eqs 2.6-2.7), the definition of the volume as invariably proportional to the total mass of macromolecules (Eq. 2.2), or the fact that degradation is assumed negligible with respect to the rates of other reactions in the system. More likely, the assumptions regarding the cost of the regulatory system could be more critical than initially thought. By only comparing the regulatory schemes with respect to their benefits, we implicitly made the assumption that their costs are negligible or at least comparable. The costs of regulatory schemes are difficult to model [211–213] since the cellular variables are not equally costly to measure and, as we showed for the ppGpp system, one mechanism may measure several variables at once. Nevertheless, relaxation of this hypothesis would definitively need to be explored before concluding that microorganisms do not optimize their biomass production during growth transitions.

In any case, exploring the costs of regulatory mechanisms might uncover

a fundamental link between the complexity of the regulatory schemes and the dynamics of the environment. As we illustrated in Chapter 2, complex regulatory schemes only prove beneficial in a dynamical context, away from steady state. At the opposite extreme, one can expect the benefits of these regulatory schemes to saturate, whereas the cost of complexity would probably not saturate, since acquiring more information is expected to require additional sensing systems, diverting away an increasingly larger part of cellular resources. Overall, these two tendencies are expected to cross at a point that would represent an evolutionary bottleneck, beyond which the costs exceed the benefits (see [214] for another example of such a bottleneck). Interestingly, the position of the bottleneck might be dependent on the environment, resulting in a law that would link the complexity of regulatory systems to the dynamics of the environment. In other words, while the current growth laws consider how the physiology of the cell varies with the nutrient quality of the medium, we could add another dimension representing the time-varying availability of this nutrient in the natural environment.

# Bibliography

1. Fox J. How to Write a Good Introduction Section. And Tell If Yours Is Bad.; 2016. Available from: <https://dynamicecology.wordpress.com/2016/04/19/many-introduction-sections-suck-heres-how-to-ensure-yours-doesnt/>.
2. Madigan MT, Martinko JM, Brock TD. Biology of Microorganisms. Upper Saddle River, NJ: Pearson Prentice Hall; 2006. OCLC: 57001814.
3. Schaechter M, Ingraham JL, Neidhardt FC. Microbe. ASM Press; 2006.
4. McArthur JV. Microbial Ecology: An Evolutionary Approach. 1st ed. Amsterdam ; Boston: Academic Press; 2006.
5. Menge DNL, Hedin LO, Pacala SW. Nitrogen and Phosphorus Limitation over Long-Term Ecosystem Development in Terrestrial Ecosystems. PLOS ONE. 2012 Aug;7(8):e42045. doi: [10.1371/journal.pone.0042045](https://doi.org/10.1371/journal.pone.0042045).
6. Hobbie JE, Hobbie EA. Microbes in Nature Are Limited by Carbon and Energy: The Starving-Survival Lifestyle in Soil and Consequences for Estimating Microbial Rates. Front Microbiol. 2013;4. doi: [10.3389/fmicb.2013.00324](https://doi.org/10.3389/fmicb.2013.00324).
7. Savageau MA. *Escherichia Coli* Habitats, Cell Types, and Molecular Mechanisms of Gene Control. Am Nat. 1983 Dec;122(6):732–44.
8. Savageau MA. Demand Theory of Gene Regulation. II. Quantitative Application to the Lactose and Maltose Operons of *Escherichia Coli*. Genetics. 1998 Jan;149(4):1677–91.
9. Blount ZD. The Unexhausted Potential of *E. Coli*. eLife. 2015 Mar;4:e05826. doi: [10.7554/eLife.05826](https://doi.org/10.7554/eLife.05826).

10. van Elsas JD, Semenov AV, Costa R, Trevors JT. Survival of *Escherichia Coli* in the Environment: Fundamental and Public Health Aspects. *ISME J.* 2011 Feb;5(2):173–83. doi: [10.1038/ismej.2010.80](https://doi.org/10.1038/ismej.2010.80).
11. Rothschild LJ, Mancinelli RL. Life in Extreme Environments. *Nature.* 2001 Feb;409(6823):1092–101. doi: [10.1038/35059215](https://doi.org/10.1038/35059215).
12. Nicholson WL, Moeller tPT Ralf, Horneck G. Transcriptomic Responses of Germinating *Bacillus Subtilis* Spores Exposed to 1.5 Years of Space and Simulated Martian Conditions on the EXPOSE-E Experiment PROTECT. *Astrobiology.* 2012 May;12(5):469–86. doi: [10.1089/ast.2011.0748](https://doi.org/10.1089/ast.2011.0748).
13. Zimmer C. *Microcosm: E. Coli and the New Science of Life.* Vintage Books USA; 2009.
14. Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martínez C, et al. EcoCyc: Fusing Model Organism Databases with Systems Biology. *Nucleic Acids Res.* 2013 Jan;41(Database issue):D605–12. doi: [10.1093/nar/gks1027](https://doi.org/10.1093/nar/gks1027).
15. Dawkins R. *The Selfish Gene.* Oxford University Press; 1976.
16. Schuetz R, Kuepfer L, Sauer U. Systematic Evaluation of Objective Functions for Predicting Intracellular Fluxes in *Escherichia Coli*. *Mol Syst Biol.* 2007 Jan;3:119. doi: [10.1038/msb4100162](https://doi.org/10.1038/msb4100162).
17. Frank SA. The Trade-off between Rate and Yield in the Design of Microbial Metabolism. *J Evol Biol.* 2010 Mar;23(3):609–13. doi: [10.1111/j.1420-9101.2010.01930.x](https://doi.org/10.1111/j.1420-9101.2010.01930.x).
18. MacLean RC. The Tragedy of the Commons in Microbial Populations: Insights from Theoretical, Comparative and Experimental Studies. *Heredity.* 2008 May;100(5):471–7. doi: [10.1038/sj.hdy.6801073](https://doi.org/10.1038/sj.hdy.6801073).
19. Schuster S, Pfeiffer T, Fell D. Is Maximization of Molar Yield in Metabolic Networks Favoured by Evolution? *J Theor Biol.* 2008 Jun;252(3):497–504. doi: [10.1016/j.jtbi.2007.12.008](https://doi.org/10.1016/j.jtbi.2007.12.008).
20. Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U. Multidimensional Optimality of Microbial Metabolism. *Science.* 2012 May;336(6081):601–4. doi: [10.1126/science.1216882](https://doi.org/10.1126/science.1216882).

21. Edwards JS, Ibarra RU, Palsson BO. In Silico Predictions of *Escherichia Coli* Metabolic Capabilities Are Consistent with Experimental Data. Nat Biotechnol. 2001 Feb;19(2):125–30. doi: [10.1038/84379](https://doi.org/10.1038/84379).
22. Ibarra RU, Edwards JS, Palsson BO. *Escherichia Coli* K-12 Undergoes Adaptive Evolution to Achieve in Silico Predicted Optimal Growth. Nature. 2002 Nov;420(6912):186–9. doi: [10.1038/nature01149](https://doi.org/10.1038/nature01149).
23. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, et al. Omic Data from Evolved *E. Coli* Are Consistent with Computed Optimal Growth from Genome-Scale Models. Mol Syst Biol. 2010 Jul;6:390. doi: [10.1038/msb.2010.47](https://doi.org/10.1038/msb.2010.47).
24. Molenaar D, van Berlo R, de Ridder D, Teusink B. Shifts in Growth Strategies Reflect Tradeoffs in Cellular Economics. Mol Syst Biol. 2009 Jan;5:323. doi: [10.1038/msb.2009.82](https://doi.org/10.1038/msb.2009.82).
25. Monod J. The Growth of Bacterial Cultures. Annu Rev Microbiol. 1949;doi: [10.1146/annurev.mi.03.100149.002103](https://doi.org/10.1146/annurev.mi.03.100149.002103).
26. Droop MR. Some Thoughts on Nutrient Limitation in Algae. J Phycol. 1973 Sep;9(3):264–272. doi: [10.1111/j.1529-8817.1973.tb04092.x](https://doi.org/10.1111/j.1529-8817.1973.tb04092.x).
27. Koch A. Why Can't a Cell Grow Infinitely Fast? Can J Microbiol. 1988; Available from: <http://www.nrcresearchpress.com/doi/abs/10.1139/m88-074>.
28. Schaechter M, MaalØe O, Kjeldgaard NO. Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of *Salmonella Typhimurium*. Microbiology. 1958;19(3):592–606. doi: [10.1099/00221287-19-3-592](https://doi.org/10.1099/00221287-19-3-592).
29. Bremer H, Dennis P. Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate. In: Neidhardt FC, III RC, Ingraham JL, Lin ECC, Low KB, Magasanik B, et al., editors. *Escherichia Coli* and *Salmonella*: Cellular and Molecular Biology. 2nd ed. Washington, DC: ASM Press; 1996. p. 1553–69.
30. Ehrenberg M, Bremer H, Dennis PP. Medium-Dependent Control of the Bacterial Growth Rate. Biochimie. 2013 Apr;95(4):643–58. doi: [10.1016/j.biochi.2012.11.012](https://doi.org/10.1016/j.biochi.2012.11.012).



31. Scott M, Hwa T. Bacterial Growth Laws and Their Applications. *Curr Opin Biotechnol.* 2011 Aug;22(4):559–65. doi: [10.1016/j.copbio.2011.04.014](https://doi.org/10.1016/j.copbio.2011.04.014).
32. Scott M, Klumpp S, Mateescu EM, Hwa T. Emergence of Robust Growth Laws from Optimal Regulation of Ribosome Synthesis. *Mol Syst Biol.* 2014 Aug;10:747. doi: [10.15252/msb.20145379](https://doi.org/10.15252/msb.20145379).
33. Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science.* 2010 Nov;330(6007):1099–102. doi: [10.1126/science.1192588](https://doi.org/10.1126/science.1192588).
34. Fishov I, Zaritsky A, Grover NB. On Microbial States of Growth. *Mol Microbiol.* 1995 Mar;15(5):789–94. doi: [10.1111/j.1365-2958.1995.tb02349.x](https://doi.org/10.1111/j.1365-2958.1995.tb02349.x).
35. Campbell A. Synchronization of Cell Division. *Bacteriol Rev.* 1957 Dec;21(4):263–72. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC180914/>.
36. Borirak O, Bekker M, Hellingwerf KJ. Molecular Physiology of the Dynamic Regulation of Carbon Catabolite Repression in *Escherichia Coli*. *Microbiology.* 2014 Jun;160(Pt6):1214–23. doi: [10.1099/mic.0.077289-0](https://doi.org/10.1099/mic.0.077289-0).
37. Orth JD, Thiele I, Palsson BØ. What Is Flux Balance Analysis? *Nat Biotechnol.* 2010 Mar;28(3):245–248. 00491. doi: [10.1038/nbt.1614](https://doi.org/10.1038/nbt.1614).
38. Swinnen IAM, Bernaerts K, Dens EJJ, Geeraerd AH, Van Impe JF. Predictive Modelling of the Microbial Lag Phase: A Review. *Int J Food Microbiol.* 2004 Jul;94(2):137–59. doi: [10.1016/j.ijfoodmicro.2004.01.006](https://doi.org/10.1016/j.ijfoodmicro.2004.01.006).
39. Ng H, Ingraham JL, Marr AG. Damage and Derepression in *Escherichia Coli* Resulting from Growth at Low Temperatures. *J Bacteriol.* 1962 Jan;84(2):331–9. Available from: <http://jb.asm.org/content/84/2/331>.
40. Dufrenne J, Delfgou E, Ritmeester W, Notermans S. The Effect of Previous Growth Conditions on the Lag Phase Time of Some Foodborne Pathogenic Micro-Organisms. *Int J Food Microbiol.* 1997 Jan;34(1):89–94. doi: [10.1016/S0168-1605\(96\)01170-1](https://doi.org/10.1016/S0168-1605(96)01170-1).

41. Shaw MK. Effect of Abrupt Temperature Shift on the Growth of Mesophilic and Psychrophilic Yeasts. *J Bacteriol.* 1967 Jan;93(4):1332–6. Available from: <http://jlb.asm.org/content/93/4/1332>.
42. McMeekin TA, Olley J, Ratkowsky DA, Ross T. Predictive Microbiology: Towards the Interface and Beyond. *Int J Food Microbiol.* 2002 Mar;73(2–3):395–407. doi: [10.1016/S0168-1605\(01\)00663-8](https://doi.org/10.1016/S0168-1605(01)00663-8).
43. Cheroutre-Vialette M, Lebert A. Application of Recurrent Neural Network to Predict Bacterial Growth in Dynamic Conditions. *Int J Food Microbiol.* 2002 Mar;73(2–3):107–18. doi: [10.1016/S0168-1605\(01\)00642-0](https://doi.org/10.1016/S0168-1605(01)00642-0).
44. Herbert D, Elsworth R, Telling RC. The Continuous Culture of Bacteria; a Theoretical and Experimental Study. *Microbiology.* 1956;14(3):601–22. doi: [10.1099/00221287-14-3-601](https://doi.org/10.1099/00221287-14-3-601).
45. Wang P, Robert L, Pelletier J, Dang WL, Taddei F, Wright A, et al. Robust Growth of Escherichia Coli. *Curr Biol CB.* 2010 Jun;20(12):1099–103. doi: [10.1016/j.cub.2010.04.045](https://doi.org/10.1016/j.cub.2010.04.045).
46. Chubukov V, Sauer U. Environmental Dependence of Stationary-Phase Metabolism in Bacillus Subtilis and Escherichia Coli. *Appl Environ Microbiol.* 2014 Jan;80(9):2901–2909. doi: [10.1128/AEM.00061-14](https://doi.org/10.1128/AEM.00061-14).
47. Stragier P, Losick R. Molecular Genetics of Sporulation in Bacillus Subtilis. *Annu Rev Genet.* 1996 Dec;30(1):297–341. doi: [10.1146/annurev.genet.30.1.297](https://doi.org/10.1146/annurev.genet.30.1.297).
48. González-Pastor JE. Cannibalism: A Social Behavior in Sporulating Bacillus Subtilis. *FEMS Microbiol Rev.* 2011 May;35(3):415–24. doi: [10.1111/j.1574-6976.2010.00253.x](https://doi.org/10.1111/j.1574-6976.2010.00253.x).
49. Servedio MR, Brandvain Y, Dhole S, Fitzpatrick CL, Goldberg EE, Stern CA, et al. Not Just a Theory—the Utility of Mathematical Models in Evolutionary Biology. *PLoS Biol.* 2014 Dec;12(12):e1002017. doi: [10.1371/journal.pbio.1002017](https://doi.org/10.1371/journal.pbio.1002017).
50. McGill B. A Calm and Balanced Case for Math in Biology; 2013. Available from: <https://dynamicecology.wordpress.com/2013/04/09/a-calm-and-balanced-case-for-math-in-biology/>.

51. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, et al. A Comprehensive Genome-scale Reconstruction of *Escherichia Coli* Metabolism—2011. *Mol Syst Biol.* 2011 Jan;7(1):535. doi: [10.1038/msb.2011.65](https://doi.org/10.1038/msb.2011.65).
52. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: CONstraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol.* 2013;7:74. doi: [10.1186/1752-0509-7-74](https://doi.org/10.1186/1752-0509-7-74).
53. Palsson B. *Systems Biology: Properties of Reconstructed Networks.* Cambridge University Press; 2011.
54. Varma A, Palsson B. Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic By-Product Secretion in Wild-Type *Escherichia Coli* W3110. *Appl Environ Microbiol.* 1994;60(10):3724–3731. Available from: <http://aem.asm.org/content/60/10/3724.short>.
55. Fong SS, Palsson BØ. Metabolic Gene-deletion Strains of *Escherichia Coli* Evolve to Computationally Predicted Growth Phenotypes. *Nat Genet.* 2004 Oct;36(10):1056–8. doi: [10.1038/ng1432](https://doi.org/10.1038/ng1432).
56. Kotte O, Zaugg JB, Heinemann M. Bacterial Adaptation through Distributed Sensing of Metabolic Fluxes. *Mol Syst Biol.* 2010 Jan;6(355):355. doi: [10.1038/msb.2010.10](https://doi.org/10.1038/msb.2010.10).
57. Peskov K, Mogilevskaya E, Demin O. Kinetic Modelling of Central Carbon Metabolism in *Escherichia Coli*. *FEBS J.* 2012 Sep;279(18):3374–85. doi: [10.1111/j.1742-4658.2012.08719.x](https://doi.org/10.1111/j.1742-4658.2012.08719.x).
58. Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival Jr B, et al. A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell.* 2012 Jul;150(2):389–401. doi: [10.1016/j.cell.2012.05.044](https://doi.org/10.1016/j.cell.2012.05.044).
59. Park JO, Rubin SA, Xu YF, Amador-Noguez D, Fan J, Shlomi T, et al. Metabolite Concentrations, Fluxes and Free Energies Imply Efficient Enzyme Usage. *Nat Chem Biol.* 2016 Jul;12(7):482–9. doi: [10.1038/nchembio.2077](https://doi.org/10.1038/nchembio.2077).
60. Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD. Absolute Metabolite Concentrations and Implied Enzyme Active Site Occupancy in *Escherichia Coli*. *Nat Chem Biol.* 2009 Aug;5(8):593–9. doi: [10.1038/nchembio.186](https://doi.org/10.1038/nchembio.186).

61. Büscher JM, Czernik D, Ewald JC, Sauer U, Zamboni N. Cross-Platform Comparison of Methods for Quantitative Metabolomics of Primary Metabolism. *Anal Chem.* 2009 Mar;81(6):2135–43. doi: [10.1021/ac8022857](https://doi.org/10.1021/ac8022857).
62. Jaqaman K, Danuser G. Linking Data to Models: Data Regression. *Nat Rev Mol Cell Biol.* 2006 Nov;7(11):813–9. doi: [10.1038/nrm2030](https://doi.org/10.1038/nrm2030).
63. Mendes P, Kell D. Non-Linear Optimization of Biochemical Pathways: Applications to Metabolic Engineering and Parameter Estimation. *Bioinforma Oxf Engl.* 1998;14(10):869–83.
64. Cho KH, Shin SY, Kolch W, Wolkenhauer O. Experimental Design in Systems Biology, Based on Parameter Sensitivity Analysis Using a Monte Carlo Method: A Case Study for the Tnf $\alpha$ -Mediated Nf- $\kappa$  b Signal Transduction Pathway. *Simulation.* 2003;79(12):726–39.
65. Brodersen R, Nielsen F, Christiansen JC, Andersen K. Characterization of Binding Equilibrium Data by a Variety of Fitted Isotherms. *Eur J Biochem.* 1987 Dec;169(3):487–95.
66. Rodriguez-Fernandez M, Mendes P, Banga JR. A Hybrid Approach for Efficient and Robust Parameter Estimation in Biochemical Pathways. *Biosystems.* 2006 Feb-Mar;83(2-3):248–65. doi: [10.1016/j.biosystems.2005.06.016](https://doi.org/10.1016/j.biosystems.2005.06.016).
67. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally Sloppy Parameter Sensitivities in Systems Biology Models. *PLoS Comput Biol.* 2007 Oct;3(10):1871–78. 00400. doi: [10.1371/journal.pcbi.0030189](https://doi.org/10.1371/journal.pcbi.0030189).
68. Ingram PJ, Stumpf MPH, Stark J. Network Motifs: Structure Does Not Determine Function. *BMC Genomics.* 2006 May;7:108. doi: [10.1186/1471-2164-7-108](https://doi.org/10.1186/1471-2164-7-108).
69. Mayo AE, Setty Y, Shavit S, Zaslaver A, Alon U. Plasticity of the Cis-Regulatory Input Function of a Gene. *PLoS Biol.* 2006 Apr;4(4):e45. doi: [10.1371/journal.pbio.0040045](https://doi.org/10.1371/journal.pbio.0040045).
70. Kauffman KJ, Prakash P, Edwards JS. Advances in Flux Balance Analysis. *Curr Opin Biotechnol.* 2003 Oct;14(5):491–6. doi: [10.1016/j.copbio.2003.08.001](https://doi.org/10.1016/j.copbio.2003.08.001).

71. Spiesser TW, Müller C, Schreiber G, Krantz M, Klipp E. Size Homeostasis Can Be Intrinsic to Growing Cell Populations and Explained without Size Sensing or Signalling. *FEBS J.* 2012 Nov;279(22):4213–30. doi: [10.1111/febs.12014](https://doi.org/10.1111/febs.12014).
72. van Dijken JP, Weusthuis RA, Pronk JT. Kinetics of Growth and Sugar Consumption in Yeasts. *Antonie Van Leeuwenhoek.* 1993;63(3-4):343–52. doi: [10.1007/BF00871229](https://doi.org/10.1007/BF00871229).
73. Vemuri GN, Altman E, Sangurdekar DP, Khodursky AB, Eiteman MA. Overflow Metabolism in Escherichia Coli during Steady-State Growth: Transcriptional Regulation and Effect of the Redox Ratio. *Appl Environ Microbiol.* 2006 Jan;72(5):3653–61. doi: [10.1128/AEM.72.5.3653-3661.2006](https://doi.org/10.1128/AEM.72.5.3653-3661.2006).
74. McKeehan WL. Glycolysis, Glutaminolysis and Cell Proliferation. *Cell Biol Int Rep.* 1982;6(7):635–50.
75. Hsu PP, Sabatini DM. Cancer Cell Metabolism: Warburg and Beyond. *Cell.* 2008 Sep;134(5):703–7. doi: [10.1016/j.cell.2008.08.021](https://doi.org/10.1016/j.cell.2008.08.021).
76. Volkmer B, Heinemann M. Condition-Dependent Cell Volume and Concentration of Escherichia Coli to Facilitate Data Conversion for Systems Biology Modeling. *PloS ONE.* 2011 Jul;6(7):e23126. doi: [10.1371/journal.pone.0023126](https://doi.org/10.1371/journal.pone.0023126).
77. Churchward G, Bremer H, Young R. Macromolecular Composition of Bacteria. *J Theor Biol.* 1982 Feb;94(3):651–70. doi: [10.1016/0022-5193\(82\)90305-8](https://doi.org/10.1016/0022-5193(82)90305-8).
78. Bremer H, Dennis P, Ehrenberg M. Free RNA Polymerase and Modeling Global Transcription in Escherichia Coli. *Biochimie.* 2003 Jun;85(6):597–609. doi: [10.1016/S0300-9084\(03\)00105-6](https://doi.org/10.1016/S0300-9084(03)00105-6).
79. Bremer H, Dennis P. Feedback Control of Ribosome Function in Escherichia Coli. *Biochimie.* 2008 Mar;90(3):493–499. doi: [10.1016/j.biochi.2007.10.008](https://doi.org/10.1016/j.biochi.2007.10.008).
80. Davey HM, Kell DB. Flow Cytometry and Cell Sorting of Heterogeneous Microbial Populations: The Importance of Single-Cell Analyses. *Microbiol Rev.* 1996 Jan;60(4):641–96. Available from: <http://mibr.asm.org/content/60/4/641>.

81. Bakshi S, Siryaporn A, Goulian M, Weisshaar JC. Superresolution Imaging of Ribosomes and RNA Polymerase in Live *Escherichia Coli* Cells. *Mol Microbiol.* 2012;85(1):21–38. doi: [10.1111/j.1365-2958.2012.08081.x](https://doi.org/10.1111/j.1365-2958.2012.08081.x).
82. Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S. Bacterial Persistence as a Phenotypic Switch. *Science.* 2004 Sep;305(5690):1622–5. doi: [10.1126/science.1099390](https://doi.org/10.1126/science.1099390).
83. Booth IR. Stress and the Single Cell: Intrapopulation Diversity Is a Mechanism to Ensure Survival upon Exposure to Stress. *Int J Food Microbiol.* 2002 Sep;78(1–2):19–30. doi: [10.1016/S0168-1605\(02\)00239-8](https://doi.org/10.1016/S0168-1605(02)00239-8).
84. Sumner ER, Avery SV. Phenotypic Heterogeneity: Differential Stress Resistance among Individual Cells of the Yeast *Saccharomyces Cerevisiae*. *Microbiology.* 2002;148(2):345–51. doi: [10.1099/00221287-148-2-345](https://doi.org/10.1099/00221287-148-2-345).
85. True HL, Lindquist SL. A Yeast Prion Provides a Mechanism for Genetic Variation and Phenotypic Diversity. *Nature.* 2000 Sep;407(6803):477–83. doi: [10.1038/35035005](https://doi.org/10.1038/35035005).
86. Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB. Noise Minimization in Eukaryotic Gene Expression. *PLOS Biol.* 2004 Apr;2(6):e137. doi: [10.1371/journal.pbio.0020137](https://doi.org/10.1371/journal.pbio.0020137).
87. Raser JM, O’Shea EK. Control of Stochasticity in Eukaryotic Gene Expression. *Science.* 2004 Jun;304(5678):1811–4. doi: [10.1126/science.1098641](https://doi.org/10.1126/science.1098641).
88. Hoskisson PA, Hobbs G. Continuous Culture – Making a Comeback? *Microbiology.* 2005;151(10):3153–9. doi: [10.1099/mic.0.27924-0](https://doi.org/10.1099/mic.0.27924-0).
89. Kremling A, Geiselmann J, Ropers D, de Jong H. Understanding Carbon Catabolite Repression in *Escherichia Coli* Using Quantitative Models. *Trends Microbiol.* 2015;23(2):99–109. doi: [10.1016/j.tim.2014.11.002](https://doi.org/10.1016/j.tim.2014.11.002).
90. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of *Escherichia Coli* K-12 In-frame, Single-gene Knockout Mutants: The Keio Collection. *Mol Syst Biol.* 2006 Jan;2(1):2006.0008. doi: [10.1038/msb4100050](https://doi.org/10.1038/msb4100050).

91. Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, et al. A Comprehensive Library of Fluorescent Transcriptional Reporters for Escherichia Coli. *Nat Methods*. 2006 Aug;3(8):623–628. doi: [10.1038/nmeth895](https://doi.org/10.1038/nmeth895).
92. Yamamoto N, Nakahigashi K, Nakamichi T, Yoshino M, Takai Y, Touda Y, et al. Update on the Keio Collection of Escherichia Coli Single-gene Deletion Mutants. *Mol Syst Biol*. 2009 Jan;5(1):335. doi: [10.1038/msb.2009.92](https://doi.org/10.1038/msb.2009.92).
93. Gerosa L, Kochanowski K, Heinemann M, Sauer U. Dissecting Specific and Global Transcriptional Regulation of Bacterial Gene Expression. *Mol Syst Biol*. 2013 Apr;9(658):1–11. 00008. doi: [10.1038/msb.2013.14](https://doi.org/10.1038/msb.2013.14).
94. Berthoumieux S, de Jong H, Baptist G, Pinel C, Ranquet C, Ropers D, et al. Shared Control of Gene Expression in Bacteria by Transcription Factors and Global Physiology of the Cell. *Mol Syst Biol*. 2013 Jan;9(1). doi: [10.1038/msb.2012.70](https://doi.org/10.1038/msb.2012.70).
95. Keren L, Zackay O, Lotan-Pompan M, Barenholz U, Dekel E, Sasson V, et al. Promoters Maintain Their Relative Activity Levels under Different Growth Conditions. *Mol Syst Biol*. 2013 Oct;9(1). 00000. doi: [10.1038/msb.2013.59](https://doi.org/10.1038/msb.2013.59).
96. Ronen M, Rosenberg R, Shraiman BI, Alon U. Assigning Numbers to the Arrows: Parameterizing a Gene Regulation Network by Using Accurate Expression Kinetics. *Proc Natl Acad Sci*. 2002 Jun;99(16):10555–60. doi: [10.1073/pnas.152046799](https://doi.org/10.1073/pnas.152046799).
97. Stefan D, Pinel C, Pinhal S, Cinquemani E, Geiselmann J, de Jong H. Inference of Quantitative Models of Bacterial Promoters from Time-Series Reporter Gene Data. *PLOS Comput Biol*. 2015 Jan;11(1):e1004028. doi: [10.1371/journal.pcbi.1004028](https://doi.org/10.1371/journal.pcbi.1004028).
98. Bekker M, Kramer G, Hartog AF, Wagner MJ, de Koster CG, Hellingwerf KJ, et al. Changes in the Redox State and Composition of the Quinone Pool of Escherichia Coli during Aerobic Batch-Culture Growth. *Microbiology*. 2007;153(6):1974–80. doi: [10.1099/mic.0.2007/006098-0](https://doi.org/10.1099/mic.0.2007/006098-0).
99. Ackerman RS, Cozzarelli NR, Epstein W. Accumulation of Toxic Concentrations of Methylglyoxal by Wild-Type Escherichia Coli K-



12. J Bacteriol. 1974 Jan;119(2):357–62. Available from: <http://jba.asm.org/content/119/2/357>.
100. Lenski RE, Riley MA. Chemical Warfare from an Ecological Perspective. Proc Natl Acad Sci. 2002 Jan;99(2):556–8. doi: [10.1073/pnas.022641999](https://doi.org/10.1073/pnas.022641999).
101. Luli GW, Strohl WR. Comparison of Growth, Acetate Production, and Acetate Inhibition of Escherichia Coli Strains in Batch and Fed-Batch Fermentations. Appl Environ Microbiol. 1990 Jan;56(4):1004–11. Available from: <http://aem.asm.org/content/56/4/1004>.
102. Myers J, Clark LB. Culture Conditions and the Development of the Photosynthetic Mechanism. J Gen Physiol. 1944 Nov;28(2):103–12. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2142662/>.
103. Novick A, Szilard L. Experiments with the Chemostat on Spontaneous Mutations of Bacteria. Proc Natl Acad Sci. 1950;36(12):708–19.
104. Kolkman A, Olsthoorn MMA, Heeremans CEM, Heck AJR, Slijper M. Comparative Proteome Analysis of Saccharomyces Cerevisiae Grown in Chemostat Cultures Limited for Glucose or Ethanol. Mol Cell Proteomics. 2005 Jan;4(1):1–11. doi: [10.1074/mcp.M400087-MCP200](https://doi.org/10.1074/mcp.M400087-MCP200).
105. Boer VM, de Winde JH, Pronk JT, Piper MDW. The Genome-Wide Transcriptional Responses of Saccharomyces Cerevisiae Grown on Glucose in Aerobic Chemostat Cultures Limited for Carbon, Nitrogen, Phosphorus, or Sulfur. J Biol Chem. 2003 Jan;278(5):3265–74. doi: [10.1074/jbc.M209759200](https://doi.org/10.1074/jbc.M209759200).
106. Betts JI, Baganz F. Miniature Bioreactors: Current Practices and Future Opportunities. Microb Cell Factories. 2006 Jan;5:21. doi: [10.1186/1475-2859-5-21](https://doi.org/10.1186/1475-2859-5-21).
107. Levy S, Ihmels J, Carmi M, Weinberger A, Friedlander G, Barkai N. Strategy of Transcription Regulation in the Budding Yeast. PLoS ONE. 2007 Feb;2(2):e250. 00041. doi: [10.1371/journal.pone.0000250](https://doi.org/10.1371/journal.pone.0000250).
108. Levy S, Barkai N. Coordination of Gene Expression with Growth Rate: A Feedback or a Feed-Forward Strategy? FEBS Lett. 2009 Dec;583(24):3974–3978. 00016. doi: [10.1016/j.febslet.2009.10.071](https://doi.org/10.1016/j.febslet.2009.10.071).



109. Madar D, Dekel E, Bren A, Zimmer A, Porat Z, Alon U. Promoter Activity Dynamics in the Lag Phase of *Escherichia Coli*. *BMC Syst Biol*. 2013 Dec;7(1):136. doi: [10.1186/1752-0509-7-136](https://doi.org/10.1186/1752-0509-7-136).
110. Kjeldgaard NO. The Kinetics of Ribonucleic Acid- and Protein Formation in *Salmonella Typhimurium* during the Transition between Different States of Balanced Growth. *Biochim Biophys Acta*. 1961 Apr;49(1):64–76. doi: [10.1016/0006-3002\(61\)90870-8](https://doi.org/10.1016/0006-3002(61)90870-8).
111. Schaechter M. Patterns of Cellular Control during Unbalanced Growth. *Cold Spring Harb Symp Quant Biol*. 1961 Jan;26:53–62. doi: [10.1101/SQB.1961.026.01.011](https://doi.org/10.1101/SQB.1961.026.01.011).
112. Johnsen K, Molin S, Karlström O, Maaløe O. Control of Protein Synthesis in *Escherichia Coli*: Analysis of an Energy Source Shift-Down. *J Bacteriol*. 1977 Jan;131(1):18–29.
113. Stengel RF. *Optimal Control and Estimation*. Mineola, NY: Dover Publications; 1994.
114. Carlson DA, Haurie AB, Leizarowitz A. *Infinite Horizon Optimal Control*. Berlin, Heidelberg: Springer; 1991.
115. Bonnans F, Martinon P, Grélard V. *Bocop - A Collection of Examples*. Inria; 2012. Eprint: <https://hal.inria.fr/hal-00726992>.
116. Bosdriesz E, Molenaar D, Teusink B, Bruggeman FJ. How Fast-Growing Bacteria Robustly Tune Their Ribosome Concentration to Approximate Growth-Rate Maximization. *FEBS J*. 2015 Mar;282:2029–204. doi: [10.1111/febs.13258](https://doi.org/10.1111/febs.13258).
117. Andersen KB, von Meyenburg K. Are Growth Rates of *Escherichia Coli* in Batch Cultures Limited by Respiration? *J Bacteriol*. 1980 Oct;144(1):114–23.
118. Kailath T, Sayed AH, Hassibi B. *Linear Estimation*. vol. 1. Prentice Hall Upper Saddle River, NJ; 2000.
119. Jazwinski AH. *Stochastic Processes and Filtering Theory*. Courier Corporation; 2007.
120. Giordano N, Mairet F, Gouzé JL, Geiselman J, de Jong H. Dynamical Allocation of Cellular Resources as an Optimal Control Problem: Novel Insights into Microbial Growth Strategies. *PLOS Comput Biol*. 2016 Mar;12(3):e1004802. doi: [10.1371/journal.pcbi.1004802](https://doi.org/10.1371/journal.pcbi.1004802).

121. Maitra A, Dill KA. Bacterial Growth Laws Reflect the Evolutionary Importance of Energy Efficiency. *Proc Natl Acad Sci USA*. 2014 Dec;112(2):406–11. doi: [10.1073/pnas.1421138111](https://doi.org/10.1073/pnas.1421138111).
122. Weiße AY, Oyarzún DA, Danos V, Swain PS. Mechanistic Links between Cellular Trade-Offs, Gene Expression, and Growth. *Proc Natl Acad Sci USA*. 2015 Mar;112(9):1038–47. doi: [10.1073/pnas.1416533112](https://doi.org/10.1073/pnas.1416533112).
123. Pavlov MY, Ehrenberg M. Optimal Control of Gene Expression for Fast Proteome Adaptation to Environmental Change. *Proc Natl Acad Sci USA*. 2013 Dec;110(51):20527–32. doi: [10.1073/pnas.1309356110](https://doi.org/10.1073/pnas.1309356110).
124. van den Berg HA, Kiselev YN, Kooijman SaLM, Orlov MV. Optimal Allocation between Nutrient Uptake and Growth in a Microbial Trichome. *J Math Biol*. 1998 Jul;37(1):28–48. doi: [10.1007/s002850050118](https://doi.org/10.1007/s002850050118).
125. Waldherr S, Oyarzún DA, Bockmayr A. Dynamic Optimization of Metabolic Networks Coupled with Gene Expression. *J Theor Biol*. 2015 Jan;365:469–85. doi: [10.1016/j.jtbi.2014.10.035](https://doi.org/10.1016/j.jtbi.2014.10.035).
126. Dalebroux ZD, Swanson MS. ppGpp: Magic beyond RNA Polymerase. *Nat Rev Microbiol*. 2012;10(3):203–12. doi: [10.1038/nrmicro2720](https://doi.org/10.1038/nrmicro2720).
127. Potrykus K, Cashel M. (P)ppGpp: Still Magical? *Annu Rev Microbiol*. 2008;62:35–51. doi: [10.1146/annurev.micro.62.081307.162903](https://doi.org/10.1146/annurev.micro.62.081307.162903).
128. Hauryliuk V, Atkinson GC, Murakami KS, Tenson T, Gerdes K. Recent Functional Insights into the Role of (p)ppGpp in Bacterial Physiology. *Nat Rev Microbiol*. 2015;13(5):298–309. doi: [10.1038/nrmicro3448](https://doi.org/10.1038/nrmicro3448).
129. Iglesias PA, Ingalls BP. *Control Theory and Systems Biology*. MIT Press; 2010.
130. Sipper M. Fifty Years of Research on Self-Replication: An Overview. *Artif Life*. 1998;4(3):237–57. doi: [10.1162/106454698568576](https://doi.org/10.1162/106454698568576).
131. Flamm C, Endler L, Müller S, Widder S, Schuster P. A Minimal and Self-Consistent in Silico Cell Model Based on Macromolecular Interactions. *Philos Trans R Soc B*. 2007;362(1486):1831–9. doi: [10.1098/rstb.2007.2075](https://doi.org/10.1098/rstb.2007.2075).

132. Klumpp S, Scott M, Pedersen S, Hwa T. Molecular Crowding Limits Translation and Cell Growth. *Proc Natl Acad Sci USA*. 2013 Oct;110(42):16754–9. doi: [10.1073/pnas.1310377110](https://doi.org/10.1073/pnas.1310377110).
133. Chubukov V, Gerosa L, Kochanowski K, Sauer U. Coordination of Microbial Metabolism. *Nat Rev Microbiol*. 2014 Mar;12:327–40. doi: [10.1038/nrmicro3238](https://doi.org/10.1038/nrmicro3238).
134. Deutscher J, Francke C, Postma PW. How Phosphotransferase System-Related Protein Phosphorylation Regulates Carbohydrate Metabolism in Bacteria. *Microbiol Mol Biol Rev*. 2006;70(4):939–1031. doi: [10.1128/MMBR.00024-06](https://doi.org/10.1128/MMBR.00024-06).
135. Görke B, Stülke J. Carbon Catabolite Repression in Bacteria: Many Ways to Make the Most out of Nutrients. *Nat Rev Microbiol*. 2008;6(8):613–24. doi: [10.1038/nrmicro1932](https://doi.org/10.1038/nrmicro1932).
136. English BP, Hauryliuk V, Sanamrad A, Tankov S, Dekker NH, Elf J. Single-Molecule Investigations of the Stringent Response Machinery in Living Bacterial Cells. *Proc Natl Acad Sci USA*. 2011;108(31):365–73. doi: [10.1073/pnas.1102255108](https://doi.org/10.1073/pnas.1102255108).
137. Dennis PP, Ehrenberg M, Bremer H. Control of rRNA Synthesis in *Escherichia Coli*: A Systems Biology Approach. *Microbiol Mol Biol Rev*. 2004 Jan;68(4):639–68. doi: [10.1128/MMBR.68.4.639-668.2004](https://doi.org/10.1128/MMBR.68.4.639-668.2004).
138. Keener J, Nomura M. Regulation of Ribosome Synthesis. In: Neidhardt FC, III RC, Ingraham JL, Lin ECC, Low KB, Magasanik B, et al., editors. *Escherichia Coli and Salmonella: Cellular and Molecular Biology*. 2nd ed. Washington, DC: ASM Press; 1996. p. 1417–31.
139. Murray HD, Schneider DA, Gourse RL. Control of rRNA Expression by Small Molecules Is Dynamic and Nonredundant. *Mol Cell*. 2003;12(1):125–34. doi: [10.1016/S1097-2765\(03\)00266-1](https://doi.org/10.1016/S1097-2765(03)00266-1).
140. Macevicz S, Oster G. Modeling Social Insect Populations II: Optimal Reproductive Strategies in Annual Eusocial Insect Colonies. *Behav Ecol Sociobiol*. 1976;1(3):265–82. doi: [10.1007/BF00300068](https://doi.org/10.1007/BF00300068).
141. Itzkovitz S, Blat IC, Jacks T, Clevers H, van Oudenaarden A. Optimality in the Development of Intestinal Crypts. *Cell*. 2012 Feb;148(3):608–19. doi: [10.1016/j.cell.2011.12.025](https://doi.org/10.1016/j.cell.2011.12.025).

142. Bartl M, Li P, Schuster S. Modelling the Optimal Timing in Metabolic Pathway Activation—Use of Pontryagin’s Maximum Principle and Role of the Golden Section. *Biosystems*. 2010 Jul;101(1):67–77. doi: [10.1016/j.biosystems.2010.04.007](https://doi.org/10.1016/j.biosystems.2010.04.007).
143. Oyarzún DA, Ingalls BP, Middleton RH, Kalamatianos D. Sequential Activation of Metabolic Pathways: A Dynamic Optimization Approach. *Bull Math Biol*. 2009;71:1851–72. doi: [10.1007/s11538-009-9427-5](https://doi.org/10.1007/s11538-009-9427-5).
144. Kiselev YN, Avvakumov SN, Orlov MV. Resource Allocation Problem in a Two-Sector Economic Model of Special Form. *Differ Equ*. 2009 Dec;45(12):1791–810. 00003. doi: [10.1134/S0012266109120106](https://doi.org/10.1134/S0012266109120106).
145. Potrykus K, Murphy H, Philippe N, Cashel M. ppGpp Is the Major Source of Growth Rate Control in *E. Coli*. *Env Microbiol*. 2011;13(3):563–75. doi: [10.1111/j.1462-2920.2010.02357.x](https://doi.org/10.1111/j.1462-2920.2010.02357.x).
146. Traxler MF, Chang DE, Conway T. Guanosine 3’,5’-Bispyrophosphate Coordinates Global Gene Expression during Glucose-Lactose Diauxie in *Escherichia Coli*. *Proc Natl Acad Sci USA*. 2006;103(7):2374–9. doi: [10.1073/pnas.0510995103](https://doi.org/10.1073/pnas.0510995103).
147. Traxler MF, Summers SM, Nguyen HT, Zacharia VM, Hightower GA, Smith JT, et al. The Global, ppGpp-Mediated Stringent Response to Amino Acid Starvation in *Escherichia Coli*. *Mol Microbiol*. 2008;68(5):1128–48. doi: [10.1111/j.1365-2958.2008.06229.x](https://doi.org/10.1111/j.1365-2958.2008.06229.x).
148. Gaca AO, Colomer-Winter C, Lemos JA. Many Means to a Common End: The Intricacies of (p)ppGpp Metabolism and Its Control of Bacterial Homeostasis. *J Bacteriol*. 2015;197(7):1146–56. doi: [10.1128/JB.02577-14](https://doi.org/10.1128/JB.02577-14).
149. Liu K, Bittner AN, Wang JD. Diversity in (p)ppGpp Metabolism and Effectors. *Curr Opin Microbiol*. 2015;24:72–9. doi: [10.1016/j.mib.2015.01.012](https://doi.org/10.1016/j.mib.2015.01.012).
150. Kalisky T, Dekel E, Alon U. Cost-Benefit Theory and Optimal Design of Gene Regulation Functions. *Phys Biol*. 2007 Nov;4(4):229–45. doi: [10.1088/1478-3975/4/4/001](https://doi.org/10.1088/1478-3975/4/4/001).
151. Poelwijk FJ, de Vos MGJ, Tans SJ. Tradeoffs and Optimality in the Evolution of Gene Regulation. *Cell*. 2011 Aug;146(3):462–70. 00023. doi: [10.1016/j.cell.2011.06.035](https://doi.org/10.1016/j.cell.2011.06.035).

152. Gausing K. Regulation of Ribosome Synthesis in *E. Coli*. In: Chambliss G, Craven GR, Davies J, Davis K, Kahan L, Nomura M, editors. Ribosomes: Structure, Function and Genetics. Baltimore: University Park Press; 1980. p. 693–718.
153. Zengel JM, Lindahl L. Transcription of Ribosomal Genes during a Nutritional Shift-up of *Escherichia Coli*. *J Bacteriol*. 1986;167(3):1095–7.
154. Friesen JD, Fiil NP, von Meyenburg K. Synthesis and Turnover of Basal Level Guanosine Tetrphosphate in *Escherichia Coli*. *J Biol Chem*. 1975;250(1):304–9.
155. Young JW, Locke JC, Altinok A, Rosenfeld N, Bacarian T, Swain PS, et al. Measuring Single-Cell Gene Expression Dynamics in Bacteria Using Fluorescence Time-Lapse Microscopy. *Nat Protoc*. 2011;7(1):80–8. doi: [10.1038/nprot.2011.432](https://doi.org/10.1038/nprot.2011.432).
156. Duncombe TA, Tentori AM, Herr AE. Microfluidics: Reframing Biological Enquiry. *Nat Rev Mol Cell Biol*. 2015;16(9):554–67. doi: [10.1038/nrm4041](https://doi.org/10.1038/nrm4041).
157. Molin S, Meyenburg KV, Maaloe O, Hansen MT, Pato ML. Control of Ribosome Synthesis in *Escherichia Coli*: Analysis of an Energy Source Shift-Down. *J Bacteriol*. 1977;131(1):7–17.
158. Venayak N, Anesiadis N, Cluett WR, Mahadevan R. Engineering Metabolism through Dynamic Control. *Curr Opin Biotechnol*. 2015;34:142–52. doi: [10.1016/j.copbio.2014.12.022](https://doi.org/10.1016/j.copbio.2014.12.022).
159. Dahl RH, Zhang F, Alonso-Gutierrez J, Baidoo E, Batth TS, Redding-Johanson AM, et al. Engineering Dynamic Pathway Regulation Using Stress-Response Promoters. *Nat Biotechnol*. 2013;31(11):1039–46. doi: [10.1038/nbt.2689](https://doi.org/10.1038/nbt.2689).
160. Xu P, Li L, Zhang F, Stephanopoulos G, Koffas M. Improving Fatty Acids Production by Engineering Dynamic Pathway Regulation and Metabolic Control. *Proc Natl Acad Sci USA*. 2014;111(31):11299–304. doi: [10.1073/pnas.1406401111](https://doi.org/10.1073/pnas.1406401111).
161. Izard J, Gomez Balderas C, Ropers D, Lacour S, Song X, Yang Y, et al. A Synthetic Growth Switch Based on Controlled Expression of RNA Polymerase. *Mol Syst Biol*. 2015;11(11):840. doi: [10.15252/msb.20156382](https://doi.org/10.15252/msb.20156382).

162. Storn R, Price K. Differential Evolution – a Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *J Glob Optim.* 1997 Dec;11(4):341–59. doi: [10.1023/A:1008202821328](https://doi.org/10.1023/A:1008202821328).
163. Borisov VF. Fuller’s Phenomenon: Review. *J Math Sci.* 2000 Jul;100(4):2311–54. doi: [10.1007/s10958-000-0001-9](https://doi.org/10.1007/s10958-000-0001-9).
164. Filippov AF. *Differential Equations with Discontinuous Right-hand Sides.* Dordrecht: Kluwer Academic Publishers; 1988. doi: [10.1007/978-94-015-7793-9](https://doi.org/10.1007/978-94-015-7793-9).
165. Cohen SD, Hindmarsh AC. CVODE, a Stiff/Nonstiff ODE Solver in C. *Comput Phys.* 1996;10(5):138–43. doi: [10.1063/1.4822377](https://doi.org/10.1063/1.4822377).
166. Hindmarsh AC, Brown PN, Grant KE, Lee SL, Serban R, Shumaker DE, et al. SUNDIALS: Suite of Nonlinear and Differential/Algebraic Equation Solvers. *ACM Trans Math Softw.* 2005;31(3):363–96. doi: [10.1145/1089014.1089020](https://doi.org/10.1145/1089014.1089020).
167. Uter NT, Perona JJ. Long-Range Intramolecular Signaling in a tRNA Synthetase Complex Revealed by Pre-Steady-State Kinetics. *Proc Natl Acad Sci USA.* 2004 Oct;101(40):14396–401. doi: [10.1073/pnas.0404017101](https://doi.org/10.1073/pnas.0404017101).
168. Freist W, Gauss DH, Ibba M, Söll D. Glutaminyl-tRNA Synthetase. *Biol Chem.* 1997 Oct;378(10):1103–17.
169. Hachiya T, Terashima I, Noguchi K. Increase in Respiratory Cost at High Growth Temperature Is Attributed to High Protein Turnover Cost in *Petunia x Hybrida* Petals. *Plant Cell Env.* 2007 Oct;30(10):1269–83. doi: [10.1111/j.1365-3040.2007.01701.x](https://doi.org/10.1111/j.1365-3040.2007.01701.x).
170. Yamamoto T, Izumi S, Gekko K. Mass Spectrometry of Hydrogen/Deuterium Exchange in 70S Ribosomal Proteins from *E. Coli*. *FEBS Lett.* 2006 Jun;580(15):3638–42. doi: [10.1016/j.febslet.2006.05.049](https://doi.org/10.1016/j.febslet.2006.05.049).
171. Zhou Y, Vazquez A, Wise A, Warita T, Warita K, Bar-Joseph Z, et al. Carbon Catabolite Repression Correlates with the Maintenance of near Invariant Molecular Crowding in Proliferating *E. Coli* Cells. *BMC Syst Biol.* 2013 Dec;7:138. doi: [10.1186/1752-0509-7-138](https://doi.org/10.1186/1752-0509-7-138).
172. Zimmerman SB, Trach SO. Estimation of Macromolecule Concentrations and Excluded Volume Effects for the Cytoplasm of *Escherichia*

- Coli*. J Mol Biol. 1991 Dec;222(3):599–620. doi: [10.1016/0022-2836\(91\)90499-V](https://doi.org/10.1016/0022-2836(91)90499-V).
173. McGuffee SR, Elcock AH. Diffusion, Crowding & Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm. PLoS Comput Biol. 2010 Mar;6(3):e1000694. doi: [10.1371/journal.pcbi.1000694](https://doi.org/10.1371/journal.pcbi.1000694).
174. Marchal C. Chattering Arcs and Chattering Controls. J Optim Theory Appl. 2013 Aug;11(5):441–68. doi: [10.1007/BF00935659](https://doi.org/10.1007/BF00935659).
175. Trélat E, Zuazua E. The Turnpike Property in Finite-Dimensional Nonlinear Optimal Control. J Differ Equ. 2015 Jan;258(1):81–114. doi: [10.1016/j.jde.2014.09.005](https://doi.org/10.1016/j.jde.2014.09.005).
176. Weinersmith Z. Biologists Are Weird [Webcomic]; 2016. Available from: <http://www.smbc-comics.com/comic/biology>.
177. Hui S, Silverman JM, Chen SS, Erickson DW, Basan M, Wang J, et al. Quantitative Proteomic Analysis Reveals a Simple Strategy of Global Resource Allocation in Bacteria. Mol Syst Biol. 2015 Feb;11(1):784.
178. Zulkower V, Page M, Ropers D, Geiselmann J, de Jong H. Robust Reconstruction of Gene Expression Profiles from Reporter Gene Data Using Linear Inversion. Bioinformatics. 2015 Jun;31(12):i71–9. doi: [10.1093/bioinformatics/btv246](https://doi.org/10.1093/bioinformatics/btv246).
179. Gefen O, Fridman O, Ronin I, Balaban NQ. Direct Observation of Single Stationary-Phase Bacteria Reveals a Surprisingly Long Period of Constant Protein Production Activity. Proc Natl Acad Sci U S A. 2014 Jan;111(1):556–61. doi: [10.1073/pnas.1314114111](https://doi.org/10.1073/pnas.1314114111).
180. Klein J, Leupold S, Biegler I, Biedendieck R, Münch R, Jahn D. TLM-Tracker: Software for Cell Segmentation, Tracking and Lineage Analysis in Time-Lapse Microscopy Movies. Bioinformatics. 2012 Jan;28(17):2276–7. doi: [10.1093/bioinformatics/bts424](https://doi.org/10.1093/bioinformatics/bts424).
181. Wang Q, Niemi J, Tan CM, You L, West M. Image Segmentation and Dynamic Lineage Analysis in Single-Cell Fluorescence Microscopy. Cytometry A. 2010 Jan;77A(1):101–10. doi: [10.1002/cyto.a.20812](https://doi.org/10.1002/cyto.a.20812).
182. Paintdakhi A, Parry B, Campos M, Irnov I, Elf J, Surovtsev I, et al. Oufiti: An Integrated Software Package for High-Accuracy, High-Throughput Quantitative Microscopy Analysis: Oufiti: Image Analysis Software. Mol Microbiol. 2016 Feb;99(4):767–77. doi: [10.1111/mmi.13264](https://doi.org/10.1111/mmi.13264).



183. Sachs CC, Grünberger A, Helfrich S, Probst C, Wiechert W, Kohlheyer D, et al. Image-Based Single Cell Profiling: High-Throughput Processing of Mother Machine Experiments. *PLOS ONE*. 2016 Sep;11(9):e0163453. doi: [10.1371/journal.pone.0163453](https://doi.org/10.1371/journal.pone.0163453).
184. Cinquemani E, Laroute V, Coccagn-Bousquet M, de Jong H, Ropers D. Estimation of Time-Varying Growth, Uptake and Excretion Rates from Dynamic Metabolomics Data. *Bioinformatics*. 2017; Accepted for publication.
185. Kalman RE. A New Approach to Linear Filtering and Prediction Problems. *J Basic Eng*. 1960 Mar;82(1):35–45. doi: [10.1115/1.3662552](https://doi.org/10.1115/1.3662552).
186. Julier SJ, Uhlmann JK. New Extension of the Kalman Filter to Non-linear Systems. vol. 3068; 1997. p. 182–93. doi: [10.1117/12.280797](https://doi.org/10.1117/12.280797).
187. Tikhonov AN, Arsenin VI. *Solutions of Ill-Posed Problems*. Winston; 1977.
188. Tikhonov A. Solution of Incorrectly Formulated Problems and the Regularization Method. In: *Soviet Math. Dokl.* vol. 5; 1963. p. 1035–8.
189. De Nicolao G, Sparacino G, Cobelli C. Nonparametric Input Estimation in Physiological Systems: Problems, Methods, and Case Studies. *Automatica*. 1997 May;33(5):851–70. doi: [10.1016/S0005-1098\(96\)00254-3](https://doi.org/10.1016/S0005-1098(96)00254-3).
190. FluoBacTracker: An ImageJ Plugin Designed to Segment and Track Growing *E. Coli* Cells; 2016. Available from: <https://fluobactracker.inrialpes.fr/index.php>.
191. Golub GH, Heath M, Wahba G. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*. 1979 May;21(2):215–23. doi: [10.1080/00401706.1979.10489751](https://doi.org/10.1080/00401706.1979.10489751).
192. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat Methods*. 2009 May;6(5):343–5. doi: [10.1038/nmeth.1318](https://doi.org/10.1038/nmeth.1318).
193. Sharan SK, Thomason LC, Kuznetsov SG, Court DL. Recombineering: A Homologous Recombination-Based Method of Genetic Engineering. *Nat Protoc*. 2009;4(2):206–23. doi: [10.1038/nprot.2008.227](https://doi.org/10.1038/nprot.2008.227).



194. Goto A, Doerr H. Open and Save SPE Image Files; 2005. Available from: <https://imagej.nih.gov/ij/plugins/spe.html>.
195. Guizar-Sicairos M, Thurman ST, Fienup JR. Efficient Subpixel Image Registration Algorithms. *Opt Lett*. 2008 Jan;33(2):156–8. doi: [10.1364/OL.33.000156](https://doi.org/10.1364/OL.33.000156).
196. Cross-Correlation (Phase Correlation); 2016. Available from: [http://scikit-image.org/docs/dev/auto\\_examples/plot\\_register\\_translation.html](http://scikit-image.org/docs/dev/auto_examples/plot_register_translation.html).
197. Mihalcescu I, Gateau MVM, Chelli B, Pinel C, Ravanat JL. Green Autofluorescence, a Double Edged Monitoring Tool for Bacterial Growth and Activity in Micro-Plates. *Phys Biol*. 2015;12(6):066016. doi: [10.1088/1478-3975/12/6/066016](https://doi.org/10.1088/1478-3975/12/6/066016).
198. Doherty GP, Bailey K, Lewis PJ. Stage-Specific Fluorescence Intensity of GFP and mCherry during Sporulation In *Bacillus Subtilis*. *BMC Res Notes*. 2010 Nov;3:303. doi: [10.1186/1756-0500-3-303](https://doi.org/10.1186/1756-0500-3-303).
199. @tomroud. Tom Roud on Twitter : "Apples Fall onto the Earth Because Natural Selection Eliminated Apples Falling towards the Sky #BiologistSpaceFacts"; 2016. Available from: <https://twitter.com/tomroud/status/708494284668977152>.
200. Falk D. *Universe on a T-Shirt: The Quest for the Theory of Everything*. Arcade Publishing; 2005.
201. Frank SA. Natural Selection. IV. The Price Equation\*: Price Equation. *J Evol Biol*. 2012 Jun;25(6):1002–19. doi: [10.1111/j.1420-9101.2012.02498.x](https://doi.org/10.1111/j.1420-9101.2012.02498.x).
202. Deamer DW, Fleischaker GR. *Origins of Life: The Central Concepts*. Jones & Bartlett Pub; 1994.
203. Benner SA. Defining Life. *Astrobiology*. 2010 Dec;10(10):1021–1030. doi: [10.1089/ast.2010.0524](https://doi.org/10.1089/ast.2010.0524).
204. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From Molecular to Modular Cell Biology. *Nature*. 1999;402:C47–52.
205. Arkin AP, Fletcher DA. Fast, Cheap and Somewhat in Control. *Genome Biol*. 2006;7:114. doi: [10.1186/gb-2006-7-8-114](https://doi.org/10.1186/gb-2006-7-8-114).

206. Guido NJ, Wang X, Adalsteinsson D, McMillen D, Hasty J, Cantor CR, et al. A Bottom-up Approach to Gene Regulation. *Nature*. 2006 Feb;439(7078):856–60. doi: [10.1038/nature04473](https://doi.org/10.1038/nature04473).
207. Geisel N. Constitutive versus Responsive Gene Expression Strategies for Growth in Changing Environments. *PLoS One*. 2011 Jan;6(11):e27033. 00005. doi: [10.1371/journal.pone.0027033](https://doi.org/10.1371/journal.pone.0027033).
208. López-Maury L, Marguerat S, Bähler J. Tuning Gene Expression to Changing Environments: From Rapid Responses to Evolutionary Adaptation. *Nat Rev Genet*. 2008 Aug;9(8):583–593. 00196. doi: [10.1038/nrg2398](https://doi.org/10.1038/nrg2398).
209. Lambert G, Kussel E. Memory and Fitness Optimization of Bacteria under Fluctuating Environments. *PLoS Genet*. 2014 Sep;10(9):e1004556. doi: [10.1371/journal.pgen.1004556](https://doi.org/10.1371/journal.pgen.1004556).
210. Kussell E, Leibler S. Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments. *Science*. 2005 Sep;309(5743):2075–2078. doi: [10.1126/science.1114383](https://doi.org/10.1126/science.1114383).
211. Shachrai I, Zaslaver A, Alon U, Dekel E. Cost of Unneeded Proteins in *E. Coli* Is Reduced after Several Generations in Exponential Growth. *Mol Cell*. 2010 Jun;38(5):758–67. doi: [10.1016/j.molcel.2010.04.015](https://doi.org/10.1016/j.molcel.2010.04.015).
212. Dong H, Nilsson L, Kurland CG. Gratuitous Overexpression of Genes in *Escherichia Coli* Leads to Growth Inhibition and Ribosome Destruction. *J Bacteriol*. 1995 Jan;177(6):1497–504. doi: [10.1128/jb.177.6.1497-1504.1995](https://doi.org/10.1128/jb.177.6.1497-1504.1995).
213. Dekel E, Mangan S, Alon U. Environmental Selection of the Feed-Forward Loop Circuit in Gene-Regulation Networks. *Phys Biol*. 2005;2(2):81. doi: [10.1088/1478-3975/2/2/001](https://doi.org/10.1088/1478-3975/2/2/001).
214. Short MB, Solari CA, Ganguly S, Powers TR, Kessler JO, Goldstein RE. Flows Driven by Flagella of Multicellular Organisms Enhance Long-Range Molecular Transport. *Proc Natl Acad Sci*. 2006 May;103(22):8315–9. doi: [10.1073/pnas.0600566103](https://doi.org/10.1073/pnas.0600566103).



# List of Figures

- 1.1 **Monod law, reproduced from [25].** Monod has been a pioneer in the formalization of microbial growth into fundamental, coarse-grained relationships. This figure displays the steady-state growth rate (in number of divisions per hour) of *Escherichia coli* in a synthetic minimal medium at 37°C containing different concentrations of glucose (a common carbon source). As the concentration of glucose increases, so does the growth rate of *E. coli* by following the hyperbolic relationship described in Eq. 1.1. The solid curve is obtained for  $\mu_K = 1.35 \text{ div h}^{-1}$ , and  $K_C = 0.22 \cdot 10^{-4} \text{ mol L}^{-1}$ . . . . . 16
- 1.2 **The growth law of ribosomal abundance (figure reproduced from Fig S1 in [33]).** For a variety of carbon sources and their corresponding steady-state growth rate, the RNA/protein ratio is linearly correlated with the growth rate, a relation that holds for many species of microorganisms. This ratio is correlated with the fraction of ribosome-affiliated proteins, and therefore with the relative abundance of ribosomes in the cell. . . . . 17

- 1.3 **The different phases of a typical growth curve.** During a typical batch growth scenario, biomass accumulates (black thick line) whereas nutrients are consumed until depletion (green dashed line) [3]. (A) The lag phase is a variable period of time during which the organism adapt to the new medium [38]. It is hard to study experimentaly and is known to be affected by the pre-culturing history of the strain [39–41], the magnitude and the rate of the change between the past and present environments [42], and other hard-to-control environmental conditions [43]. (B) The steady-state (or balanced-growth) phase is characterized by an exponential production of biomass. Its characteristics are time-invariant and quite robust accross conditions, which has made it a standard for microbial growth studies [3]. This phase can be extended for hundreds of generations in continuous cultures by the constant renewal of the medium [36, 44, 45]. As represented here, nutrients are quickly depleted in batch conditions, which does not allow to maintain steady-state growth for a long period of time. (C) The stationary phase occurs after the depletion of the limiting nutrients in the medium [3, 46]. In natural conditions, microorganisms usually encounter poor media and spend most of their time in stationary phase [4–6]. Some species that have evolved long-term resistance mechanisms, like sporulation [47] or cannibalism [48], can survive particularly long stationary phases. . . . . 19
- 2.1 **Self-replicator model of bacterial growth.** External substrates  $S$  enter the cell and are transformed into precursors  $P$  through the action of the metabolic machinery  $M$ . The precursors are used by the gene expression machinery  $R$  to make the proteins composing both the metabolic machinery (transporters, enzymes, ...) and the gene expression machinery itself (RNA polymerase, ribosomes, ...).  $\alpha (1 - \alpha)$  is the mass proportion of precursors converted into  $R$  ( $M$ ). Thick arrows denote reactions and thin, dashed arrows denote catalytic activities. The rate of synthesis of precursors and the rate of synthesis of proteins from precursors are denoted by  $V_M$  and  $V_R$ , respectively. . . . . 36

- 2.2 **Analysis of self-replicator model of bacterial growth.**  
*A*: Phase-plane analysis of the self-replicator model of Eqs 2.3 and 2.4. The nullclines for  $p$  and  $r$  are shown as solid and dashed curves, respectively. Parameter values are  $e_M = 3.6 \text{ h}^{-1}$ ,  $k_R = 3.6 \text{ h}^{-1}$ ,  $K_R = 1 \text{ g L}^{-1}$ ,  $\beta = 0.003 \text{ L g}^{-1}$ ,  $\alpha = 0.45$ . *B*: Dependence of the growth rate at steady state  $\mu^*$  on the resource allocation parameter  $\alpha$ , for two different environmental conditions (solid line,  $e_M = 4.76 \text{ h}^{-1}$ ; dashed line,  $e_M = 1.57 \text{ h}^{-1}$ , other parameter values are  $k_R = 2.23 \text{ h}^{-1}$ ,  $K_R = 1 \text{ g L}^{-1}$ , and  $\beta = 0.003 \text{ L g}^{-1}$ ). The maximal growth rate is attained for a unique  $\alpha$ , called  $\alpha_{opt}^*$ . . . . . 39
- 2.3 **Self-replicator model accounts for bacterial growth laws.**  
*A*: Predicted quasi-linear relation between the maximal growth rate  $\mu_{opt}^*$  and the corresponding optimal resource allocation  $\alpha_{opt}^*$ , for different values of  $e_M$  (different colors). The colored dots indicate  $\alpha_{opt}^*$  and  $\mu_{opt}^*$  for  $k_R = 2.23 \text{ h}^{-1}$  and different  $e_M$ , and the dashed black line the relation for all intermediate values of  $e_M$ . The dashed colored lines indicate the relation between  $\alpha_{opt}^*$  and  $\mu_{opt}^*$  obtained when, for a given value of  $e_M$ , the value of  $k_R$  is decreased (lower  $k_R$  leads to lower  $\mu_{opt}^*$ ). The solid grey curves correspond to  $(\mu^*, \alpha)$ -profiles like those shown in Fig. 2.2*B*. *B*: Measured relation between the total RNA/protein mass ratio and the growth rate, in different growth media with different doses of a translation inhibitor (data from [33]). For each medium, indicated by a color, five different concentrations of inhibitor were used (higher dose leads to lower growth rate). Growth-medium compositions are given in the original publication and error bars represent standard deviations. The dashed black and colored lines are the same as in panel *A*, indicating the good quantitative correspondence between model predictions and experimental data for the chosen parameter values, obtained by fitting the model to the data points (see *Methods* for details). . . . . 41

- 2.4 **Optimal control of the self-replicator during a nutrient upshift.** *A:* Optimal trajectory in the phase plane for the nondimensionalized model of Eqs 2.11-2.12, with streamlines. The optimal trajectory is shown as a solid, red curve. The solid, black curve represents the  $\hat{p}$ -nullcline. The dashed, black curve is the switching curve  $\varphi(\hat{p})$ . The optimal solution was obtained by numerical optimization using `bocop` [115] (see *Methods* for details), using the parameter values  $E_M = 1$  and  $K = 0.003$ , and starting from the initial state  $(0.024, 0.18)$  at  $t = 0$  (optimal steady state for  $E_M = 0.2$ ). *B:* Time evolution of the control variable  $\alpha_{opt}(\cdot)$  (thick, red line) and the environment  $E_M$  (dashed, black line). . . . . 44
- 2.5 **Alternative strategies for controlling the self-replicator of bacterial growth.** The feedback control strategies, shown in red and superposed on the self-replicator of Fig. 2.1, exploit information on system variables and the environment to adjust the value of  $\alpha$ , and thus the relative allocation of resources to the metabolic machinery and gene expression machinery. . . . 46

- 2.6 **Comparison of the performance of the nutrient-only and precursor-only strategies after a nutrient upshift.** *A:* Trajectory in the phase plane for the nutrient-only strategy (green curve). The solid, black curve represents the  $\hat{p}$ -nullcline. The dashed, black curve is the  $\hat{r}$ -nullcline. The solution is obtained by numerical simulation of the system of Eqs 2.11-2.12, supplemented with  $\alpha = f(E_M)$  as specified by Eq. 2.27 in the *Methods* section and plotted in S1 Figure. The initial state corresponds to the steady state attained for an environment given by  $0.2 E_M$ . While converging to the new steady state after the upshift, the precursor concentration makes a large overshoot. *B:* As above, but for the precursor-only strategy. The feedback control strategy is now defined by  $\alpha = g(\hat{p})$  as specified by Eq. 2.28 in the *Methods* section and plotted in S1 Figure. The solution trajectory (blue curve) exhibits a lower overshoot. *C:* Evolution of the control variable  $\alpha(\cdot)$  as a function of time, for each of the above two strategies. Notice that in the nutrient-only strategy  $\alpha(\cdot)$  immediately jumps to the optimal value for the post-upshift steady state (green curve), whereas in the precursor-only strategy it depends on the (time-varying) precursor concentration (blue curve). *D:* Evolution of the ratio  $\text{Vol}/\text{Vol}_{opt}$  as a function of time, where  $\text{Vol}$  is the volume of the self-replicator and  $\text{Vol}_{opt}$  the volume of the same replicator following the optimal strategy shown in Fig. 2.4. In all of the above simulations, the parameter values  $E_M = 1$  and  $K = 0.003$  were used. . . . . 48



- 2.7 **Comparison of the performance of the precursor-only and the on-off strategies after a nutrient upshift.** *A*: Trajectory in the phase plane for the on-off strategy (yellow curve). The solid, black curve represents the  $\hat{p}$ -nullcline and the dashed, black curve the function  $g$ . The solution is obtained by numerical simulation of the system of Eqs 2.11-2.12, supplemented with the equation  $\alpha = h(\hat{p}, \hat{r})$  defined in Eq. 2.19 and plotted in Fig. 2.8A. The initial state corresponds to the optimal steady state attained for an environment given by  $0.2 E_M$ . *B*: Trajectory in the phase plane for the precursor-only strategy (same as in Fig. 2.6B, added for comparison). *C*: Evolution of the control variable  $\alpha$  for each strategy as a function of time. Both strategies stabilize the system at the optimal steady state, but only the on-off strategy (yellow curve) exhibits bang-bang behavior. *D*: Evolution of the ratio  $\text{Vol}/\text{Vol}_{opt}$  for the on-off and precursor-only strategies as a function of time, where  $\text{Vol}$  is the volume of the self-replicator and  $\text{Vol}_{opt}$  the volume of the same replicator following the optimal strategy shown in Fig. 2.4. The final values of  $\text{Vol}/\text{Vol}_{opt}$  attained by the two strategies are 0.9831 and 0.9413, respectively. The on-off strategy is thus hardly distinguishable from the optimal control strategy in the plot. In all of the above simulations, the parameter values  $E_M = 1$  and  $K = 0.003$  were used. . . . . 52
- 2.8 **ppGpp regulation implements an on-off control strategy of resource allocation.** *A*: Response surface of the on-off control strategy, defined by  $\alpha = h(\hat{p}, \hat{r})$  in Eq. 2.19. *B*: Response surface of the ppGpp control strategy, as defined by Eq. 2.20 and the simplified kinetic model defining ppGpp in terms of the total amino acid concentration and the ribosomal protein fraction (S4 Text). The shape of the response surface of the ppGpp control strategy is seen to be in very good agreement with the on-off strategy leading to near-optimal performance of the self-replicator during a nutrient upshift. . . . . 54
- 2.9 **Local stability of the on-off strategy.** The on-off strategy sets  $\alpha$  to a value of 0 (1) when  $\hat{r} > g(\hat{p})$  ( $\hat{r} < g(\hat{p})$ ). The solid, black curve is the  $\hat{p}$ -nullcline. The dashed, black curve is the curve  $\hat{r} = g(\hat{p})$ . The arrows represent the vector fields for  $\alpha = 0$  (in blue) and  $\alpha = 1$  (in red). The intersection of the  $\hat{p}$ -nullcline and the curve  $\hat{r} = g(\hat{p})$  corresponds to a unique non-trivial stable steady state, which is equal to  $(\hat{p}_{opt}^*, \hat{r}_{opt}^*)$  by Eq. 2.60. . . . . 70

- 2.10 **Transitions between regions in the phase-plane for the adjoint system.** A switch occurs when a trajectory crosses the  $\lambda_p$ -axis. Left: abnormal case. An extremal trajectory cannot have more than two switches. Right: normal case.  $(\lambda_p^s, 0)$  corresponds to the singular arc. After the first switch, an extremal trajectory cannot have two consecutive switches if it stays in the region  $\{(\hat{p}, \hat{r}) \in \Omega \mid \hat{p} < \hat{p}_{opt}^*\}$  or  $\{(\hat{p}, \hat{r}) \in \Omega \mid \hat{p} > \hat{p}_{opt}^*\}$ . . . . . 77
- 2.11 **ppGpp concentration is a function of total ribosome and amino acid concentrations.** We assume the dynamics of ppGpp to be fast on the time-scale of changes in the ribosome and amino acid concentrations. The concentration of ppGpp can thus be expressed as a function of the latter two variables, using the model of Bosdriesz *et al.* [116]. Parameters are taken from Table 2.1. . . . . 82
- 2.12 **Simple control strategies for the self-replicator of bacterial growth.** *A:* Nutrient-only strategy:  $\alpha = f(E_M)$ . The dashed, black curve is the (unique) strategy driving the system exactly to the optimal steady state, that is, the state in which growth occurs at the maximal rate supported by  $E_M$ . The function  $f$  is defined by Eq. 2.27 in the *Methods* section of the main text. The solid, red curve is an approximation of this function by the simple Michaelis-Menten curve of Eq. 2.16, with  $K_{mE} = 1.0$ . *B:* Precursor-only strategy:  $\alpha = g(\hat{p})$ . The dashed, black curve is the (unique) strategy driving the system exactly to the optimal steady state. The function  $g$  is defined by Eq. 2.28 in the *Methods* section of the main text. The solid, red curve is an approximation of this function by the simple Hill curve of Eq. 2.18, with  $K_{mp} = 0.06$  and a cooperativity coefficient 2. . . . . 85

- 3.1 **Schematic outline of the upshift experiment.** The goal is to measure the fluorescence and length/area of *E. coli rpsB-gfp* cells during an acetate-to-glucose upshift. We use M9 minimal medium supplemented with 0.2% of acetate or glucose (see Material and Methods 3.4.2). A preculture was started from glycerol stock for 2.5 days on 0.2% acetate in batch condition (shake flask). The day of the experiment, the cells were injected into the mothermachine and fed by a constant flow of fresh 0.2% acetate medium (see Material and Methods 3.4.3). Fluorescence and phase contrast images were taken every 5 minutes. After 20 h, the feeding media was switched to 0.2% glucose and maintained for 20 h while continuing image acquisition. Time 0 corresponds to the moment of the acetate-glucose upshift. . . . . 94
- 3.2 **Results of data acquisition.** We imaged 6 fields (X[1-3]Y[1-2]) each containing 15 wells (W[0-14]) for ~40 h. From the 90 wells, 68 were suitable for further analysis (the others being empty, out of frame for some period of time, or plugged). For all wells, data points are missing in the interval between [-720,-150] because the camera was out of focus. We stopped the analysis at 550 min, when about 3/4 of the bacteria were still growing, before the entire population died within a few hours for an unknown reason. The image labeled "Raw image" is the last image analyzed for the highlighted well. The bacterium on the left of this image was manually segmented by selecting two pixels at the poles on the fluorescence images (red cross). A 6-pixel-wide rectangular mask was computed for each image, resulting in the "Segmented image" on the right. Fluorescence intensities are expressed in Relative Fluorescence Units (RFU) on a 16-bit image and were corrected for camera background, but not autofluorescence background (Material and Methods 3.4.4 and Discussion 3.3). The fluorescence intensity of the cell, expressed in units RFU/pixel/cell, was computed by dividing the sum of the fluorescence intensities of the pixels in the rectangle by the total number of pixels in the rectangle. The cell length is the distance in pixels between the two poles (red line). The thick lines in green and black highlight the time-varying length and fluorescence density for the cell visible in the top images, labeled "X3Y2, W2". The vertical dashed lines represent the time of the upshift from growth on acetate to growth on glucose. . . . . 96

- 3.3 Growth-rate estimation using Kalman smoothing based on measurement of the length of bacteria growing in a microfluidic device.** Gray lines represent the estimation of the time-varying length (upper plot) and growth rate (lower plot) of 68 cells by the unscented Kalman smoothing procedure. The solid red lines highlight the result for one particular cell, located at the bottom of the well labelled "X3Y2, W2", while black crosses in the top graph are the data points for this cell. The vertical dashed lines represent the time of the upshift from growth on acetate to growth on glucose. As prior for the algorithm, we used an observation variance of 9 pixels<sup>2</sup> for the length  $L$ . The transition variance  $\theta^2$  (*i.e.*, the smoothing factor for  $\mu$ ) is fixed at  $10^{-8} \text{ min}^{-6}$ . Inheritance between mother and daughter cells is taken into account by systematically choosing an initial mean growth rate equal to the last estimated value for  $\mu$  before cell division, and to half the last estimated value for  $V_\gamma$ , bearing in mind that *E. coli* cells divide symmetrically. At the start of the experiment, when no mother cell is available to provide initial estimates, the above values were fixed at 15 pixels for  $V_\gamma$  and  $0.004 \text{ min}^{-1}$  for  $\mu$ . The variances associated with these means are 16 pixels<sup>2</sup> and  $10^{-4} \text{ min}^{-2}$ , respectively for  $V_\gamma$  and  $\mu$ . The initial mean of  $v$  is set equal to 0, with an initial variance of  $10^{-8} \text{ min}^{-6}$ . All the cross-covariances are set to 0 because the system variables are independent by construction. . . . . 102
- 3.4 Growth-rate estimation using Kalman smoothing for the normal cells only.** The top graph is the same as the bottom graph of Fig. 3.3, except that the dying and pausing cells were removed. The bottom graph shows the 25% (lower gray curve), 50% (solid black curve) and 75% (upper gray curve) quartiles, computed at each time step. The gray area represents the interquartile range. . . . . 103

- 3.5 Performance of the Kalman smoothing procedure on synthetic data simulating an acetate-glucose upshift.** (A) Synthetic data simulating an upshift from acetate to glucose, with and without additive white noise, as well as the results obtained by Kalman smoothing. The synthetic data were generated by simulating the model presented in Eqs 2.3-2.4 with the on-off regulatory strategy (Eq. 2.19 and Fig. 2.7). The model parameters used for the simulation are  $e_{M,Ac} = 0.18 \text{ h}^{-1}$ ,  $e_{M,Glu} = 0.9 \text{ h}^{-1}$ ,  $k_R = 3.6 \text{ h}^{-1}$ ,  $\beta = 0.003 \text{ L g}^{-1}$ ,  $K_R = 1 \text{ g L}^{-1}$ . The predicted  $r(t)$  profile was multiplied by a factor  $\gamma = 0.02 \text{ RFU L g}^{-1}$  in order to obtain the corresponding fluorescence intensity profile  $F$  (dashed black curve). The noise level was estimated from the data (S7 Text) and added to  $F$ . The choice of the parameters of the Kalman smoothing procedure is discussed in the Material and Methods 3.4.5. (B) Estimation of the resource allocation profile  $\gamma\alpha/\beta$  based on the data in (A). Following Chapter 2,  $\alpha(t)$  displays a bang-bang-singular profile during the upshift (dashed black curve). While the Kalman smoother is not able to capture the discontinuous variations in  $\gamma\alpha/\beta$ , it qualitatively reproduces the input quite well (red solid curve). (C) The predicted growth rate during the upshift experiment. This information is used as an input in the smoothing procedure, since it is supposed to have been independently estimated from the measurements  $\{L(t_0), \dots, L(t_{N-1})\}$ . . . . . 105
- 3.6 Estimation of the resource allocation profile using Kalman smoothing based on the fluorescence density measurements and the estimated growth rates from Fig. 3.3.** (A-B) Gray lines represent the estimation of the fluorescence density (RFU/pixel)  $F(\cdot)$  (in A) and the resource allocation profile  $\gamma\alpha/\beta$  (in B) by the Kalman smoothing procedure for the 45 normal cells. The solid red curves highlight the result for one particular cell, , located at the bottom of the well labelled "X3Y2, W2", while black crosses in the top graph are the data points for this cell. The vertical dashed lines represent the time of the upshift from growth on acetate to growth on glucose. The prior values for the parameters of the smoothing algorithm are exactly the same as those used for Fig. 3.5 and are reported in the Material and Methods 3.4.5. . . . . 106

- 3.7 **Robust statistics for the estimation results presented in Fig. 3.6.** Each graph shows the 25% (lower gray curve), 50% (solid black curve) and 75% (upper gray curve) quartiles, computed at each time step for the signals reconstructed in Fig. 3.6. The gray area represents the interquartile range. Interestingly, while most oscillations in the resource allocation profile  $\gamma\alpha(\cdot)/\beta$  cancel out at the population level, the first peak after the upshift is conserved. . . . . 108
- 3.8 **Global overview of all estimated resource allocation profiles presented in Fig. 3.6.** The individual  $\gamma\alpha(\cdot)/\beta$  curves have been normalized with respect to their maximum in the interval  $[0, 400]$  and then ordered with respect to their maximum in the interval  $[0, 120]$ . . . . . 109
- 3.9 **Construction of the gfp-tagged ribosome subunit.** The translational fusion of *gfp\_mut2* with *rpsB* was constructed on the chromosome of *E. coli* (top line) by homologous recombination of *gfp\_mut2* followed by a selection “cassette” (second line). A small linker was inserted between *rpsB* and *gfp\_mut2* in order to minimize interference of the fluorescent protein with ribosome functioning. The selection cassette consists of a positive selection marker, the gene coding for the resistance to the antibiotic kanamycine, and a negative selection marker, the gene coding for the toxin CcdB. The latter is transcribed from the *p<sub>BAD</sub>* promoter, which is only activated in the presence of arabinose in the culture medium. Homologous recombination is indicated by the grey shaded lines. The resulting recombination product (third line) contains the desired fusion protein followed by the selection cassette on the chromosome of the bacterium. A second homologous recombination, using an oligonucleotide complementary to the end of *gfp\_mut2* and the beginning the original region downstream of *rpsB* removes the selection cassette. The resulting strain (line four) carries the translational fusion of *gfp\_mut2* to *rpsB* without any other modification of the chromosome. . . . . 114

- 3.10 **Growth curves in M9 0.2% glucose for the *rpsB-gfp* and *rpsB-mCherry* strains.** Growth and measurements were performed at 37°C on a 96-well microplate in a Tecan infinite 200 pro. We monitored 4 wells for each of the WT, *rpsB-gfp*, and *rpsB-mCherry* strains. Curves were time shifted to correct for variability in the inoculation process (-50 min for WT, and -20 min for *rpsB-mCherry*). (A) Absorbance at 600 nm, corrected for background by subtracting the absorbance of the M9 medium. Curve superposition indicates similar growth rates between the WT and the modified strains. (B) Fluorescence measured in each well in Relative Fluorescence Units (RFU). Green fluorescence (485 nm excitation, 535 nm emission) is measured for the WT (dashed lines) and the *rpsB-gfp* (solid lines) strains. Red fluorescence (560 nm exc., 635 nm em.) is measured for the WT (dashed lines) and the *rpsB-mCherry* (solid lines) strains. Fluorescence levels of the modified strains are far above the autofluorescences measured on the WT strain. (C) Ratio of fluorescence over absorbance (proxy for the fluorescence concentration in the cells). Autofluorescence background was corrected by removing the fluorescence measured on the WT. We observed a strange increase in fluorescence concentration for the *rpsB-mCherry* strain after entry into stationary phase, something that does not occur on the *rpsB-gfp* strain. At each measurement cycle, the following procedure was applied: shaking (Orbital 6mm) for 30 s, shaking (Linear 6mm) for 30 s, waiting for 5 s. . . . . 141
- 3.11 **Estimation of the maturation and degradation rates of the reporter proteins in the *rpsB-gfp* and *rpsB-mCherry* strains.** The strains were grown in the same condition as in Fig. 3.10, except that a high concentration of Chloramphenicol was added in mid-exponential phase ( $t = 0$ , vertical dashed grey line). The fluorescence was normalized by the value at  $t = 0$  and the degradation and maturation rates were estimated by using the model in Eq. 3.22. Parameters of the best fit (solid black line) are reported in Tab. 3.1. Data are the result of the aggregation of 9 and 8 independent growth curves for *rpsB-gfp* and *rpsB-mCherry*, respectively. . . . . 142

**3.12 Points of interest for the noise estimation.**  
*(Top graph)* RFU/pixel/cell data points for the cell at the bottom of the well labeled X3Y2,W2. Noise estimation was performed on the points just before the upshift (red points), where the bacteria are assumed to grow at steady state. *(Bottom graph)* Cumulative density function (CDF, in red) and probability density function (PDF, in gray) of the points highlighted in red on the top graph. They are visually compared with CDF (solid green line) and PDF (dashed green line) of a Gaussian fit with mean 548.88 and standard deviation 17.316. . . . . 145

**3.13 Distribution of the means and standard deviations in the region of interest for the normal cells.** The empirical mean and standard deviation were evaluated in the region before the upshift (red points on Fig. 3.12) for each of the 45 normal cells. They appear to be slightly correlated (Pearson  $R^2$ : 0.3213, p-value:  $4.923 \cdot 10^{-5}$ ) which could indicate a multiplicative instead of an additive noise. . . . . 146

**3.14 Distribution of the noise residues after normalization and aggregation for the 45 normal cells.** Two noise models were considered for the normalization: either the temporal mean  $m_k(\cdot)$  was removed from the measurements independently for each cell (additive noise model in Eq. 3.25), or 1 was removed from the ratio of the measurements with the temporal mean  $m_k(\cdot)$  (multiplicative noise model in Eq. 3.27). For the additive noise model, the fitted Gaussian distribution has a mean of  $1.256 \cdot 10^{-15}$  and a standard deviation of 28.29. For the multiplicative noise model, the fitted Gaussian distribution has a mean of  $2.666 \cdot 10^{-18}$  and a standard deviation of 0.05383. . . . . 148

**3.15 Complete analysis for the cell located at the bottom of the well labelled "X1Y1, W11".** (Figure description available in the introduction of S8 Text.)  
 Growth rate and resource allocation are particularly unstable at the beginning of the experiment, but seem to stabilize before the upshift. Oscillations in the RFU/pixel/cell signal are clearly visible and result in oscillations in the resource allocation signal reconstruction. . . . . 150



- 3.16 **Complete analysis for the cell located at the bottom of the well labelled "X2Y1, W7".** (Figure description available in the introduction of [S8 Text.](#))  
 Here, the resource allocation is unstable at the beginning of the experiment despite the apparent regularity of the RFU/pixel/cell. This could indicate a smoothing factor that is too low for this particular cell. . . . . 151
- 3.17 **Complete analysis for the cell located at the bottom of the well labelled "X3Y2, W5".** (Figure description available in the introduction of [S8 Text.](#))  
 We see for this cell that the reconstruction of the growth rate on the acetate medium (before 0) is affected by the huge gap in the acquisition, the increase before -150 min clearly being an artifact. However, because the Kalman smoothing procedure we used allows for flexibility between the mother and the daughter cells, we quickly recover a more realistic growth rate before the upshift. Despite an instability at the beginning of the experiment, the resource allocation reconstruction exhibits a remarkable stability during growth on the acetate medium, followed by oscillations after the upshift on glucose (also visible directly on the RFU/pixel/cell data). . . . 152
- 3.18 **Complete analysis for the cell located at the bottom of the well labelled "X2Y2, W3".** (Figure description available in the introduction of [S8 Text.](#))  
 The reconstruction on this cell is a good example of what was expected: the growth rate is stable on acetate, than quickly increases after the upshift on glucose. The resource allocation is stable on acetate (indicating a steady state) than starts to oscillate after the upshift on glucose, even though the smoothing factor seems a little too high for this particular cell (some oscillations are poorly predicted after the upshift, in particular in the interval [200,400] min). 153
- 3.19 **Complete analysis for the cell located at the bottom of the well labelled "X1Y2, W6".** (Figure description available in the introduction of [S8 Text.](#))  
 The results on this cell are particularly questionable: we do not seem to obtain a good steady state on acetate since the reconstructed resource allocation appears either negative or oscillating before the upshift. It is thus impossible to conclude if the oscillations observed after the upshift are real or due to overfitting. This cell highlights why robust steady states before and after the transition are important for a reliable signal reconstruction. . . . 154

3.20	<b>Cell categories identified in the microscopy analysis.</b> As stated in Section 3.2.4, the reconstruction of the growth rate for the 68 available cells motivated their classification in three categories: dying cells stop growing (lowest graphs) , pausing cells present growth rate oscillations (middle graph), and normal cells that do not present any of the above (highest graph). . . . .	155
3.21	<b>Robust statistics for the cell categories identified in the microscopy analysis.</b> Each graph shows the 25% (lower gray curve), 50% (solid black curve) and 75% (upper gray curve) quartiles, computed at each time step for the growth rates presented in S2 Fig. The gray area represents the interquartile range. . . . .	156

## List of Tables

2.1	<b>Parameters and variables reused from Bosdriesz <i>et al.</i> [116].</b> The abbreviation aa denotes amino acids. . . . .	83
2.2	<b>Parameter values of self-replicator model</b> The parameter values in the model were obtained by fitting Eq. 2.25 to the data of Scott et al [33] (Fig. 2.3 in the main text), as described in the <i>Methods</i> section. They are compared with order-of-magnitude estimates from the literature (S2 Text). . . . .	84
3.1	<b>Fitted parameters for the degradation and maturation of GFP mut2 and mCherry, according to the model in Eq. 3.22 and data in Fig. 3.11</b> . . . . .	143