Quantification of bacterial resource allocation in changing environments on the single-cell level

Antrea Pavlou

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Virologie - Microbiologie - Immunologie

Arrêté ministériel : 25 mai 2016

Présentée par

Antrea PAVLOU

Thèse dirigée par **Hans GEISELMANN**
et codirigée par **Hidde DE JONG**, Inria

préparée au sein du **Laboratoire Interdisciplinaire de Physique** et de l'**INRIA**
dans **l'École Doctorale Chimie et Sciences du Vivant**

Quantification de l'allocation des ressources bactériennes lors des changements d'environnement au niveau des cellules individuelles.

Quantification of bacterial resource allocation in changing environments on the single-cell level.

Thèse soutenue publiquement le **12 juillet 2022**,
devant le jury composé de :

Monsieur Wolfram LIEBERMEISTER
DIRECTEUR DE RECHERCHE, INRAE, Rapporteur
Monsieur Pascal HERSEN
DIRECTEUR DE RECHERCHE, Institut Curie, Rapporteur
Madame Rosalind ALLEN
PROFESSEUR, University of Jena, Examinatrice
Monsieur Franz BRUCKERT
PROFESSEUR DES UNIVERSITES, Grenoble INP - UGA, Président
Monsieur Hidde DE JONG
DIRECTEUR DE RECHERCHE, INRIA, Directeur de thèse
Monsieur Hans GEISELMANN
PROFESSEUR, UGA - LIPhy, Directeur de thèse
Monsieur Eugenio CINQUEMANI
CHARGE DE RECHERCHE, INRIA, Invité

Acknowledgments

As I reflect back to write these lines, I realize how many people were there for me throughout my thesis. Here, I wish to take the opportunity to thank those who have been instrumental in making this work possible.

First, I would like to thank my thesis committee, and especially the two reviewers of this manuscript, Wolfram Liebermeister and Pascal Hersen, that took the time to assess my work. I would like to thank my two supervisors, Hans Geiselmann and Hidde de Jong, for guiding me, teaching me and believing in me through the years. I will always remember the long and extremely interesting discussions we had about how the world works. I would also like to express my gratitude to Eugenio Cinquemani whom has been kind of a third supervisor to me and taught me so many great things.

I would like to express my gratitude to Célia Boyat, Corinne Pinel, Mathilde Gateau and Noel Scaramozzino for their priceless help on the experimental part. More generally, I am grateful to all members of the BIOP and Microcosme teams for all these years of support, with special thanks to Delphine Ropers, Aline Marguet, Erik Geissler, Nassab Reda and Thibault Clavier.

I would also like to thank some LIPhy members that helped me find solutions to problems related to my thesis, especially Lionel Bureau, Irène Ventrillard and the whole MicroTiss team. I will always be grateful for your involvement.

I also want to thank all the PhD students at LIPhy that made me want to go to work everyday and that constantly supported me, especially Luiza, Laila, Maaike and Saranath, whom I consider friends for life.

More generally, I am thankful to all my friends from all around the world. Special thanks to my friends from Cyprus: Andrea, Andrea, Azap, Denise, Koulla, Kristia and Sebastian whom always inspired me and brought out the best in me. I am grateful also to Andria, Kiki and Dorina for always believing in me. Lastly, I appreciate all friends that shared moments with me in the water or underground. It has been a crazy ride.

I would like to thank my family who stood by me at all the times, and especially my mother, who sacrificed so many things so I can get where I am today. Mamma, you helped me become the person I am today and I am forever grateful for that.

Lastly, I would like to thank the person who helped me fight the battles, with patience, support and kindness. Thank you, Adrien. Here's to new adventures.

Abstract

The ability of microbes to colonize the most improbable places can be partly attributed to the efficient coordination between growth and metabolism. Over the last 50 years, the relationship between growth and the environment has been intensely studied, and has lead to general empirical relationships or 'growth laws'. In most studies, however, bacteria are maintained at steady-state growth even though such conditions are rarely found in a natural environment. To investigate bacterial adaptation in changing environments, we have tracked growth and gene expression of single cells of *Escherichia coli* bacteria growing in a microfluidics device in changing environments. We have examined the behavior of key ribosomal and metabolic genes using fluorescent protein tags. Using inference algorithms, along with models accounting for the maturation kinetics of reporters, we were able to derive dynamic resource allocation profiles of each protein of interest from the time-lapse measurements. The experimental results provide a detailed view of resource allocation strategies of individual bacteria in dynamically changing environments. Even though the average behavior of the bacteria precisely matches known growth laws during steady-state, resource allocation deviates from the classical growth laws during growth transitions. Furthermore, we identified a considerable heterogeneity between bacteria that manifests itself by different strategies for adapting to a new environment. Our results reveal new principles of dynamical resource allocation and could be helpful in improving biotechnological processes involving microorganisms.

Extended abstract

The ability of microbes to colonize the most improbable places relies on efficient coordination between growth and metabolism. Over the last 50 years, the relationship between growth and the environment has been well characterized in the form of general empirical relationships or 'growth laws', conserved in different microbial species and conditions. These growth laws describe or predict how the available resources are allocated to different cellular functions. However, our knowledge is limited to scenarios of balanced growth where the

environment remains unchanged, a condition rarely found in natural environments. Moreover, even though genetically identical, the members of a bacterial population differ by random fluctuations, i.e., stochastically, and by their epigenetic state. In order to characterize resource allocation in a precise manner, we need to study the response of individual bacteria. Coarsed-grained models can provide predictions about resource allocation in dynamic conditions but experimental evidence is lacking. The aim of this thesis is to provide insights into resource allocation in changing environments at the single-cell level.

To overcome the experimental challenges associated with this task, we chose to measure the allocation to the production of ribosomes, and two key metabolic proteins representative of important cellular functions, via fusions to fluorescent reporters in *Escherichia coli*. The goal was to quantify the amount of resources allocated dynamically towards the expression of each protein of interest for individual bacteria. We focused on the specific context of nutrient upshifts and downshifts from acetate to glucose, and back. The fluorescence and length of individual bacteria growing in a mother machine were monitored experimentally during growth transitions. The microscope images were segmented and the time-lapse measurements were extracted using a custom-made pipeline. The derived length and fluorescence data were used to reconstruct the single-cell growth rates and resource allocation profiles for each protein, using inference methods developed for this purpose. To be able to provide an accurate quantification of the resource allocation from fluorescence data generated from different reporters, we needed to evaluate whether the physical characteristics of the reporter genes bias the calculation of the derived quantities, such as promoter activities. For this, we constructed protein specific models that take into account the maturation kinetics of each reporter. We tested the influence of maturation on fluorescence data in a specific experimental framework at the population level in *E. coli*. Using the same inference methods mentioned above, we reconstructed the underlying promoter activities from fluorescence data. We concluded that maturation can introduce a significant bias in the analysis of fluorescence data, in particular during dynamical growth transitions. The derived protein-specific models are able to correct for the maturation bias in the reconstruction of promoter activities.

Consequently, the maturation models were used within the inference algorithms for the reconstruction of single-cell resource allocation profiles from the time-lapse fluorescence data. The experimental and technical work behind this project gave us access to the concentrations of ribosomes and other key proteins for individual bacteria in a dynamical context. We were able to characterize the relationship between ribosomes and growth rate during upshifts and downshifts. The resulting resource allocation profiles confirm previous population and steady-state measurements, but also demonstrate that the average resource allocation to ribosomes of a bacterial population does not follow the classic growth laws. The adaptation pattern is complex and different for nutrient upshifts or downshifts. Furthermore, the behavior of individual bacteria reveals a heterogeneous response, with some cell adaptation responses that are closer and some that are further from the expected static growth law. The work conducted here provides a deeper understanding of the adaptation principles in changing environments and may be useful for biotechnological purposes, notably for the optimization of processes involving bacteria producing valuable goods.

Non-technical summary of the work

The growth of bacteria is linked to the external conditions by quantitative relationships that have been studied extensively in constant environments and for cell populations. In this thesis, we are interested in how individual bacteria adapt in changing environments. We conduct experiments where we track individual cells growing in changing conditions and extract important measurements over time. We use computational methods to extract the growth rates and the amount of resources allocated towards making some important proteins. We report that the average behavior of the bacteria reproduces the existing relationships in constant environments but is very different when the environment changes abruptly. Additionally, the behavior of individual cells is variable during environment changes, with some bacteria adapting their resources in a more optimal way than others. These results provide new insights on the bacterial adaptation and may be useful for biotechnological purposes.

Résumé

La capacité des microbes à coloniser les endroits les plus improbables peut être en partie attribuée à la coordination efficace entre la croissance et le métabolisme. Au cours des 50 dernières années, la relation entre la croissance et l'environnement a fait l'objet d'études approfondies, qui ont abouti à des relations empiriques générales ("lois de croissance"). Cependant, dans la plupart des études, les bactéries se trouvent dans des conditions de croissance à l'état stationnaire, alors que de telles conditions sont rarement rencontrées dans un environnement naturel. Pour étudier l'adaptation des bactéries dans des environnements changeants, nous avons suivi la croissance et l'expression génétique de cellules uniques de bactéries *Escherichia coli* se développant dans un dispositif microfluidique dans des environnements changeants. Nous avons examiné le comportement de gènes ribosomaux et métaboliques clés en utilisant des fusions avec des protéines fluorescentes. En utilisant des algorithmes d'inférence, ainsi que des modèles tenant compte de la cinétique de maturation des rapporteurs, nous avons pu obtenir des profils d'allocation de ressources dynamiques de chaque protéine d'intérêt à partir des mesures temporelles. Les résultats expérimentaux fournissent une vue détaillée des stratégies d'allocation des ressources des bactéries individuelles dans des environnements changeants. Même si en moyenne, le comportement des bactéries correspond précisément aux lois de croissance connues en régime permanent, l'allocation des ressources s'écarte des lois de croissance classiques pendant les transitions de croissance. De plus, nous avons identifié une hétérogénéité considérable entre les bactéries qui se manifeste par différentes stratégies d'adaptation à un nouvel environnement. Nos résultats révèlent de nouveaux principes d'allocation dynamique des ressources et pourraient être utiles pour améliorer les processus biotechnologiques impliquant des micro-organismes.

Résumé détaillé

La capacité des microbes à coloniser les endroits les plus improbables repose sur une coordination efficace entre la croissance et le métabolisme. Au cours des 50 dernières années, la relation entre la croissance et l'environnement

a été bien caractérisée sous la forme de relations empiriques générales ou "lois de croissance", conservées dans différentes espèces microbiennes et conditions. Ces lois de croissance décrivent ou prédisent comment les ressources disponibles sont allouées aux différentes fonctions cellulaires. Cependant, nos connaissances sont limitées à des scénarios de croissance à l'état stationnaire où l'environnement reste inchangé, une condition rarement rencontrée dans les environnements naturels. De plus, même s'ils sont génétiquement identiques, les membres d'une population bactérienne diffèrent par des fluctuations aléatoires, c'est-à-dire de manière stochastique, et par leur état épigénétique. Afin de caractériser l'allocation des ressources de manière précise, nous devons étudier la réponse des bactéries individuelles. Les modèles dit "coarsed-grained" peuvent fournir des prédictions sur l'allocation des ressources dans des conditions dynamiques, mais les preuves expérimentales manquent. L'objectif de cette thèse est de fournir des informations sur l'allocation des ressources dans des environnements changeants au niveau des cellules individuelles.

Pour surmonter les défis expérimentaux associés à cette tâche, nous avons choisi de mesurer l'allocation à la production des ribosomes, et de deux protéines métaboliques clés représentatives de fonctions cellulaires importantes, via des fusions à des rapporteurs fluorescents chez *Escherichia coli*. L'objectif était de quantifier la quantité de ressources allouées dynamiquement à l'expression de chaque protéine d'intérêt pour chaque bactérie. Nous nous sommes concentrés sur le contexte spécifique des "upshifts" et "downshifts", c'est à-dire pendant les transitions de l'acétate au glucose, et inversement. La fluorescence et la longueur des bactéries individuelles se développant dans un appareillage microfluidique ont été suivies expérimentalement pendant les transitions de croissance. Les images de microscopie ont été segmentées et les mesures temporelles ont été extraites à l'aide d'un pipeline adapté. Les données de longueur et de fluorescence ont été utilisées pour reconstruire les taux de croissance unicellulaires et les profils d'allocation des ressources pour chaque protéine, en utilisant des méthodes d'inférence développées à cette fin.

Pour être en mesure de fournir une quantification précise de l'allocation des ressources à partir des données de fluorescence générées par différents rapporteurs, nous avons dû évaluer si les caractéristiques physiques des gènes rappor-

teurs peuvent introduire des biais dans le calcul des quantités dérivées, telles que les activités des promoteurs. Pour cela, nous avons construit des modèles qui prennent en compte la cinétique de maturation de chaque rapporteur. Nous avons testé l'influence de la maturation sur les données de fluorescence dans un cadre expérimental spécifique au niveau des populations de *E. coli*. En utilisant les mêmes méthodes d'inférence mentionnées ci-dessus, nous avons reconstruit les activités sous-jacentes des promoteurs à partir des données de fluorescence. Nous avons conclu que la maturation peut introduire un biais significatif dans l'analyse des données de fluorescence, en particulier pendant les transitions dynamiques de croissance. Les modèles dérivés spécifiques aux protéines sont capables de corriger le biais de maturation dans la reconstruction des activités des promoteurs.

Par conséquent, les modèles de maturation ont été utilisés dans les algorithmes d'inférence pour la reconstruction des profils d'allocation des ressources des cellules individuelles à partir des données de fluorescence temporelles. Le travail expérimental et technique derrière ce projet nous a donné accès aux concentrations de ribosomes et d'autres protéines clés pour les bactéries individuelles dans un contexte dynamique. Nous avons pu caractériser la relation entre les ribosomes et le taux de croissance dans des conditions externes variables. Les profils d'allocation des ressources qui en résultent confirment les mesures précédentes de la population et du régime permanent, mais démontrent également que l'allocation moyenne des ressources aux ribosomes d'une population bactérienne ne suit pas les lois classiques de la croissance. Le schéma d'adaptation est complexe et différent selon qu'il s'agit d'un "upshift" ou d'un "downshift". De plus, le comportement de chaque bactérie révèle une réponse hétérogène, certaines réponses d'adaptation cellulaire étant plus proches et d'autres plus éloignées de la loi de croissance statique attendue. Les travaux menés ici permettent de mieux comprendre les principes d'adaptation dans des environnements changeants et peuvent être utiles à des fins biotechnologiques, notamment pour l'optimisation des processus impliquant des micro-organismes.

Résumé pour un public non spécialiste

La croissance des bactéries est liée aux conditions extérieures par des relations quantitatives, largement étudiées dans des environnements constants et pour des populations de cellules. Dans cette thèse, nous étudions l'adaptation des bactéries individuelles dans des environnements changeants. Nous menons des expériences dans lesquelles nous suivons des cellules individuelles se développant dans des conditions changeantes. Nous utilisons des méthodes computationnelles pour extraire les taux de croissance et la quantité de ressources allouées à la fabrication de certaines protéines. Le comportement moyen des bactéries reproduit les relations existantes dans des conditions constants mais est très différent lorsque l'environnement change. De plus, le comportement des cellules individuelles est variable lors des changements d'environnement, avec certaines bactéries adaptant leurs ressources de manière plus optimale que d'autres.

Contents

Chapter 1

Introduction

"Science, for me, gives a partial explanation for life. In so far as it goes, it is based on fact, experience and experiment."

Rosalind Franklin

1.1 Context

1.1.1 Microbial adaptation requires economic investments

Microorganisms are essential for life [1]. These unicellular organisms are the reason Earth's biosphere has its current, complex form. They explore the physicochemical boundaries of life and they help us define what life is in case we were to search for it elsewhere [2–5].

This success of colonizing the biosphere can be partly attributed to the fact that microbes are social beings that very often prefer living in communities [6–8]. These communities are formed between different microbial species but also between microbes and other more complex life forms (plants and animals, notably [9–12]). An important example for this cooperation is the symbiosis of multiple bacterial species in the human gut, also known as the human gut microbiome [13]. These communities of bacteria, yeast, and viruses are constantly adapting to the ever-changing gut environment via metabolic regulation and cross-feeding [14, 15], while providing essential services to the hosts [16, 17]. The communication between these organisms is mainly conducted via quorum sensing, [18] but also uses electrical signaling for long distance communication [19]. The ability of microbes to change, exchange and adapt relies on adapting patterns in gene expression.

Bacteria constantly sense their surroundings and adapt gene expression to produce fewer, more, or new proteins, signaling molecules and other important macromolecules [1]. Furthermore, rearranging gene expression is one of the most effective survival strategies in the case of stress [20–22]. One of the most general response strategies to a stress is called stringent response with the signaling molecule ppGpp at its center. One major condition detected by the stringent response is the depletion of the pool of amino acids. This signal triggers the production of ppGpp, which binds to the RNA polymerase and decreases the rate of production of ribosomes [23, 24]. The production, and degradtion, of ppGpp is very rapid and the physiological effects extend directly or indirectly to the expression of hundreds of genes simultaneously.

Optimally rearranging gene expression is a resource allocation problem that all organisms must solve in order to thrive in a given environment. The amplitude and rapidity of each particular molecular response needs to be carefully modulated in order to optimize the amount of resources invested in a task. In order to grow, bacteria need to invest in costly cellular machineries. To be able to uptake a specific nutrient, transporters need to be fabricated, additionally to the enzymes necessary for converting the nutrient into precursors for macromolecules and energy co-factors. The synthesis of proteins requires protein complexes responsible for translation and transcription: ribosomes and RNA polymerases. Considering that the cell has a limited amount of resources, available precursors need to be attributed efficiently to each task and gene expression needs to be optimally rearranged. All these investments can be quite "risky" for the cell, with the possibility of allocating too many resources on fruitless tasks. The question that naturally arises is, what strategies bacteria use to make the most out of the available resources when facing a change in the environment?

1.1.2 Microbial physiology and its growth laws

Gene expression is indeed one of the most important processes in bacterial adaptation and very few methods and tools were available to understand the relation between gene expression, growth and the environment on the molec-

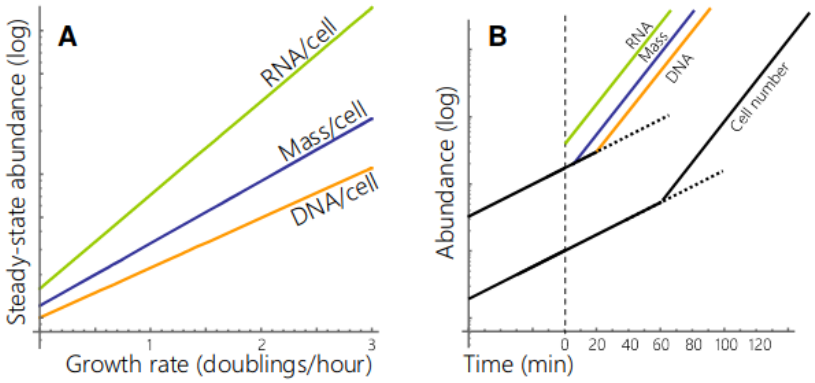


Figure 1.1: Schaechter-Maaløe-Kjeldgaard 1958 experiments, reproduced from [25]. (A) Nutrient growth law during balanced growth of a population. **(B)** Physiological response of a culture to a new medium after a transition from a poorer to a richer carbon source.

ular level, so more phenomenological approaches were used.

Most bacteria are not visible to the naked eye ([26]), therefore the most common way to monitor how fast they grow consisted in measuring the turbidity of cell cultures. By measuring some phenotypical parameters, scientists in the mid-20th century were able to observe some remarkable relationships. Monod was the first to establish a dependency between the growth rate and the concentration of a given nutrient [27]. It is quite fascinating that with one simple equation, the growth dynamics of a population can be predicted, even though this process is controlled by very complex molecular mechanisms.

After this breakthrough, many physiologists started studying bacterial growth and the physiological states of the cell. The two 1958 publications of Maaløe, Kjeldgaard and Schaechter characterized physiology during and after steady-state growth using *Salmonella typhimurium* as model organism (Fig. 1.1). In their first publication, experiments were conducted where the nutrient quality or the temperature were modified while some physiological parameters were monitored during balanced growth [28]. These experiments revealed robust de-

pendencies between nutrient-imposed growth rates and molecular composition of the cell (DNA and RNA content of the cell for example). This relation was later named the "nutrient growth law" since it holds for most bacterial species and for many conditions and does not depend on the chemical composition of the medium.

In their second publication, M, K and S focused on transitions between conditions of balanced growth [29]. By abruptly switching the carbon source contained in the growth medium, the bacteria are forced out of their current state of balanced growth and need to adapt their physiology to the new conditions. Such changes in carbon source are called upshifts or downshifts, depending on whether the quality of the carbon source increases or decreases. In this work, the authors observed that the RNA, mass and DNA contents change to a new rate of accumulation in that order, followed by cell division more than an hour after the upshift. These adaptation dynamics are observed for transitions between different media supporting different growth rates.

Following this experiment, other robust growth laws have been discovered that link balanced growth with molecular composition. Many physiologists contributed to the establishment of a growth law that states the total cell mass and number increase exponentially in balanced growth [25, 30]. Most importantly, Neidhardt and Magasanik [31] discovered a linear relationship between RNA content (and thus ribosomal content, since most of the cellular RNA is rRNA) and balanced growth. This work was pivotal for the discovery of the implication of ribosomes in protein synthesis and has lead to the discovery of mRNA, among other contributions (Jacob, F. in [32]).

The ribosomal growth law was recently revived by Scott *et al.* [33] where the ribosome concentrations were carefully measured for different nutrient imposed growth rates in *E. coli*. These measurements were then extended to other situations (limited translation capacity by adding antibiotics). They also included literature data from other prokaryotic and eukaryotic species, conditions and temperatures (Fig. 1.2) that prove the universality of this relationship.

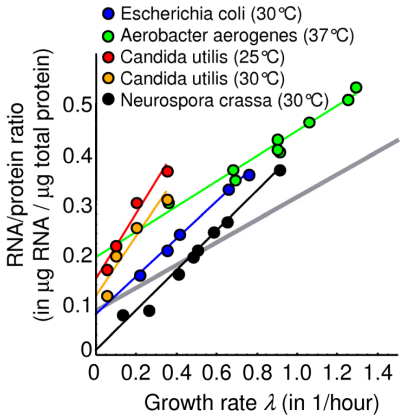


Figure 1.2: Ribosomal growth law, reproduced from [33]. The ribosomal protein fraction is linearly dependent on the nutrient-imposed growth rate. The RNA/protein ratio can be used to quantify the amount or ribosomes since most RNA in the cell is ribosomal. This relationship holds for many species, conditions and temperatures.

1.1.3 Proteome partitioning and resource allocation

The described ribosomal growth law opened new perspectives in the study of resource allocation. The question that naturally arose was: how does the concentration of ribosomes increases with the growth rate, in a cell that has limited space to store proteins and limited protein synthesis capacity? To this problem of space is added the problem of coordination: if ribosome concentration increases during faster growth, how does the cell coordinate the response for the rest of the proteins necessary for growth? The answer, at least in part, lies with the signaling molecule cAMP. This molecule, coordinates the expression of catabolic and ribosomal proteins [34, 35].

The observation that certain proteins are jointly regulated during growth led to the definition of the so-called proteome sectors [36–38]. The idea is that instead of looking at individual proteins, we can group proteins depending on their (known) function and quantify their accumulation at different growth

rates. Proteomics studies revealed proteome sectors that are associated with ribosomes, energy and amino acid metabolism and substrate transport/stress [36]. The constraints imposed on protein composition result in a linear variation of the proteome sectors with the growth rate (growth law of proteome partitioning) [25]. These relations hold for balanced growth but it is unclear whether this tight coordination is maintained during growth in dynamic conditions.

1.1.4 Growth in static and dynamic conditions

The definition of balanced growth enabled the realization of reproducible experiments and led to rapid developments in the field of microbial physiology. Balanced growth is attained in stable environments over long periods of time where a microbial population grows exponentially [1, 39]. Consequently, the notions of balanced [40] and steady-state growth [41] were introduced to complement the term exponential growth. Steady-state growth, is defined as the state in which "the distribution of each intensive random variable does not depend on the time" (by Painter and Marr 1968 [41]). Balanced growth, defined by Campbell [40], is the state in which 'every extensive property of the growing system increases by the same factor over a time interval'. The main difference between the two is that balanced growth does not require the cell number to be increasing proportionally with the other cellular constituents [42]. Therefore, steady-state growth encapsulates both exponential and balanced growth.

Steady-state growth is very convenient for the experimental and theoretical study of microorganisms. However, in reality most microorganisms are rarely found in such conditions [1]. If we go back to the example of the gut microbiome, the environment in the gut is highly variable, with phases where nutrients are flowing in the gut and others where they are scarce [15]. Gut bacteria (like most bacteria) lead a starving-survival lifestyle and have evolved to quickly uptake carbon, nitrogen or phosphorus when they become available [43].

However, not many studies exist with a focus on growth in dynamic conditions, even though the strains most commonly used in laboratories, for example

E. coli, are strains that are naturally found in fluctuating environments [44]. The research conducted over the past 60 years has established simple empirical relationships that are robust for balanced growth in a variety of static environments. The question that arises now is, taking advantage of the technology and methods we have today, can we establish experimentally which relationships hold outside of steady-state growth?

1.1.5 Stochasticity in the bacterial realm

In biology we traditionally considered that a group of individuals produces a population response without considering any deviations individuals might have from the mean. Stochastic factors that might influence individual responses only recently became a topic of study, notably with the contributions of theoretical and experimental investigations from theoretical ecology and evolution theory [45–48]. This is mostly because stochastic factors like genetic mutations, genetic drift and immigration constitute the only way to explain the neutral theory of biodiversity [49].

Stochastic factors also affect microbial populations along with deterministic ones [50]. Within a growing bacterial culture, individual cells are subject to random events that affect their growth and metabolism [51]. Notably, genetic mutations play a key role in bacterial heterogeneity and speciation. They can lead to the gain of new functions that might help individual bacteria survive in a stressful environment and thus increase the fitness of the population [52, 53].

For example, persister and viable but non-culturable cells (VBNC) constitute a small fraction of a bacterial population that is resistant or tolerant to antibiotics [42, 54]. These cells undergo a stochastic phenotypic switch and have a slower growth rate and metabolism prior to antibiotic stress that consequently leads to a decreased uptake of the antibiotic [55–60]. This strategy is beneficial for the survival of bacteria in stressful environments [54, 61].

Isogenic cells growing in a given environment can exhibit fluctuating growth rates, division and generation times [62–65]. There are multiple potential physiological origins for these variations. Gene expression is intrinsically stochastic because it depends on hundreds of random microscopical events and reactions

(intrinsic noise), along with the state, localization and concentration of the molecules that govern it (extrinsic noise) [66, 67]. These fluctuations have been proven to influence growth [51, 68].

The growth laws described above characterize the behavior of bacterial populations, however, in reality, it is unclear whether these laws translate to the single-cell level. Given that gene expression has evolved to be highly stochastic, to which degree do individual bacteria deviate from the population means of physiological parameters and thus from the robust growth laws that have been described?

1.2 Problem statement

Most of our understanding of bacterial resource allocation comes from the simple growth laws relating growth rate with the macromolecular composition of the cell [25]. All of these relations, however, have been described for steady-state growth. Only a few studies focus on observations outside of steady-state growth, and therefore the adaptation of bacteria to dynamic environments is still poorly understood. Consequently, quantification of resource allocation in dynamic environments would provide us with useful information about bacterial adaptation without looking in detail into the molecular mechanisms that govern it.

Moreover, as we have seen, the phenotypical response of bacteria in various situations is heterogeneous, with gene expression being the main generator of stochasticity. However, previous studies of resource allocation have focused on the response of populations and not of individual cells.

The questions that emerge from the above interrogations are the following: **How do bacteria (re)allocate their resources in dynamic environments? How much cell-to-cell variability is observed in these adaptation patterns?** To answer these questions one must take into account the following considerations.

How can we measure resource allocation of individual bacteria in changing environments? Even though omics technologies can provide a detailed quantification of the molecular contents of cells at specific points in time, they are limited by the amount of data points that can be collected. As a consequence, they are of limited interest in dynamic conditions. Moreover, they usually provide population-level measurements. So, in order to quantify resource allocation dynamically and on the single-cell level, we must develop a novel approach relying on a combination of experimental and computational approaches. First, the problem can be simplified by quantifying (changes in) resource allocation by means of proteins that have pivotal roles in cell growth and metabolism. Second, in order to measure protein abundances on the level of individual cells, along with the growth rates, the proteins of interest fused to fluorescent reporters can be tracked in microfluidics devices via microscopy.

Third, in dynamic conditions, protein abundances cannot instruct us about the physiological state of the cell. For this, the amount of resources allocated towards the synthesis of these key proteins should be quantified using estimation methods incorporating specific mathematical models and via measurements of cell growth rate and protein abundances.

As simple as this plan might sound, it comes with a multitude of challenges, both on the experimental and computational level, that need to be resolved in order to increase experimental repeatability and avoid artifacts related to the set-up and methods used. The goal of this thesis was to find solutions to all of these theoretical and experimental challenges.

1.3 Approach

In order to achieve the task of measuring resource allocation in individual bacteria, three key elements are required. Firstly, we need an experimental system that would allow us to track dynamically the abundance of proteins in individual cells. I will describe below the available methods for the quantification of physiological parameters in bacteria. Secondly, we need an approach to estimate from dynamic measurements of protein abundance the underlying resource allocation of specific proteins, notably the ribosomes. To tackle this challenge, we can rely on statistical methods that are able to reconstruct underlying signals from primary data. Thirdly, in order to obtain estimates that are biologically relevant and that correctly quantify resource allocation in the cell, we need a mathematical model, used along the inference methods, to correct for biases introduced by the experimental system and to correctly rely the input and output quantities. These three components, detailed in the following sections, will allow us to successfully estimate resource allocation strategies from dynamical single-cell data.

1.3.1 Quantification of bacterial growth and gene expression

1.3.1.1 Dynamical monitoring of parameters in microbiology

The simplest way to grow bacteria is to introduce them in a liquid medium containing water, a carbon source, salts and trace elements [69]. Liquid cell cultures have been the standard way of studying microorganisms on the population level for a very long time. However, the practices needed to evolve with the search of uniform experimental conditions and conservation of balanced growth for longer periods [25]. This was done initially by serially diluting cultures growing exponentially. This technique, even if effective, can become tedious when searching for precise quantitative results. In order to alleviate these chores, the development of microplate readers and bioreactors have come into the rescue. Microplates introduced the parallelization of data acquisition with plates that usually allow multiple conditions and strains to be tested per experiment. Bioreactors allow the continuous culture of micro-organisms for long periods of time with automatic monitoring of parameters. Bioreactors are currently quite popular because besides measuring, we can also control growth of microorganisms via feedback loops [70–73].

Using these instruments, many parameters can be measured and others inferred [74]. With measures of absorbance or optical density we can infer the population growth rate, doubling rates and generation times [75]. Using molecular biology techniques we can measure specific protein abundances in the cell via fluorescent or luminescent reporters. Lastly, biochemistry tools [76, 77] can help us quantify the total amount of proteins, nucleic acids and other macromolecules, along with concentrations of metabolites excreted by the cells [78]. There are some parameters however that cannot be determined using these techniques, an obvious example being the 'age' of cells.

To determine this type of parameters, one must rely on techniques that monitor individual cells. A well-known method to do so is to monitor single cells growing on an agarose pad by microscopy [79, 80]. This technique not only allows the tracking of cell lineages, and the exploration of single-cell phenotypes and cell-variability, but also facilitates the study of localization of

macromolecules inside cells [76, 81]. The limitation of this technique comes from the fact that the bacteria colonize the pad quite fast and hence the experiments cannot last for more than a dozen generations. To track parameters that require growth for a large number of generations, for example the process of cell aging, setups like the mother machine must be used [82].

The combination of microfluidics, microscopy and high-throughtput imaging and analysis allowed for the conception of sophisticated experimental setups. Mother machines are PDMS (polydimethylsiloxane) devices used to follow individual bacteria for longer periods of time. These devices contain closed-end channels that "trap" the bacteria and thus a "mother" cell dividing at the bottom of the channel can be tracked by microscopy for the entire experiment. The medium is refreshed continuously and so the cells remain in constant environmental conditions throughout the experiment. The PDMS devices can be designed using various architectures, making the technique very versatile for the study of bacteria [83–85] and yeast [86].

1.3.1.2 Time-lapse measurements using fluorescent reporters

The discovery of the green fluorescent protein (GFP) from *Aequorea victoria* has made it possible to measure and visualize numerous phenomena in living cells [87]. Since then, many new variants have been discovered or have been synthetically created that possess a range of physical properties [88, 89]. The new and improved variants of fluorescent proteins are generally very bright, have faster maturation times, and emit light at other wavelengths than the original GFP, making dynamic visualization of up to three macromolecules in the cell possible.

This breakthrough came with new challenges, though. The main problem, which is more general when quantifying a phenomenon that we cannot see with the naked eye, is whether the fluorescence emitted by the fluorescent reporter exactly corresponds to the phenomenon we are trying to measure, or whether there are side-effects distorting the observed signal [77]. For example, in order to quantify the activity of a promoter of interest that regulates the accumulation of a certain protein in the cell, the easiest way to proceed is to

replace the gene of interest with a reporter gene [90] and then assume that the fluorescence emitted by the corresponding reporter protein corresponds to the activity of the promoter. However, by making this assumption we ignore the intermediate process of gene expression, including post-transcriptional and post-translational modifications of the protein of interest along with its folding and localization in the cell.

A more principled way to approach this problem is to introduce the fluorescent protein as a tag either via a transcriptional or a translational fusion on the bacterial chromosome. The former produces one mRNA transcript containing both the sequence of the gene of interest along with the fluorescent tag and the latter produces a fusion protein containing both reporter and protein of interest [81, 91]. Translational fusions can thus provide the most accurate measurements among all the existing methods, although it must be verified that the tag does not interfere with protein function [92].

Furthermore, it is possible that the physicochemical properties of the fluorescent reporters influence the signal we observe [93]. Factors such as maturation and photobleaching have be shown to vary significantly across proteins and experimental conditions [93–95]. It is thus very important that this is taken into account when working with fluorescence data, especially in dynamic conditions since the origin of fluctuations in fluorescence intensity might not always be biological. In this work, we constructed translational fusions of three key proteins in the metabolic and gene expression machineries of the cell with fluorescent reporters. To avoid introducing biases in the analysis associated with the properties of the reporters, we developed models that take into account the characteristics of the reporter constructs such as, maturation and degradation.

1.3.2 Modeling bacterial growth, gene expression and metabolism

Undoubtedly, the introduction of mathematical models into our scientific practice has changed our entire perception of biology [96, 97]. In bacteria, the existence of quantitative conserved relationships between growth and other parameters have made modeling a core part of our practices from early on

[27]. The usefulness of modeling in biology is presented when we can choose an appropriate framework of more or less complexity to study a phenomenon [98, 99]. The versatility of this practice can be seen through the contributions of modeling in the understanding of certain bacteria, notably *E. coli.*

For example, a simple ODE growth model with just two parameters can capture the dynamics of a population of exponentially growing cells, regardless of the thousands underlying mechanisms controlling growth [99, 100]. The models can be more 'fine-grained' according to the interest of each study. Equations can be added to model the metabolism, or other essential processes of the cell like transcription and translation. By taking into account biological knowledge available for the organism of interest, we can simplify some relationships that can help us even derive more fine-grained mathematical models [101]. This approach can also be used to model growth between communities of different bacterial species [102].

Considering the quantity and variaty of data that exists for an organism, the fine-grainedness of the models can be pushed to the point of obtaining whole-cell models. This approach has lead to some models [103–107] that take into account entire libraries of accumulated data and have increased our understanding of the organisms in question.

Modeling all the biochemical reactions in the cell is challenging, both mathematically and computationally. Therefore, approximate models like those used in Flux Balance Analysis (FBA) have been developed by introducing some modeling assumptions [99, 100, 108, 109]. These models have proven useful for the fundamental understanding of cell metabolism and are now also used for biotechnological purposes [110]. Additionally, extensions or modified versions of the method have also been developed, for example, Resource Balance Analysis (RBA) [111].

Lastly, all of the described methods focus on growth on the population level and as we have seen, biology is fundamentally stochastic. In order to unravel the dynamics of certain stochastic phenomena and describe the heterogeneous behavior different cells often exhibit, stochastic modeling can be used instead [112]. These models come with a price: they are generally more computationally expensive and more difficult to fit to experimental data than deterministic

ones [113]. However, they are necessary in order to understand and quantify noise in biological systems. In some cases, a stochastic component can be added in deterministic models in order to capture the dynamics of stochastic processes, for example in gene expression [114].

1.3.2.1 Coarsed-grained resource allocation models

Several coarse-grained models exist that link growth rate and macromolecular composition in bacteria [115–117]. A good example is the re-examination of the ribosomal growth law by Scott *et al.* [33], via a phenomenological study that divided the proteome in three parts or sectors, the first comprising the ribosomes; the second, the metabolism and the third the rest of the cellular components which represent a constant fraction of the total proteome independent of growth rate. This simple model was able to reconstruct the growth law of ribosome synthesis without taking into account any molecular details that can vary between species and conditions but rather relies on the core principles imposed by natural selection.

Using the same principles of proteome partitioning, resource allocation models were developed that can predict responses in dynamic environments. Erickson *et al.* [118] developed a quantitative model of proteome allocation independent on kinetic parameters, which was calibrated on proteomics data and relied on qualitative molecular assumptions. With this model, the resource allocation regimes of different protein categories were predicted during a nutrient upshift and downshift.

Lastly, along the same lines, Giordano *et al.* [119] developed a dynamic self-replicator model that is based on the principles of the phenomenological model of Scott *et al.* In this case too, the proteome dependent on growth is partitioned into two sectors corresponding to ribosomes and metabolism, as before. The amount of precursors allocated to one of the two tasks is quantified by the resource allocation profile $\alpha(t)$ (Fig. 1.3, left). For a nutrient upshift or downshift, growth is considered as an optimization criterion and methods from the field of optimal control theory are used to derive the possible resource allocation strategies adopted by bacteria. Using growth rate as an

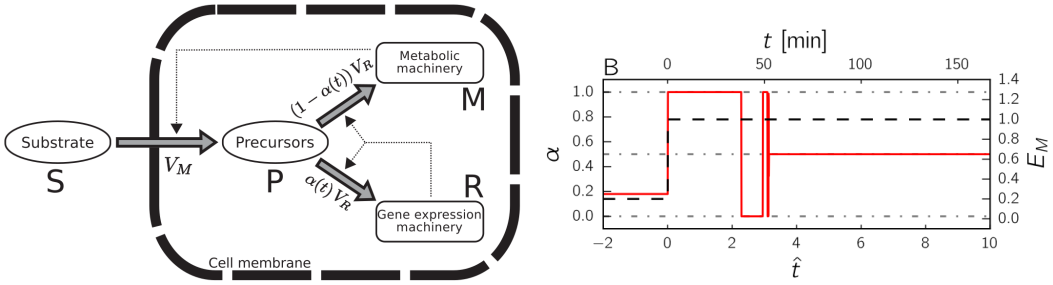


Figure 1.3: The self-replicator model of dynamic resource allocation, reproduced from [119]. In this simple model, the cell is considered as a self-replicating entity where precursors need to be allocated to the gene expression machinery (R) and the metabolic machinery (M). The amount of resources allocated to each part are controlled by the resource allocation profile $\alpha(t)$, ranging from 0 to 1. Using the maximization of biomass accumulation as an optimization criteron, the predicted optimal resource allocation profile for $\alpha(t)$ for a nutrient upshift (poor to rich carbon source) follows an on-off strategy.

optimality criterion, the predicted response follows an on-off strategy, where, during a nutrient upshift, all resources are iteratively allocated to the synthesis of the gene expression machinery and the metabolic enzymes needed in the new environment, before reaching the steady-state resource allocation strategy (Fig. 1.3, right)

These theoretical predictions raise the question whether bacteria actually adopt this kind of regime when they adapt to a new environment and have prompted the investigation of this subject experimentally. In our work, we use the self-replicator model to define resource allocation in the context of the ribosomes and other key proteins, and to relate this quantities with the obtained fluorescence measurements.

1.3.3 Reconstruction of biological quantities from primary data

The development of technologies that allow the quantification of biological phenomena has lead to an increased need for methods that allow the processing

of the obtained signals. In a dynamical context, the quantification of specific quantities in cells, for example promoter activities and protein concentrations, rely on the use of reporter gene systems, as I explained in a previous section. The quantities of interest can be estimated from the fluorescence measurements using statistical inference methods [120–122].

The main problem with estimating a quantity from observable data, is that, as with most experimental systems, the data are noisy. The challenge of these methods is thus to distinguish which part of the observed signal is due to noise (including all different types of noise) and which is due to the biological reactions taking place. In addition, since the observed signal is usually produced via the use of reporters, the processes determining the accumulation and activity of the reporters, including maturation, protein degradation, transcription and translation dynamics (depending on what the reporter actually quantifies) need to be taken into account [77, 93]. Consequently, models taking into account all these factors and relating the quantities need to be used for the correct interpretation of the data. The parameters of these models can be either known (calculated experimentally) or estimated along with the reconstructed signal. The resulting inference problems are ill-posed, in the sense that they do not have a unique solution [123, 124]. To tackle such problems, some assumptions or constraints need to be added and exploited by the statistical inference procedure. Various approaches can be used for this.

One of the most commonly-used approaches, is regularization. The problem can be formulated in the context of regularized inversion where small and large fluctuations are penalized through a regularization parameter presented in the cost function to be minimized [123–125]. This parameter is usually automatically tuned on a given dataset using Cross-Validation methods [126]. This approach has been used for treating linear inverse problems in biology, especially in the context of reconstructing promoter activities from fluorescence measurements [77, 127–131]. For our purpose, a method based on regularization was developed for estimating growth rates of individual cells.

For more complex linear problems, other statistical methods can be used to reconstruct the signal more accurately. Bayesian methods like Kalman smoothing also rely on regularization, but a series of measurements along with

other known characteristics like the noise profiles can be used to estimate a joint probability distribution of all the variables at each time-point and so can provide a more accurate reconstruction of the desired signal [132]. For the purpose of our study, we implemented a Kalman smoothing algorithm to reconstruct resource allocation profiles from fluorescence data. This method exploits the fact that we have datasets with many time-points and that we have access to the noise properties of the data.

1.4 Organization of the thesis

The goal of this thesis is to quantify resources allocation experimentally in the context of a nutrient upshift or downshift and for individual *E. coli* cells. We are mainly interested in the adaptation of the ribosomes, and of two key proteins of the energy and amino-acid metabolism to the new growth regime. To quantify these proteins dynamically, we created translational fusions with various fluorescent reporters. The observed fluorescence was then used to reconstruct the resource allocation profiles of each protein of interest using inference methods.

In order to conduct such experiments in a robust manner, we made sure that the fluorescence signal we observe is not influenced by the physicochemical properties of the reporters. The work of chapter 2 consists in developing and calibrating mathematical models that are specific to each fluorescent reporter. We took into account maturation kinetics and protein degradation to correct for any effects that might influence the interpretation of the measured fluorescence signals. As a proof of concept, we developed an experimental system where two fluorescent proteins, respectively with emission spectra in the green and red, with very different maturation dynamics, are under the influence of the same promoter. To reconstruct the dynamics of the promoter from dynamical fluorescence measurements in batch, we used a Kalman smoothing method [133, 134] along with the maturation models we defined above. This method allowed us to obtain the corrected underlying promoter activity from a fluorescence signal. Lastly, this work helped us pinpoint specific maturation problems of some fluorescent reporters at high optical densities that introduce

huge biases in fluorescence data.

In chapter 3, three strains containing reporter fusions with three proteins of interest were created and validated experimentally using multiple techniques. Numerous microfluidics experiments were then conducted where the strains in steady-state growth are shifted into a medium containing either a poorer or richer carbon source (acetate and glucose). The important physiological parameters were monitored throughout the experiment using time-lapse microscopy.

Much work went to establishing a pipeline to efficiently segment the cells and extract the relevant physiological parameters. To reconstruct the resource allocation profiles we used the models and methods developed in chapter 2 with some variations on the single-cell level. We used a simple self-replicator model [119] to derive a measurement model that links observed fluorescence with a quantity corresponding to the resources allocated to the expression of a certain gene. For estimating the growth rate from length data, we developed a fitting method based on regularization [135].

With the pipeline summarized above we were able to reconstruct growth rates and resource allocation profiles in individual cells for each protein of interest. We were first interested to see how the ribosomal growth law translates to the single-cell and to dynamical contexts. We then correlated the results for the ribosomes with those for two metabolic proteins. Moreover, we were interested in investigating whether the observed responses, both dynamically and in steady-state, are more or less heterogeneous for different cells in a population.

Chapter 2

Maturation models are necessary to obtain unbiased estimates of promoter activity

"Every component of the organism is as much of an organism as every other part"

Barbara McClintock

2.1 Context

In this chapter, I will present our work on the maturation of fluorescent reporters. As I have previously mentioned, the use of various fluorescent reporters for the quantification of resource allocation prompted the investigation on whether the physico-chemical properties of each reporter influence the signal we observe. In steady state, we can neglect these effects. In dynamic conditions, however, factors like maturation and degradation might need to be taken into consideration. Additionally to this, we suspected that some red fluorescent proteins (RFPs) can exhibit strange kinetics in some dynamic conditions. Therefore, a second goal in this project was to see in which conditions RFPs might exhibit unusual dynamics.

For this work, we developed a proof-of-concept experimental system where a red and a green fluorescent protein were put under the influence of the same promoter. The goal was to measure the fluorescence of the reporters in

batch growth experiments, and to reconstruct the underlying promoter activities using statistical methods. This was done in collaboration with Eugenio Cinquemani (INRIA Grenoble-Alpes) where we developed a Kalman smoothing method that we can use with fluorescence data. We also developed and calibrated maturation models that could potentially correct for maturation dynamics.

Through this project, we concluded that maturation can introduce bias in fluorescence data, especially in dynamic conditions. The derived models and inference methods can correct for maturation and can provide promoter activities that are unbiased in a robust manner. Lastly, we did detect strange dynamics for RFPs during the entry in stationary phase, characterized by an artificial dip in the reconstructed promoter activities. We speculated that this is an oxygenation problem at high densities that might cause delays in the maturation. The problem was solved by avoiding reaching high ODs in batch.

The work described here is currently submitted for publication and some minor revisions need to be conducted before publication. The code will be available upon publication.

2.2 Introduction

Since the discovery of GFP in 1992 [87], hundreds of fluorescent proteins (FPs) covering the visible light spectrum have been developed and have become essential for the visualization and quantification of biological phenomena in living cells [76]. FPs notably allow the quantitative investigation of gene expression in a dynamical context [89]. Accordingly, many computational methods have been developed in order to quantitatively reconstruct the promoter dynamics from FP timelapse measurements [77, 130, 131, 136–138].

An observed fluorescence signal is influenced by the distinct physical characteristics of a FP, notably maturation, whose mechanisms and kinetics vary significantly across proteins [93]. The maturation kinetics may decouple FP production and fluorescence emission, thus introducing significant biases in data analysis and interpretation. No method currently exists that incorporates this information in the reconstruction of promoter activities.

In this paper, we use a combination of mathematical modeling, experimental calibration and statistical inference to integrate maturation dynamics in the reconstruction of promoter activities from fluorescence data. We focus on the two most used FPs, green fluorescent protein (GFP) and red fluorescent protein (RFP) which have very different maturation dynamics [93]. We developed appropriate ODE models taking into account what is currently known about the maturation mechanism of each FP and calibrated the models using experimental data. We also developed a Bayesian inference approach to robustly reconstruct promoter activities from timelapse fluorescence data. To validate our approach, we constructed an experimental system in *Escherichia coli* where the two reporters are upstream of the same constitutive promoter. We show that, with correction for maturation, the inference procedure yields the same promoter activities for the two FPs. Without correction, however, the promoter activities differ, thus demonstrating that accounting for maturation is essential for the correct analysis and interpretation of fluorescence data.

Our principled and practically applicable approach can be used for the robust reconstruction of promoter activities in multiple growth conditions and for both prokaryote and eukaryote reporter systems.

2.3 Results and Discussion

2.3.1 Mechanistic maturation models

To correctly assess maturation effects in dynamic conditions, we first define mechanistic models for GFP and RFP, in particular the chosen variants GFPmut2 and mScarlet-I, taking into account maturation mechanisms reported in the literature.

For GFPs in general, an immature colorless species is transformed into a mature green species [94]. Conversely, maturation of mScarlet-I and other RFP variants is complex: a colorless species is transformed into a mature red species via the formation of a blue intermediate absorbing at around 400nm [139, 140]. The different maturation mechanisms of GFP and RFP lead to

distinct maturation kinetics [93], and correspondingly, to different maturation models. Whilst a GFP model has a single maturation step, an RFP model requires two steps in order to account for the blue intermediate (Fig. 2.1A-B).

Little is known about the reversibility of each reaction, so we have constructed and compared several variants of each model that include or exclude backflow reactions (Fig. 2.1A-B, Text S1). Statistical model selection showed that the best model for each protein is the most parsimonious one, without backflows (Materials and methods, Text S1). The models are defined on the population-level: they describe the total quantity of protein species (red, green, blue, colorless) in a growing population of cells. Protein quantities are expressed in relative fluorescence units (RFU), assumed proportional to molar units. Biomass is quantified by absorbance (Abs), assumed proportional to the volume of the bacterial population.

GFP model:

The GFP model is composed of two ordinary differential equations describing the dynamics of the quantity of immature proteins $Im(t)$ [RFU] and the quantity of mature green proteins $M(t)$ [RFU]. The rate of production of immature proteins is given by $\alpha(t) \cdot V(t)$, where $\alpha(t)$ [RFU min^{-1} Abs^{-1}] is the specific production rate, per unit population volume, and $V(t)$ [Abs] the volume of the growing population. The specific production rate consists of the maximum production rate modulated by the promoter activity, where it is assumed that the dynamics of the intermediate mRNA species can be ignored [77]. The conversion of immature to mature protein occurs at a rate proportional to the quantity of immature protein with constant k_m [min^{-1}], and all proteins are degraded at the same rate with rate constant γ [min^{-1}].

$$\frac{d}{dt}Im(t) = \alpha(t) \cdot V(t) - (\gamma + k_m) \cdot Im(t), \tag{2.1}$$

$$\frac{d}{dt}M(t) = k_m \cdot Im(t) - \gamma \cdot M(t). \tag{2.2}$$

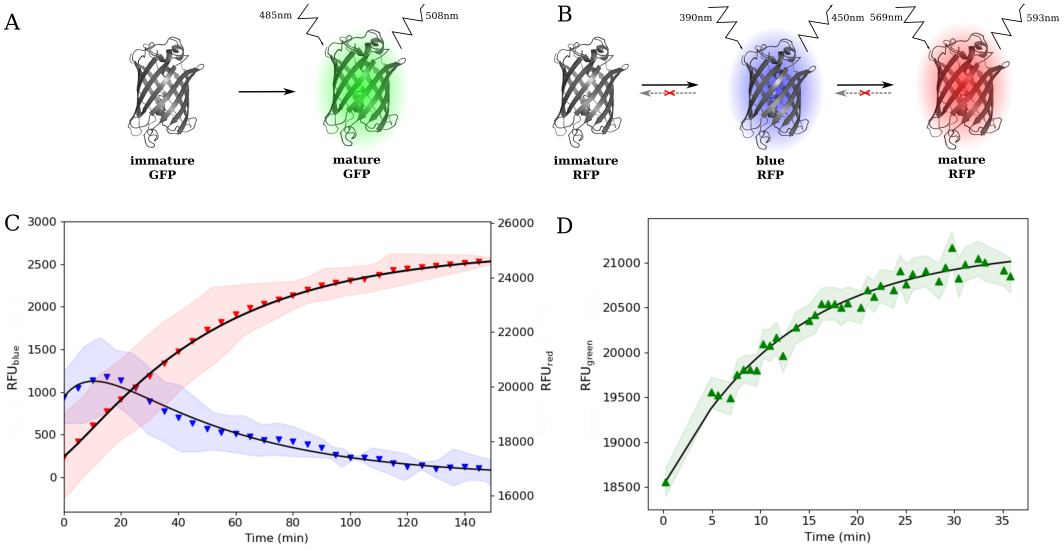


Figure 2.1: Mechanistic maturation models for RFP and GFP. (**A**) Maturation mechanism of GFP. (**B**) Maturation mechanism of RFP. (**C**) Calibration of a mechanistic model of RFP with experimental data: a strain expressing mScarlet-I was grown in MOPS medium supplemented with glucose. At time zero Chloramphenicol was added to the medium to stop translation. Blue and red fluorescence (blue and red triangles respectively) were measured. The plot shows the mean of 6 replicates. Confidence intervals are given by two times the standard error of the mean (SEM). These data were used to fit the model of Eqs 2.3-2.5 and estimate its parameters (best fit: black solid lines). The plot shows that the mechanistic maturation model captures the maturation dynamics well ($R^2 = 0.93$ for blue fluorescence, $R^2 = 0.99$ for red fluorescence).

RFP model:

The RFP model describes the dynamics of the quantities of immature proteins $Im(t)$, intermediate blue proteins $Hm(t)$, and mature red proteins $M(t)$, with appropriate rate constants for inter-species conversion k_{hm}, k_m. The units and

assumptions defined above are also valid for this model (see Text S1 for detailed model definitions).

$$\frac{d}{dt}Im(t) = \alpha(t) \cdot V(t) - (\gamma + k_{hm}) \cdot Im(t), \tag{2.3}$$

$$\frac{d}{dt}Hm(t) = k_{hm} \cdot Im(t) - (\gamma + k_m) \cdot Hm(t), \tag{2.4}$$

$$\frac{d}{dt}M(t) = k_m \cdot Hm(t) - \gamma \cdot M(t). \tag{2.5}$$

To estimate the kinetic maturation parameters of the two models, we conducted growth-arrest experiments where we quantified the fluorescence intensity of each species after adding Chloramphenicol to stop translation. As a consequence $\alpha(t) = 0$, and every increase in fluorescence after adding antibiotics is due to maturation. Whilst extensive growth-arrest experiments have been conducted by Balleza *et al.* [93] to quantify the kinetics of mature FPs, we did our own experiments to also track and quantify the blue intermediate in RFP maturation (Materials and methods).

The maturation curves for RFP are shown in Fig. 2.1C along with the dynamics of its blue intermediate. The two curves were used to estimate the model parameters (Materials and methods, Text S1), giving rise to an excellent fit ($R^2_{blue} = 0.93$, $R^2_{red} = 0.99$). In parallel, the same experiment was conducted with a green strain. The green curve was used to calibrate the GFP model (Fig. 2.5, $R^2_{green} = 0.97$). The parameters obtained demonstrate the difference in maturation time for the two proteins: the GFP model parameter indicates fast maturation (8 min), whereas for the RFP model we find significantly slower maturation (35 min) (Text S1).

We thus obtained kinetic models, tailored for each FP and its maturation mechanisms, and calibrated by means of targeted experiments.

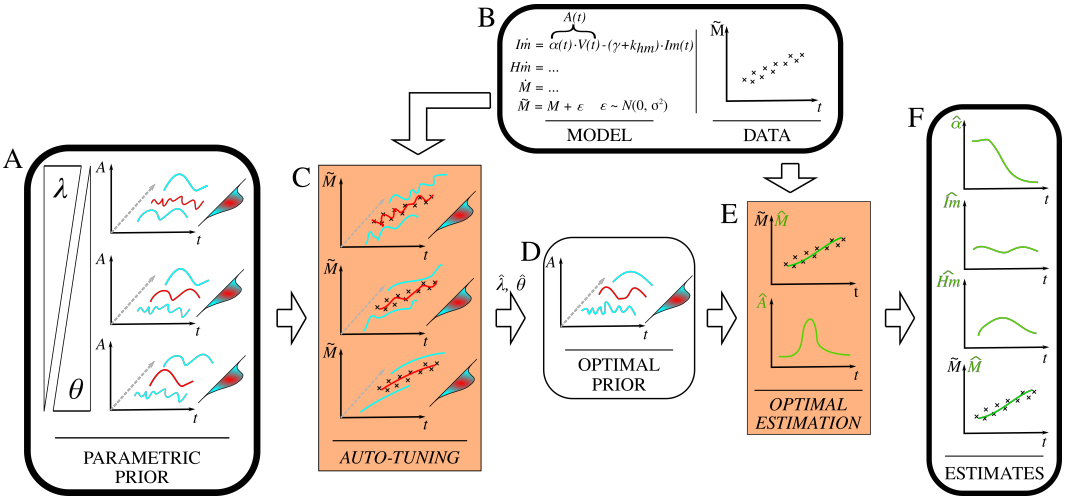


Figure 2.2: Schematic outline of the Bayesian approach for estimating promoter activities. A parametric family of priors (**A**) expresses expected properties of the profile $A(t)$. Larger values of θ (respectively smaller values of λ) assign highest probability to fast-fluctuating (smaller magnitude) profiles (red solid lines). Together with the gene expression model and fluorescence data (**B**), the auto-tuning step (**C**) selects the best values of θ and λ by evaluating the overall match of the distribution of model-predicted fluorescence profiles with the available data via a maximum-likelihood approach. The resulting optimal prior (**D**) is used to produce estimates of the gene expression dynamics and of $A(t)$ via Kalman filtering/smoothing (**E**). Normalization by $V(t)$ eventually yields the estimates of the promoter activity $\alpha(t)$ (**F**). Rounded boxes represent inputs, intermediate results and outputs of the method; rectangular boxes represent procedural steps. The procedure provides robust estimates of promoter activities based on a rigorous, automated selection of regularization parameters θ and λ.

2.3.2 Bayesian approach for the reconstruction of promoter activities from fluorescence data.

Leveraging the models presented above, we now describe our approach to reconstruct an unknown promoter activity profile $\alpha(t)$ from measurements of fluorescent reporter abundance $M(t)$. We assume that K time-sampled, noisy measurements $\tilde{M}_k = M(t_k) + e_k$ are available, where t_k are the measurement times and e_k random measurement errors with assigned variance (for $k = 1, \ldots, K$). $V(t)$ can be determined directly from absorbance data. Therefore, we first address the problem of estimating the time profile of the whole term $A(t) = \alpha(t) \cdot V(t)$ (the ensemble synthesis rate of immature proteins over the whole population), and then deduce the promoter activity $\alpha(t)$, which is the biologically relevant quantity, in a post-processing step.

Reconstructing the input (here $A(t)$) of an ODE system from sampled output measurements (here $\tilde{M}_k$) is a nontrivial inverse problem [123]. A robust estimation of $A(t)$ at any time t within the measurement period can be obtained by Bayesian regularization [124], where the expected qualitative properties of the unknown profile are expressed by a probabilistic prior. Borrowing from Rasmussen *et al.* [141], we fix our prior by assuming that $A(t)$ is the outcome of a stationary stochastic process, characterized by parameters λ and θ (*Materials and methods*). Smaller values of λ (respectively, θ) assign higher probability to small (respectively, slower) fluctuations of $A(t)$.

For the purpose of estimating A from the data, λ and θ play the role of regularization parameters: small values of λ and θ enforce stronger regularity of the solution, but the correspondingly smooth predictions of $M(t)$ may not match fast transitions in the observed fluorescence time-series. Conversely, large values of λ and θ lead to perfect data interpolation at the price of overly irregular estimates of $A(t)$. An optimally balanced choice $\hat{\theta}$ and $\hat{\lambda}$ of regularization parameters can be determined directly from the data by maximum likelihood using fast numerical procedures (*Materials and methods*).

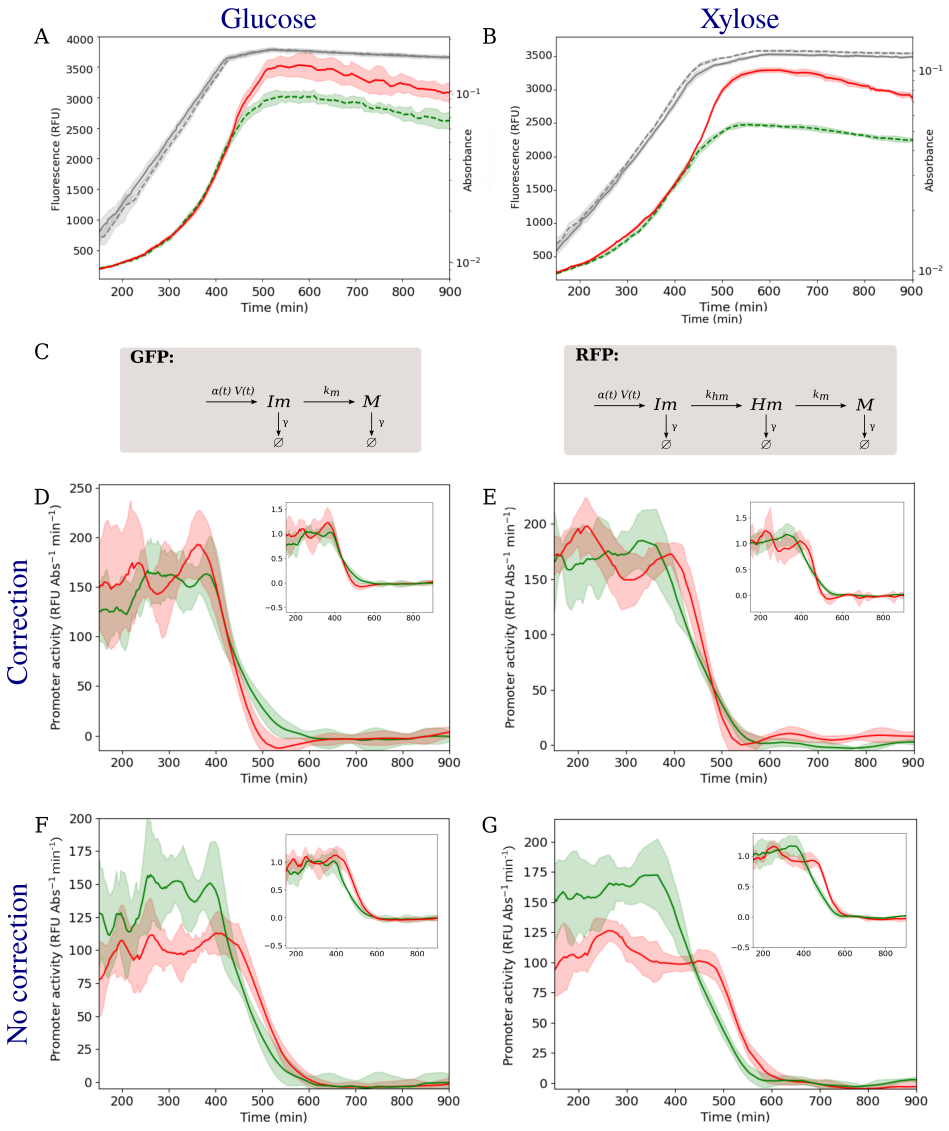


Figure 2.3: Reconstruction of promoter activities from dynamic fluorescence measurements. (**A-B**) Absorbance (grey curves), and red and green fluorescence measurements for a strain expressing either a GFP or an RFP (dotted and solid curves, respectively). Bacteria were grown in a microplate in MOPS minimal medium supplemented with either glucose (A) or xylose (B). The plots show the mean of eight replicates and a confidence interval given by twice the standard error of the mean. (**C**) Model reaction schemes used to reconstruct promoter activities $\alpha(t)$ for GFP and RFP. (**D-E**) Reconstructed promoter activities for data in glucose and xylose (D and E, respectively) for GFP (green) and RFP (red). The obtained signals were plotted on the same scale after appropriate normalization. The plots demonstrate that the Bayesian estimation approach results in promoter activity profiles from green and red FPs that are compatible both qualitatively and quantitatively.

For the above values of regularization parameters, the optimal Bayesian estimate of $A(t)$ from the available data is then [142]

$$\hat{A}(t) = \mathbb{E}[A(t)|\tilde{M}_1, \ldots, \tilde{M}_K], \tag{2.6}$$

where the right-hand side is by definition the mean of the conditional distribution of $A(t)$ given $\tilde{M}_1, \ldots, \tilde{M}_K$. This function of the data statistically minimizes the estimation error. When combined with the maturation model (Eqs 2.1-2.2 or Eqs 2.3-2.5), the model of $A(t)$ gives rise to a linear stochastic differential equation system for which Eq. 2.6 can be efficiently computed via so-called Kalman filtering/smoothing ([132, 142] and *Materials and methods*).

In summary, our method for the estimation of the promoter activity profile $\alpha(t)$ is as follows (Fig. 2.2). Given time-course data from a reporter gene experiment, we first estimate optimal regularization parameters $\hat{\lambda}$ and $\hat{\theta}$ from Eq. 2.10, then use these to calculate the optimal estimate $\hat{A}_{\hat{\lambda},\hat{\theta}}(t)$ at all times t of interest as per Eq. 2.6. Finally, we define our estimate of $\alpha(t)$ as $\hat{\alpha}(t) = \hat{A}_{\hat{\lambda},\hat{\theta}}/\hat{V}(t)$, where $\hat{V}(t)$ is given by the absorbance data. The robust performance of the approach has been verified on synthetic data (Fig. 2.6).

2.3.3 Protein-specific maturation models provide unbiased promoter activity estimates in bacteria.

To investigate whether maturation should be taken into account for the correct reconstruction of promoter activities, we used the Bayesian estimation method alongside the calibrated maturation models for each protein to interpret the fluorescence data obtained from a controlled experimental setup.

Specifically, we have designed an experimental system in *Escherichia coli* where a red and a green FP are under the influence of the same constitutive promoter, proC [143], so it is possible to directly and quantitatively compare the promoter activity profiles for each protein (*Materials and methods*).

The two newly-constructed strains were used to perform kinetic experiments in batch where the absorbance, red and green fluorescence were measured for four different carbon sources: glucose, xylose, acetate and pyruvate.

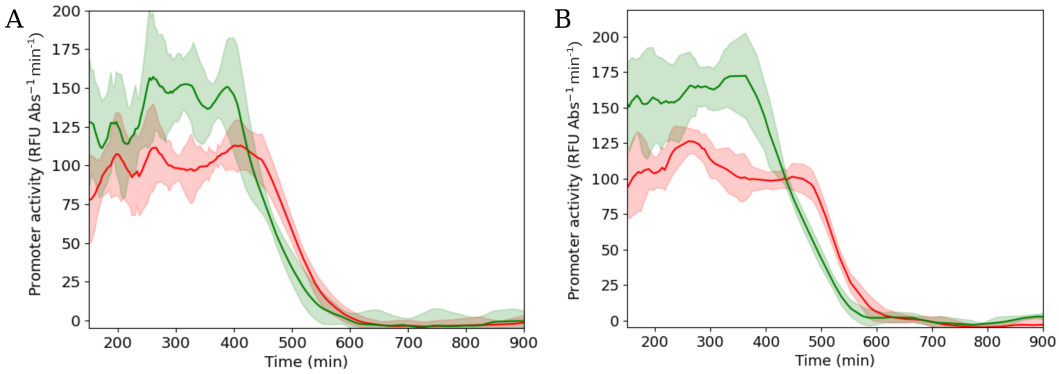


Figure 2.4: Reconstruction of promoter activities without correcting for maturation effects. Reconstruction of promoter activities from the fluorescence data in Figs 2.3A-B (**A** and **B**, respectively), without correcting for maturation. The obtained signals were plotted on the same scale after appropriate normalization. Comparison with the results in Figs 2.3D-E shows that the promoter activities are no longer comparable: the RFP promoter activity is delayed and consistently underestimated.

Absorbance (grey lines) and fluorescence curves (red and green lines) for glucose and xylose are shown in Fig. 2.3A-B. Each curve was then used to reconstruct the underlying promoter activity ($\alpha(t)$) from the models schematized in Fig. 2.3C, following the approach described above.

The reconstructed promoter activity profiles from red and green fluorescence curves are qualitatively and quantitatively identical (Fig. 2.3D-E) despite the fact that the fluorescence profile are quite different. The same observation can be made for the other growth conditions (Figs 2.7-2.14). As a first conclusion, this indicates that our method is capable of correcting for specific maturation effects in the reconstruction of promoter activities from fluorescence data for distinct FPs.

As a second conclusion, we found that accounting for maturation is essential for the correct analysis of fluorescence data. If maturation correction is

neglected, the promoter activities are no longer comparable. Specifically, there is an expected delay in the promoter activity for FPs that are slow-maturing, such as RFP (Fig. 2.4). We also observed a quantitative difference in the sense that the promoter activity from RFP is underestimated in the absence of maturation correction (Fig. 2.4). Proteins with simple maturation like GFP are less dependent on a correction model, but most RFP variants have complex dynamics and require model-based correction [93]. The same can be said for some cyan and blue FPs [93].

The differences in FP maturation mechanisms entail a different sensitivity to environmental conditions. For instance, whereas maturation of GFP requires a single oxygen molecule, RFP maturation requires two molecules [88, 144]. As a consequence, oxygenation problems at high bacterial densities [95] may strongly impact gene expression monitoring by RFP (Fig. 2.15).

In conclusion, we provide a powerful approach for the reconstruction of promoter activities from gene expression data by incorporating FP maturation in the analysis of fluorescence data. While validated on a bacterial model system, the approach is directly applicable to reporter proteins in other prokaryotes and in eukaryotes.

2.4 Materials and Methods

2.4.1 Bacterial strains and plasmids

For all experiments, an *Escherichia coli* BW25113 strain was used that contains a deletion of the *fhuA* gene, making it resistant to phage contaminations (Δ*fhuA*). The *pEB2-mScarlet-I* reporter plasmid [93] was transformed into the above strain using $CaCl_2$. A corresponding *pEB2-GFPmut2* plasmid was constructed using Gibson assembly [145]. In both plasmids, transcription of the reporter gene is controlled by the same constitutive promoter, *proC* [143]. The plasmids have a pSC101 origin of replication, meaning they are low-copy and the number of copies does not vary over different growth phases [146]. This above resulted in the strains: BW25113-Δ*fhuA-pEB2-mScarlet-I* and BW25113-Δ*fhuA-pEB2-GFPmut2*, which were used for all experiments. For

primers and sequences, see Fig. 2.16.

2.4.2 Experimental conditions

The two strains described above, along with a control without plasmids, were grown overnight at 37°C, with shaking at 200rpm in MOPS minimal medium, supplemented with 0.2% glucose. Pre-cultures were inoculated at an OD600 of 0.001 in a 96-well microplate where each well contained MOPS minimal medium [69] supplemented with 0.2% glucose. In the calibration experiments, bacteria were incubated at 37°C in a microplate reader (Tecan Infinite M200 Pro) until reaching an uncorrected absorbance of approximately 0.3. 300 μg/mL of Chloramphenicol (Cm) was then added in each well to stop protein translation. The absorbance, green, blue and red fluorescence intensity were tracked for another 10 hours after translation arrest (details in Text S1). The residual growth after Cm addition was negligible.

For the validation experiments, we used MOPS minimal medium supplemented with 0.05% of glucose, xylose, acetate or pyruvate. This time, the bacteria were incubated in a 96-well microplate without the addition of antibiotics at 37°C for 24 hours. Absorbance, red, green and blue fluorescence were measured. For all experiments, the background was corrected using fluorescence measurements from the strain without plasmids. Glass beads were added to improve oxygenation and the resulting outliers were filtered using WellInverter [147].

2.4.3 Parameter estimation and model selection.

The parameters of our maturation models, which are ordinary differential equation systems in the variables Im, Hm and M (abundance of immature, half-mature and mature fluorescent reporter proteins) were estimated using the

least-squared method in Python. Since we added Cm in the medium, in the calibration experiment, no new protein was produced and therefore the promoter activity for this experiment is zero. A scaling factor z was used and estimated along with the other parameters (k_{hm} and k_m) to relate blue and red fluorescence measurements, so that $Hm = z \cdot \overline{Hm}$, where Hm is expressed in units $\mathrm{RFU_{red}}$ and $\overline{Hm}$ in units $\mathrm{RFU_{blue}}$. We also estimated the initial concentration of immature protein Im_0. Four variations of the two-step maturation model were considered, having zero, one or two backflows (equations in Text S1). We used the Akaike Information Criterion (AIC) statistic to select the best model among the four candidates, which resulted in the model with no backflows. An identifiability analysis by means of a bootstrapping procedure showed that the parameters of this model are identifiable (Fig. 2.17). The model was then used in the Kalman filtering/smoothing algorithm for the reconstruction of promoter activities. More details can be found in Text S1.

2.4.4 Implementation of promoter activity estimation via Kalman smoothing.

In section *Results and discussion* of the the main text, promoter activity is expressed in terms of a time-varying profile $A(t)$. Abundance of mature fluorescent reporter proteins, $M(t)$, depends on $A(t)$ via the maturation models of Eqs 2.1-2.2 and Eqs 2.3-2.5, written in the form $\dot{x}(t) = Bx(t) + CA(t)$, where $M(t)$ is an entry of vector $x(t)$ together with $Im(t)$ and (where applicable) $Hm(t)$. Matrices B and C are defined by the kinetic parameters occurring in the maturation models.

In order to estimate $A(t)$ from fluorescence measurements $\tilde{M}_k = M(t_k) + e_k$ taken at discrete times t_k and corrupted by (Gaussian) noise e_k ($k = 1, \ldots, K$), we rely on the usage of a probabilistic prior on $A(t)$ expressed in the form of a stochastic differential equation. Together with a maturation model of interest, this allows us to write an augmented system of linear stochastic differential equations and to apply Kalman filtering and smoothing [142] for the statistically optimal, computationally efficient solution of the estimation problem.

We define the prior on $A(t)$ by the stochastic differential equation

$$dA(t) = -\theta \cdot A(t) \cdot dt + \lambda \cdot dW(t), \tag{2.7}$$

where $W(t)$ is the standard Wiener (white noise) process, and $\lambda, \theta > 0$ are parameters describing the (magnitude and time-scale of) fluctuations in $A(t)$. Choosing $A(0)$ as a zero-mean Gaussian random variable with variance $\lambda^2/2\theta$ ensures process stationarity. For the estimation problem, λ and θ play the role of regularization parameters, *i.e.* they ensure that reconstruction of $A(t)$ is robust to measurement noise and sparse sampling.

Assume that optimal values $\hat{\lambda}$, $\hat{\theta}$ of λ and θ have been determined (see below). Combining Eq. 2.7 with the maturation model of interest yields the augmented system

$$d\begin{bmatrix} x(t) \\ A(t) \end{bmatrix} = \begin{bmatrix} B & C \\ 0 & -\hat{\theta} \end{bmatrix} \cdot \begin{bmatrix} x(t) \\ A(t) \end{bmatrix} dt + \begin{bmatrix} 0 \\ \hat{\lambda} \end{bmatrix} dW(t), \tag{2.8}$$

$$\tilde{M}_k = \begin{bmatrix} D & 0 \end{bmatrix} \cdot \begin{bmatrix} x(t_k) \\ A(t_k) \end{bmatrix} + e_k, \tag{2.9}$$

for an appropriate 0-1 row vector D such that $D \cdot x(t_k) = M(t_k)$. Let $v(t_k)$ be the augmented state vector comprising $x(t_k)$ and $A(t_k)$. For a generic index j, the optimal Bayesian estimate of $v(t_k)$ from the data up to time t_j is the conditional expectation $\hat{v}^j(t_k) = \mathbb{E}[v(t_k)|\tilde{M}_1, \ldots, \tilde{M}_j]$.

In particular, we are interested in the calculation of $\hat{v}^K(t_k)$ (optimal estimates of v from all the K measurements) at all times t_k, from which optimal estimates of A as well as of Im and (where applicable) Hm follow (they are all entries of v). This is obtained by the following two-sweep data processing algorithm.

In a first sweep (Kalman filtering), estimates $\hat{v}^k(t_k)$ along with one-step predictions $\hat{v}^{k-1}(t_k)$ are computed iteratively for k from 1 up to K, together with corresponding estimation error variance matrices $P^k(t_k)$ and $P^{k-1}(t_k)$. The iteration is initialized by a noninformative prior on $v(t_1)$ (arbitrarily large variance $P^0(t_1)$ and $\hat{v}^0(t_1)$ null). In a second sweep (Kalman smoothing), estimates $\hat{v}^K(t_k)$ exploiting both past and future measurements are computed

at all times t_k by back-processing $\tilde{M}_k$ and the estimates from the first sweep in a time-reversed iteration (k from K down to 1). This second (often neglected) sweep provides a significant refinement of the estimates [148]. The filtering and smoothing formulas are standard [134, 142].

Given several experimental replicates (several datasets $\tilde{M}_1, \ldots, \tilde{M}_K$), we calculate promoter activity estimates $\hat{A}$ separately for every replicate. Confidence intervals on the estimates $\hat{A}$ are then calculated as twice the standard error of the mean of the estimates obtained over the different replicates. Confidence intervals for the estimates of Im, Hm and M are obtained in a similar manner.

2.4.5 Optimal choice of regularization parameters.

An optimal choice of λ and θ can be determined by solving the maximum likelihood problem

$$(\hat{\lambda}, \hat{\theta}) = \arg\max_{(\lambda,\theta)} f_{\lambda,\theta}(\tilde{M}_1, \ldots, \tilde{M}_K), \tag{2.10}$$

where $f_{\lambda,\theta}(\,\cdot\,)$ is the probability density function for the observed data under the parameter-dependent prior in Eq. 2.7. In practice, for any value of (λ, θ), evaluation of this likelihood can be performed by the same filtering tools described above. Generalizing previous definitions, let $\hat{v}^{k-1}_{\lambda,\theta}(t_k)$ and $P^{k-1}_{\lambda,\theta}(t_k)$ be the optimal one-step prediction of $v(t_k)$ and corresponding error variance matrix under generic values of λ and θ. Their calculation uses the same filtering sweep described in the previous section. In the light of Eq. 2.9, one can then evaluate at any k the conditional Gaussian densities

$$f_{\lambda,\theta}(\tilde{M}_k | \tilde{M}_1, \ldots, \tilde{M}_{k-1}) = \frac{1}{\sqrt{2\pi\Lambda_k}} \exp\left[-\frac{1}{2} \cdot \frac{\left(\tilde{M}_k - \widehat{M}^{k-1}_k\right)^2}{\Lambda_k} \right],$$

where $\widehat{M}^{k-1}_k = \begin{bmatrix} D & 0 \end{bmatrix} \cdot \hat{v}^{k-1}_{\lambda,\theta}(t_k)$ and $\Lambda_k = \begin{bmatrix} D & 0 \end{bmatrix} \cdot P^{k-1}_{\lambda,\theta}(t_k) \cdot \begin{bmatrix} D & 0 \end{bmatrix}^T + \sigma^2_k$. In turn, σ^2_k is the variance of the measurement error e_k, as determined from the

variance of $\tilde{M}_k$ across multiple experimental replicas. Thus, using Bayes' law, the likelihood in Eq. 2.10 can be evaluated in terms of the above densities as

$$f_{\lambda,\theta}(\tilde{M}_1, \ldots, \tilde{M}_K) = \prod_{k=1}^{K} f_{\lambda,\theta}(\tilde{M}_k | \tilde{M}_1, \ldots, \tilde{M}_{k-1}).$$

For numerical convenience, optimization (Eq. 2.10) is rather solved by the minimization of the negative logarithm of the likelihood, using Python solver `minimize` of the `scipy.optimize` module.

2.5 Supplementary Information on Chapter 2

2.5.1 Supplementary Figures

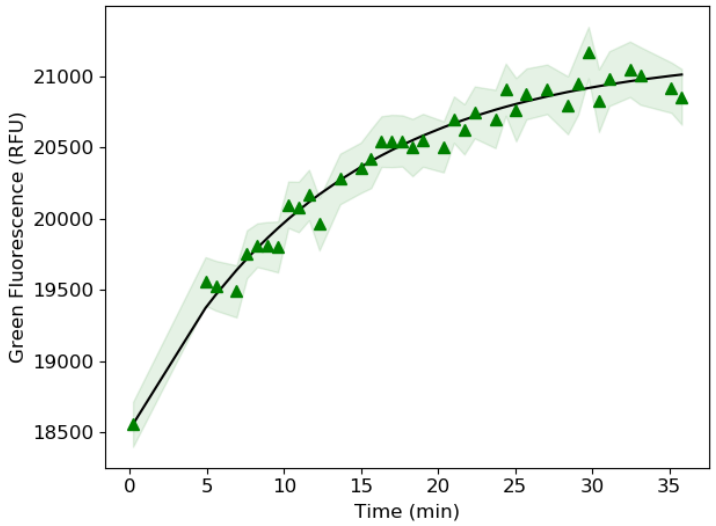


Figure 2.5: Calibration of mechanistic model of GFP with experimental data. A strain expressing GFPmut2 was grown in MOPS medium supplemented with glucose. At time zero Chloramphenicol was added to the medium to stop translation. Green fluorescence (green triangles) was measured. The plot shows the mean of 6 replicates. Confidence intervals are given by two times the standard error of the mean (SEM). These data were used to fit the model of Eqs 2.1-2.2 and estimate its parameter (best fit: black solid line).

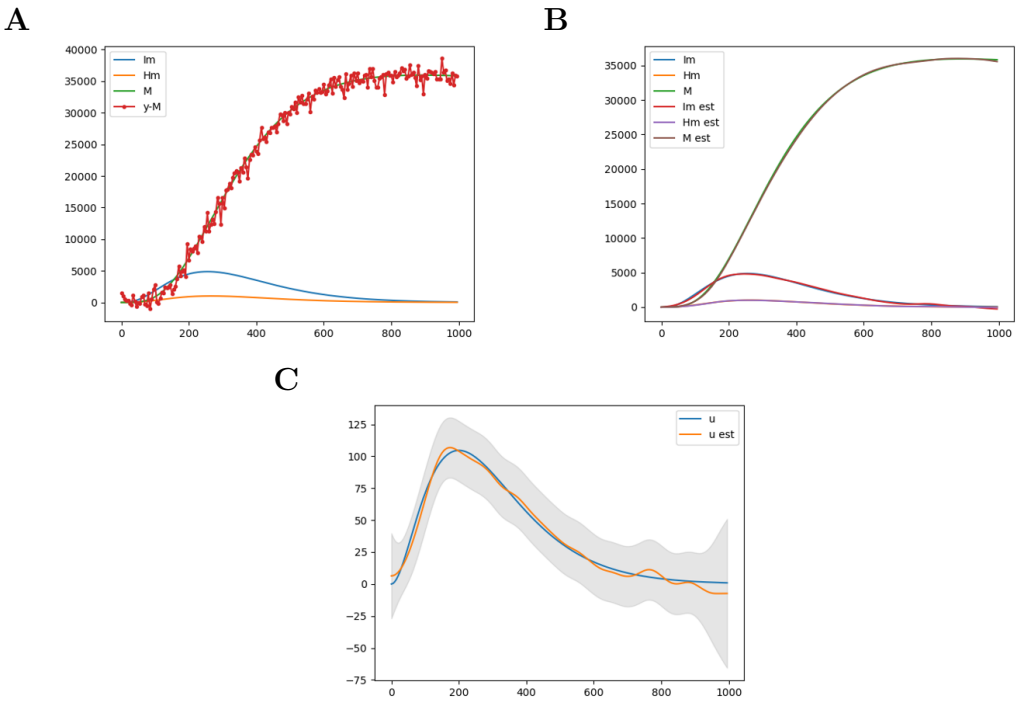


Figure 2.6: Performance of the Bayesian inference approach on synthetic data. (**A**) Synthetic data generated from a fixed promoter activity profile, using the model of Eqs 2.3-2.5 with added random white noise $\sim N(0, 1000)$ (red curve). (**B**) Estimated quantities for each species (Im est, Hm est, M est) versus generated quantities (Im, Hm, M). The estimation procedure uses the Kalman filtering/smoothing approach described in the *Materials and methods*. (**C**) Reconstructed promoter activity profile and confidence interval compared to the promoter activity used to generate the data (u and u est, respectively). The algorithm is capable of robustly reconstructing the promoter activity profiles from synthetic data.

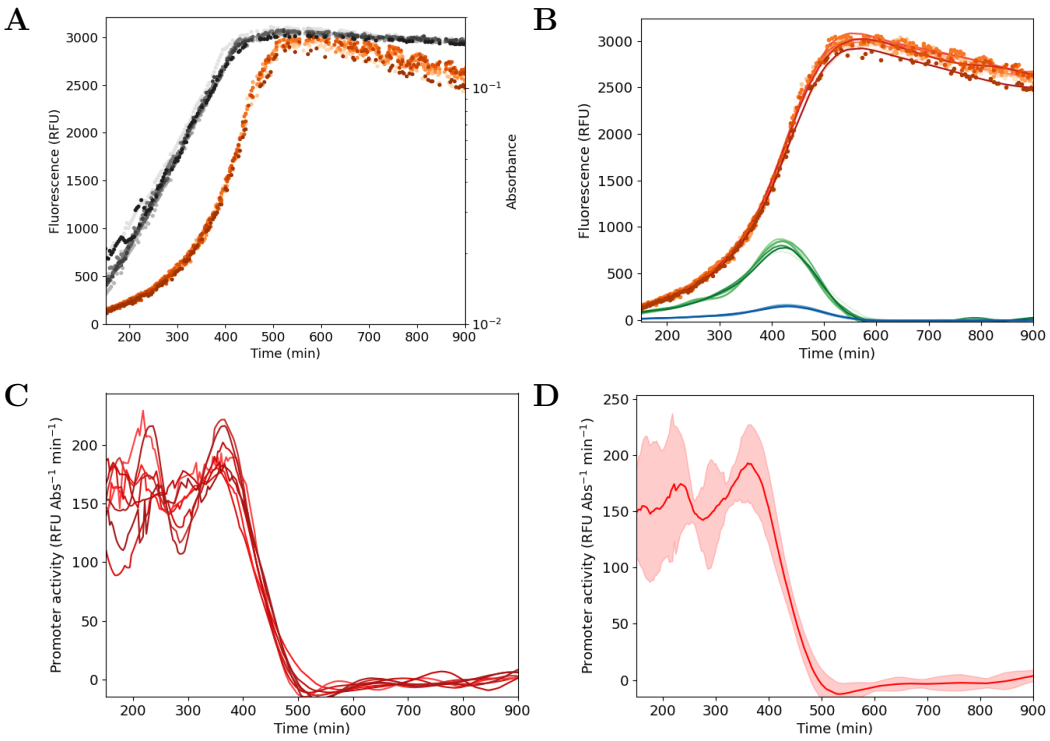


Figure 2.7: Reconstruction of promoter activities from RFP data of bacteria growing in minimal medium with glucose. (**A**) Red fluorescence (orange to red dots) and absorbance (grey to black dots) for 8 replicates. (**B**) Estimation of $Im(t)$ (green curves), $Hm(t)$ (blue curves) and $M(t)$ (orange to red curves) using the RFP maturation model and the data from panel A, also presented here (orange to red dots). (**C**) Reconstruction of individual promoter activities for each replicate using the Bayesian inference procedure. (**D**) Mean of reconstructed promoter activities and confidence intervals given by two times the standard error of the mean.

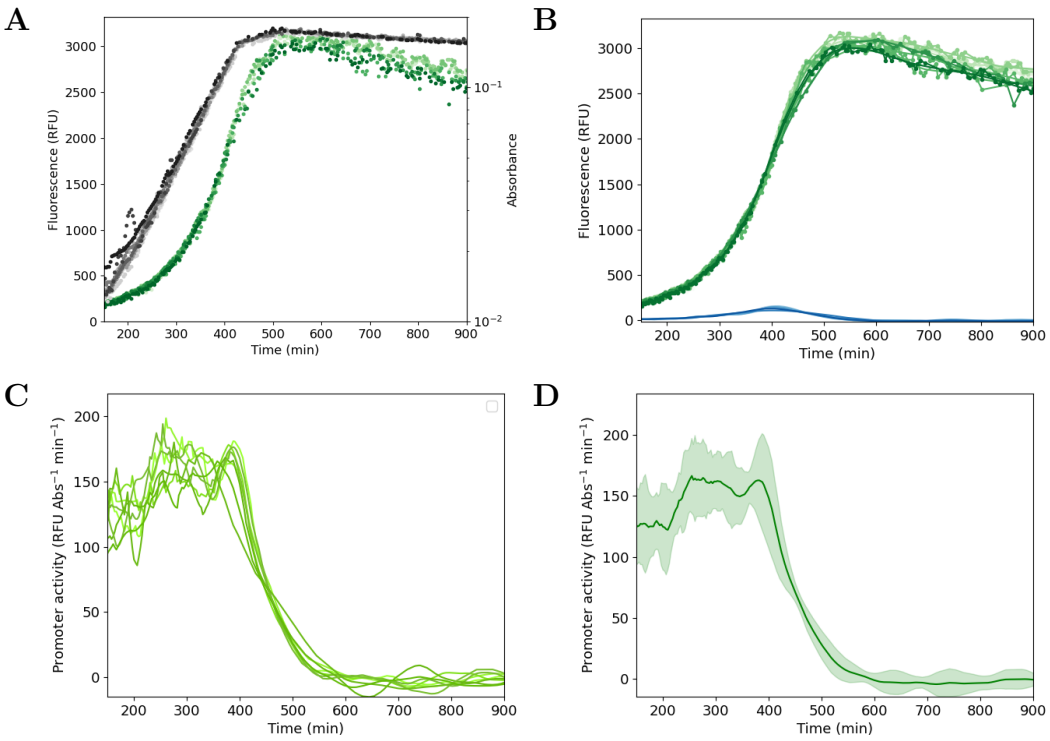


Figure 2.8: Reconstruction of promoter activities from GFP data of bacteria growing in minimal medium with glucose. (**A**) Green fluorescence (green dots) and absorbance (grey to black dots) for 8 replicates. (**B**) Estimation of $Im(t)$ (blue curves) and $M(t)$ (green curves) using the GFP maturation model and the data from panel A, also presented here (green dots). (**C**) Reconstruction of individual promoter activities for each replicate using the Bayesian inference procedure. (**D**) Mean of reconstructed promoter activities and confidence intervals given by two times the standard error of the mean.

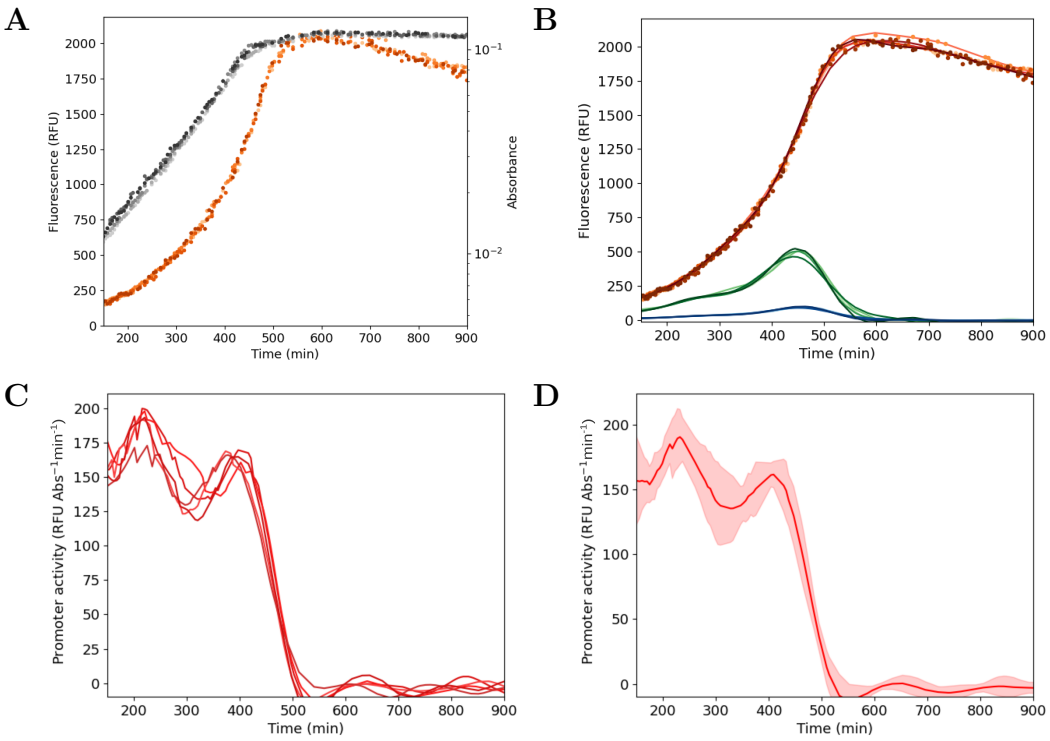


Figure 2.9: Reconstruction of promoter activities from RFP data of bacteria growing in minimal medium with xylose. (**A**) Red fluorescence (orange to red dots) and absorbance (grey to black dots) for 4 replicates. (**B**) Estimation of $Im(t)$ (green curves), $Hm(t)$ (blue curves) and $M(t)$ (orange to red curves) using the RFP maturation model and the data from panel A, also presented here (orange to red dots). (**C**) Reconstruction of individual promoter activities for each replicate using the Bayesian inference procedure. (**D**) Mean of reconstructed promoter activities and confidence intervals given by two times the standard error of the mean.

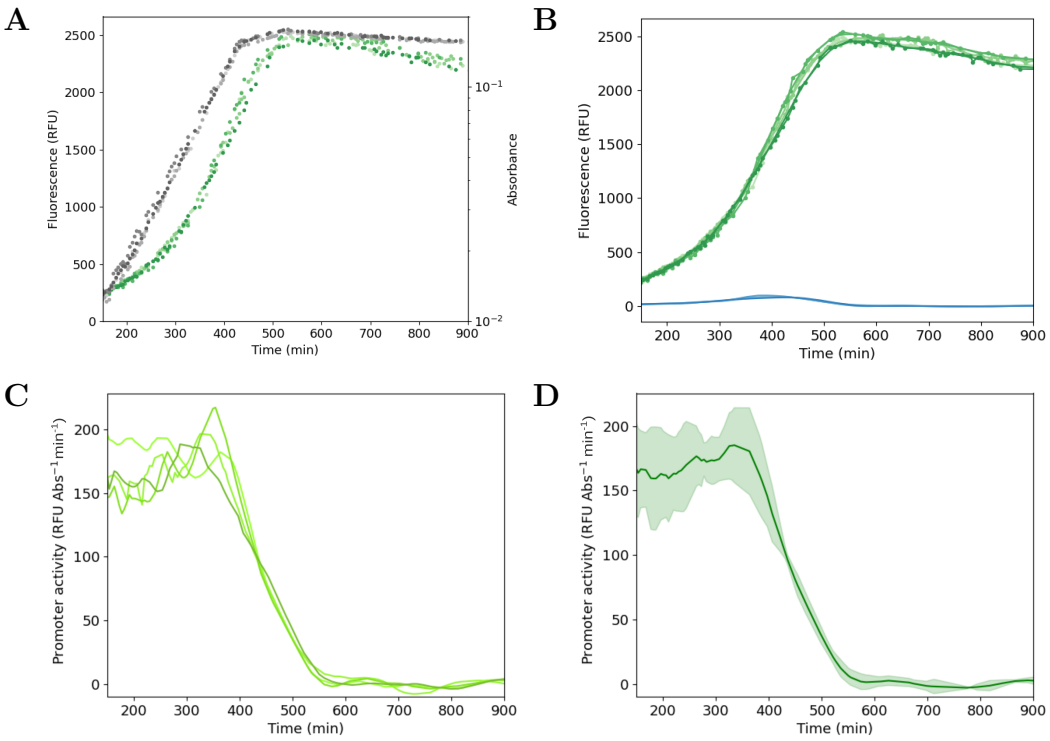


Figure 2.10: Reconstruction of promoter activities from GFP data of bacteria growing in minimal medium with xylose. (**A**) Green fluorescence (green dots) and absorbance (grey to black dots) for 4 replicates. (**B**) Estimation of $Im(t)$ (blue curves) and $M(t)$ (green curves) using the GFP maturation model and the data from panel A, also presented here (green dots). (**C**) Reconstruction of individual promoter activities for each replicate using the Bayesian inference procedure. (**D**) Mean of reconstructed promoter activities and confidence intervals given by two times the standard error of the mean.

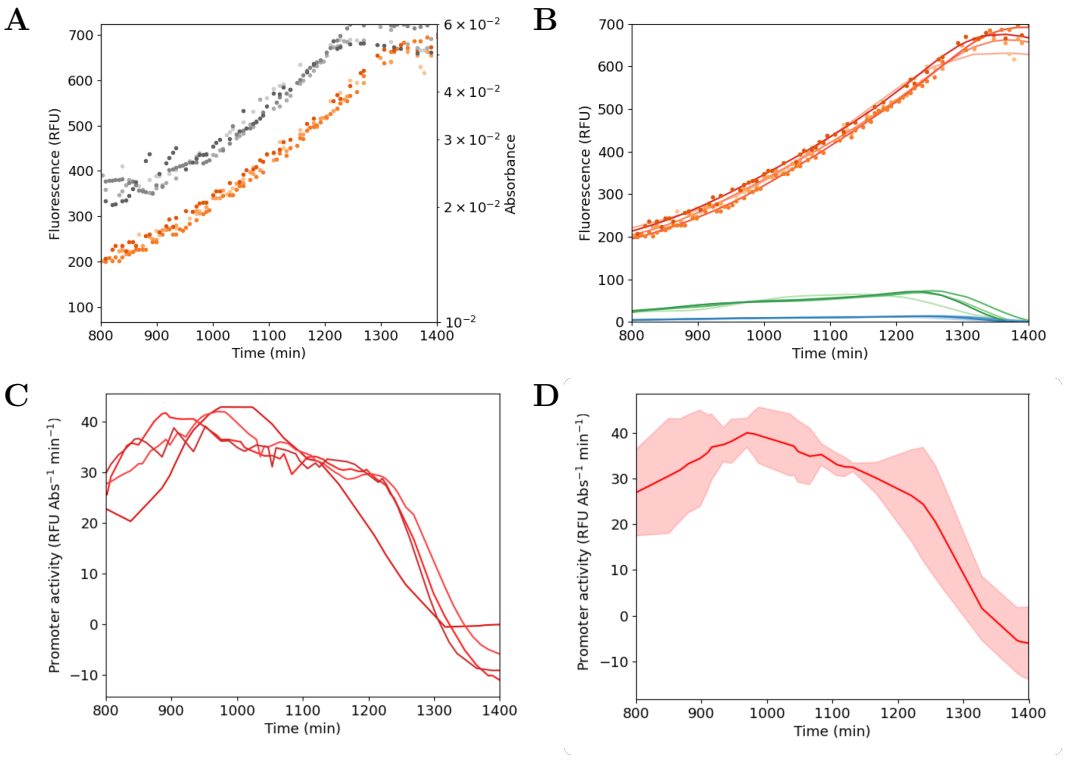


Figure 2.11: Reconstruction of promoter activities from RFP data of bacteria growing in minimal medium with acetate. (**A**) Red fluorescence (orange to red dots) and absorbance (grey to black dots) for 4 replicates. (**B**) Estimation of $Im(t)$ (green curves), $Hm(t)$ (blue curves) and $M(t)$ (orange to red curves) using the RFP maturation model and the data from panel A, also presented here (orange to red dots). (**C**) Reconstruction of individual promoter activities for each replicate using the Bayesian inference procedure. (**D**) Mean of reconstructed promoter activities and confidence intervals given by two times the standard error of the mean.

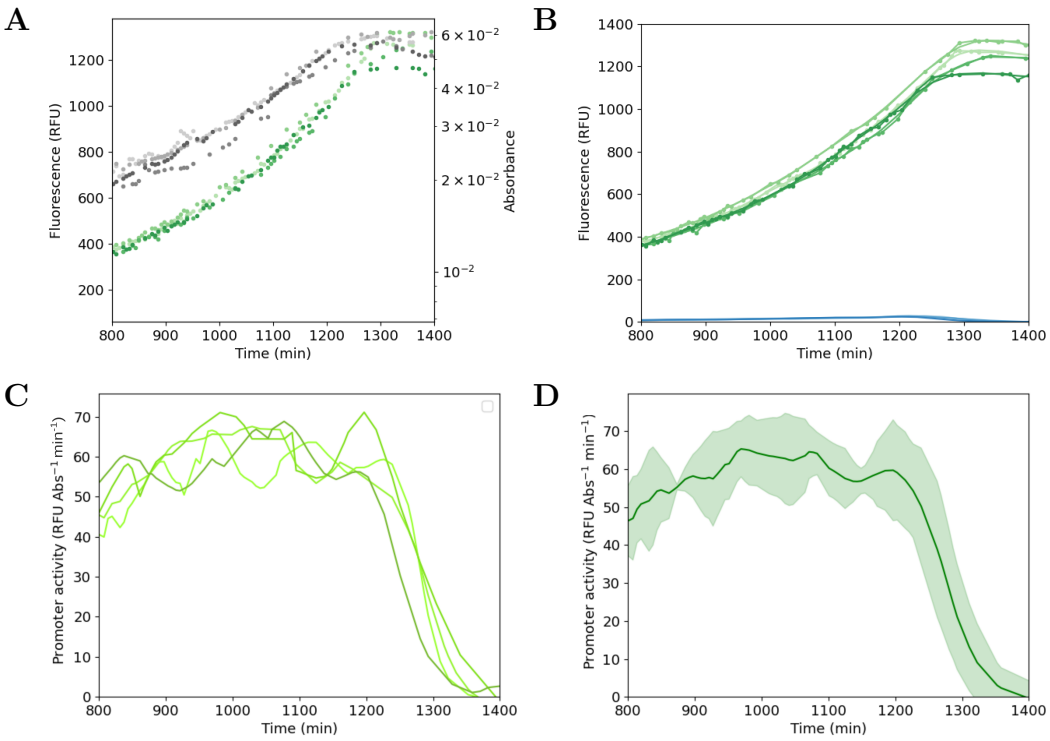


Figure 2.12: Reconstruction of promoter activities from GFP data of bacteria growing in minimal medium with acetate. (**A**) Green fluorescence (green dots) and absorbance (grey to black dots) for 4 replicates. (**B**) Estimation of $Im(t)$ (blue curves) and $M(t)$ (green curves) using the GFP maturation model and the data from panel A, also presented here (green dots). (**C**) Reconstruction of individual promoter activities for each replicate using the Bayesian inference procedure. (**D**) Mean of reconstructed promoter activities and confidence intervals given by two times the standard error of the mean.

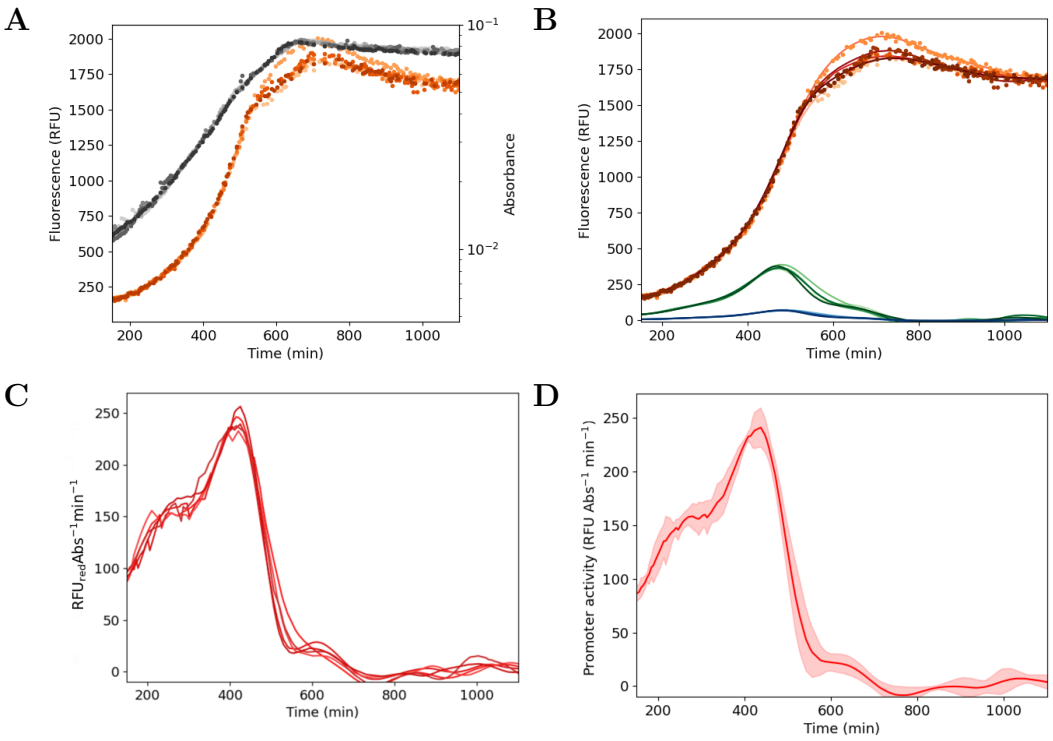


Figure 2.13: Reconstruction of promoter activities from RFP data of bacteria growing in minimal medium with pyruvate. (**A**) Red fluorescence (orange to red dots) and absorbance (grey to black dots) for 4 replicates. (**B**) Estimation of $Im(t)$ (green curves), $Hm(t)$ (blue curves) and $M(t)$ (orange to red curves) using the RFP maturation model and the data from panel A, also presented here (orange to red dots). (**C**) Reconstruction of individual promoter activities for each replicate using the Bayesian inference procedure. (**D**) Mean of reconstructed promoter activities and confidence intervals given by two times the standard error of the mean.

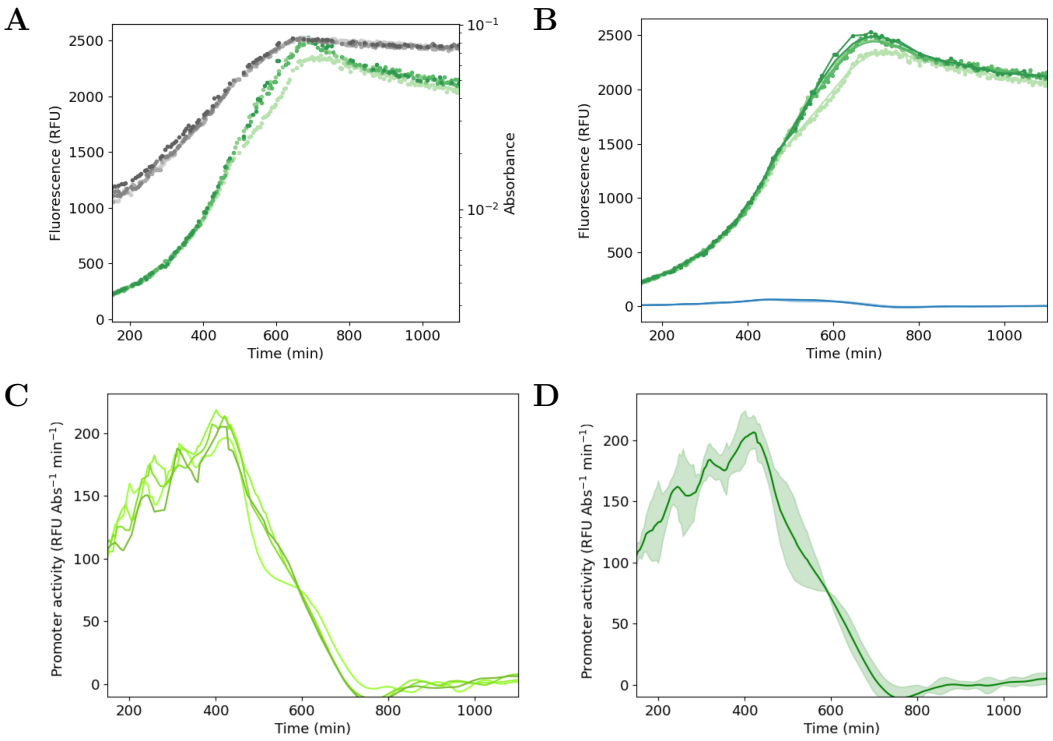


Figure 2.14: Reconstruction of promoter activities from GFP data of bacteria growing in minimal medium with pyruvate. (**A**) Green fluorescence (green dots) and absorbance (grey to black dots) for 4 replicates. (**B**) Estimation of $Im(t)$ (blue curves) and $M(t)$ (green curves) using the GFP maturation model and the data from panel A, also presented here (green dots). (**C**) Reconstruction of individual promoter activities for each replicate using the Bayesian inference procedure. (**D**) Mean of reconstructed promoter activities and confidence intervals given by two times the standard error of the mean.

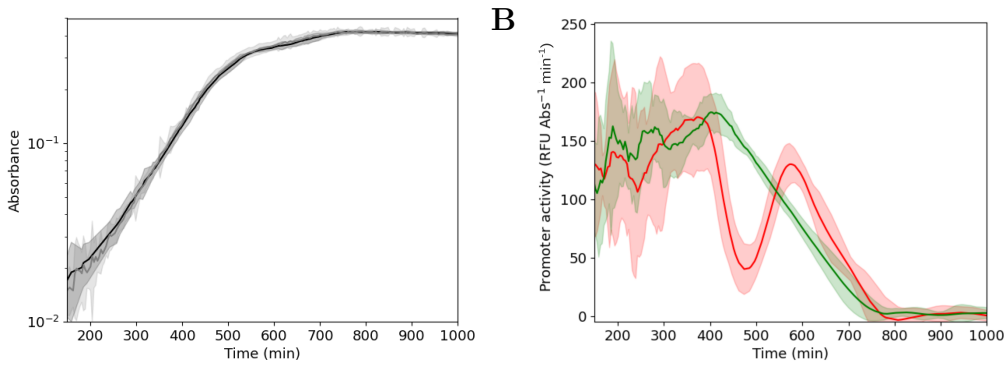
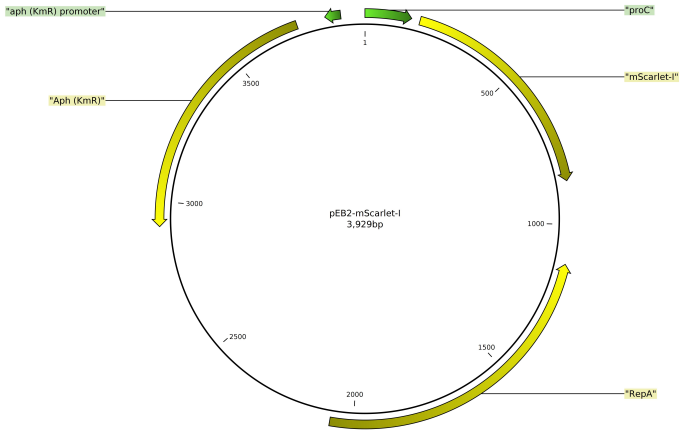


Figure 2.15: Reconstruction of promoter activities in bacterial cultures grown to high biomass densities. (**A**) Strains expressing RFP and GFP were grown in minimal medium supplemented with 0.2% glucose (black and grey curves, respectively). The corresponding fluorescence curves were used to reconstruct the promoter activities using the GFP (green curve) and RFP (red curve) maturation models. The two reconstructed curves were plotted on the same axis after appropriate normalization (**B**). A limitation of oxygen at high biomass densities most likely entails delays in the maturation of RFP and causes a dip in the promoter activity.

A



B

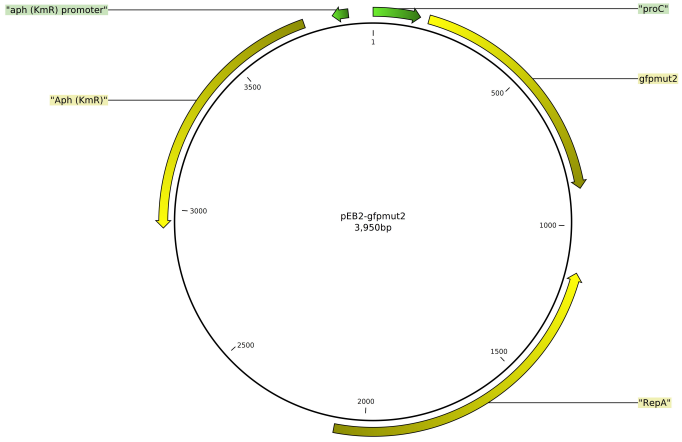


Figure 2.16: Plasmid design. (**A**) pEB2-mScarlet-I plasmid. (**B**) pEB2-gfpmut2 plasmid. Details on plasmid construction in Materials and methods. We used the following primer sequences: gfpmut2_gib_pEb2_fw: GAACTATACAAATAAATGTCCAGACCTGCA, gfpmut2_gib_pEb2_rv: GAAGGAGATATACATATGAGTAAAGGAGAAGAACTTTTC, pEB2_gib_gfpmut2_fw: GAACTATACAAATAAATGTCCAGACCTGCAG and pEB2_gib_gfpmut2_rv: TTCTCCTTTACTCATATGTATATCTCCTTCTTAAATCTAGAAAAG

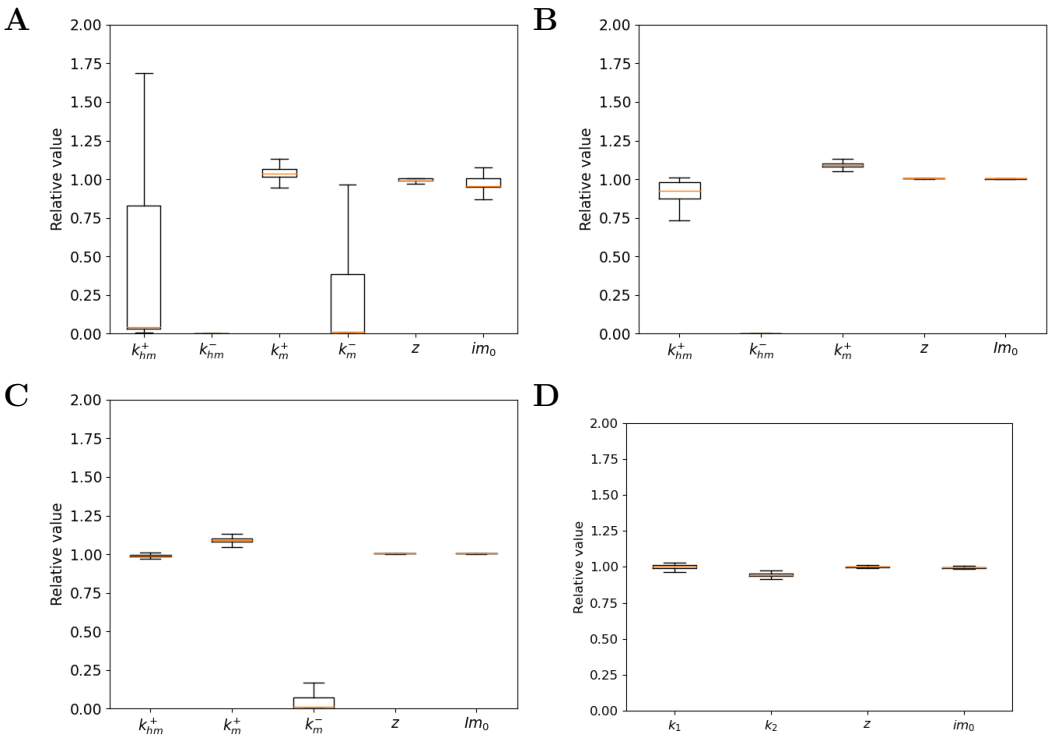


Figure 2.17: Results of identifiability analysis using a bootstrapping procedure. The four RFP models were used alongside a bootstrapping procedure described to check the robustness of the estimation procedure and signal identifiability issues (Text S1). The parameter values returned by the bootstrapping procedure were divided by the parameter value obtained from the actual dataset. The center of each box represents the median and its height the interquantile range. The whiskers represent the minimum and maximum of each estimate. Models I-IV: panels **A-D**, respectively. The parameter values in Model IV are precisely determined, whereas the other models have identifiability issues.

2.5.2 Text S1

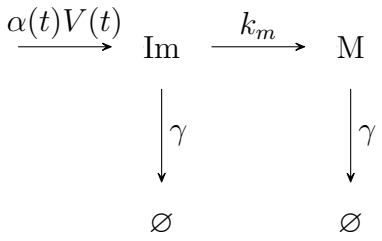
2.5.2.1 Model definition, calibration and selection

Model equations

The models are defined on the population-level: they describe the total quantity of protein species (red, green, blue, colorless) in a growing population of cells. Protein quantities are expressed in relative fluorescence units (RFU), assumed proportional to molar units. Biomass is quantified by absorbance (Abs), assumed proportional to the volume of bacterial population.

GFP model:

I



$$\frac{d}{dt}Im(t) = \alpha(t) \cdot V(t) - (\gamma + k_m) \cdot Im(t), \tag{2.11}$$

$$\frac{d}{dt}M(t) = k_m \cdot Im(t) - \gamma \cdot M(t). \tag{2.12}$$

For GFP, we assume that the abundance of immature protein $Im(t)$ is produced at a rate $\alpha(t) \cdot V(t)$, where $\alpha(t)$ [RFU min^{-1} Abs^{-1}] is the specific production rate, per unit population volume, and $V(t)$ [Abs] the volume of the growing population. The specific production rate consists of the maximum production rate modulated by the promoter activity, where it is assumed that

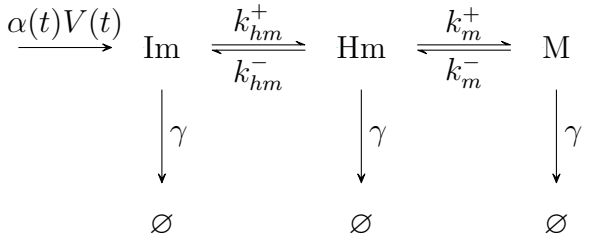
the dynamics of the intermediate mRNA species can be ignored [77]. The conversion of immature to mature protein occurs at a rate proportional to the quantity of immature protein with constant k_m [min^{-1}] and all proteins are degraded at the same rate with rate constant γ [min^{-1}].

RFP models:

The RFP model is composed of three variables, modeling the quantity of immature protein $Im(t)$, intermediate blue protein $Hm(t)$, and mature red protein $M(t)$, with appropriate rate constants for inter-species conversion $k_{hm}^{+}, k_{hm}^{-}, k_{m}^{+}$, k_{m}^{-}. The units and assumptions described above are also valid for this model (see Table 2.1 for parameters and units).

We consider four variants of the RFP model (I-IV) having zero, one or two backflows, since we have no prior information on the reversibility of each reaction.

I

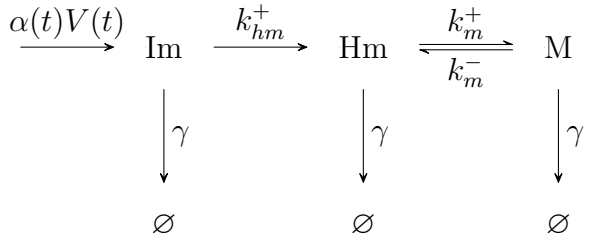


$$\frac{d}{dt}Im(t) = \alpha(t) \cdot V(t) - (k_{hm}^{+} + \gamma) \cdot Im(t) + k_{hm}^{-} \cdot Hm(t), \tag{2.13}$$

$$\frac{d}{dt}Hm(t) = k_{hm}^{+} \cdot Im(t) - (k_{hm}^{-} + k_{m}^{+} + \gamma) \cdot Hm(t) + k_{m}^{-} \cdot M(t), \tag{2.14}$$

$$\frac{d}{dt}M(t) = k_{m}^{+} \cdot Hm(t) - (k_{m}^{-} + \gamma) \cdot M(t). \tag{2.15}$$

II

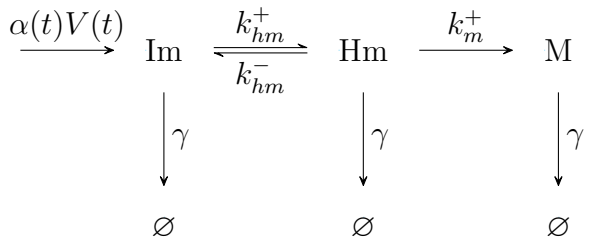


$$\frac{d}{dt}Im(t) = \alpha(t) \cdot V(t) - (k^+_{hm} + \gamma) \cdot Im(t), \tag{2.16}$$

$$\frac{d}{dt}Hm(t) = k^+_{hm} \cdot Im(t) - (k^+_m + \gamma) \cdot Hm(t) + k^-_m \cdot M(t), \tag{2.17}$$

$$\frac{d}{dt}M(t) = k^+_m \cdot Hm(t) - (k^-_m + \gamma) \cdot M(t). \tag{2.18}$$

III

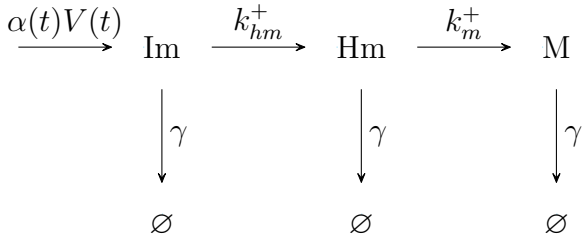


$$\frac{d}{dt}Im(t) = \alpha(t) \cdot V(t) - (k^+_{hm} + \gamma) \cdot Im(t) + k^-_{hm} \cdot Hm(t), \tag{2.19}$$

$$\frac{d}{dt}Hm(t) = k^+_{hm} \cdot Im(t) - (k^-_{hm} + k^+_m + \gamma) \cdot Hm(t), \tag{2.20}$$

$$\frac{d}{dt}M(t) = k^+_m \cdot Hm(t) - \gamma \cdot M(t). \tag{2.21}$$

IV



$$\frac{d}{dt}Im(t) = \alpha(t) \cdot V(t) - (k_{hm}^{+} + \gamma) \cdot Im(t), \tag{2.22}$$

$$\frac{d}{dt}Hm(t) = k_{hm}^{+} \cdot Im(t) - (k_{m}^{+} + \gamma) \cdot Hm(t), \tag{2.23}$$

$$\frac{d}{dt}M(t) = k_{m}^{+} \cdot Hm(t) - \gamma \cdot M(t). \tag{2.24}$$

Parameter	Definition	Meaning	Units
k_{hm}^{+}	$\in \mathbb{R}_{+}$	maturation constant	min^{-1}
k_{hm}^{-}	$\in \mathbb{R}_{+}$	maturation constant	min^{-1}
k_{m}^{+}	$\in \mathbb{R}_{+}$	maturation constant	min^{-1}
k_{m}^{-}	$\in \mathbb{R}_{+}$	maturation constant	min^{-1}
k_{hm}	$\in \mathbb{R}_{+}$	maturation constant	min^{-1}
k_{m}	$\in \mathbb{R}_{+}$	maturation constant	min^{-1}
$\alpha(t)$	$\in \mathbb{R}$	maximum protein production	RFU min^{-1} Abs^{-1}
$V(t)$	$\in \mathbb{R}$	population volume	Abs
z	$\in \mathbb{R}$	fluorescence correction factor	$\text{RFU}_{\text{red}}/\text{RFU}_{\text{blue}}$
$Im(t)$, $Hm(t)$, $M(t)$	$\in \mathbb{R}$	fluorescence intensity	RFU_{red}
$\overline{Hm}(t)$	$\in \mathbb{R}$	fluorescence intensity	RFU_{blue}

Table 2.1: Definition and units of variables and parameters in the maturation models I-IV.

Equations used for calibration

For the calibration of the models and model selection, we performed experiments where we added Cm, a translation inhibitor in growing cultures of bacteria containing the two reporter proteins (*Materials and methods*). When Cm is added, no new proteins are synthesized. Therefore, all increase in fluorescence can only result from maturation of residual immature protein. For calibration, we thus use models without protein synthesis ($\alpha = 0$). We additionally consider that γ is negligible in this case, since maturation dynamics are very fast compared to protein degradation.

Since there is no new protein synthesis, in the case of RFP models, we can write the conservation relation

$$m_t = Im(t) + Hm(t) + M(t), \tag{2.25}$$

where m_t is the total amount of proteins. Through equation 2.25 we can write that $Im(t) = m_t - Hm(t) - M(t)$. The $Im(t)$ species at each instance of time is equal to $Im_0+Hm_0+M_0-Hm(t)-M(t)$, given that $m_t = Im_0+Hm_0+M_0$. We can eliminate variable $Im(t)$ and the corresponding equation for each model variant to obtain a system with two equations. In the case of variant I this gives:

$$\frac{d}{dt}Hm(t) = k^+_{hm}\cdot(Im_0+Hm_0+M_0-Hm(t)-M(t))-(k^-_{hm}+k^+_m)\cdot Hm(t)+k^-_m\cdot M(t), \tag{2.26}$$

$$\frac{d}{dt}M(t) = k^+_m \cdot Hm(t) - k^-_m \cdot M(t). \tag{2.27}$$

As explained above, the quantities of fluorescent proteins in the growing cell population have units RFU, corresponding to measured fluorescence intensities. The different proteins distinguished in the model, emit fluorescence in different colors (blue, red) and each color has its own scale. In order to make

quantities comparable, we decided to express all variables in the red fluorescence scale (RFU_{red} or RFU for short). This notably implies the use of a constant re-scaling factor (z) to convert blue RFU to red RFU. We define:

$$Hm = z \cdot \overline{Hm}, \tag{2.28}$$

where the variable Hm is expressed in units $\mathrm{RFU}_{\mathrm{red}}$ and the variable $\overline{Hm}$ in units $\mathrm{RFU}_{\mathrm{blue}}$. The conversion factor z has unit $\mathrm{RFU}_{\mathrm{red}} \cdot \mathrm{RFU}_{\mathrm{red}}{}^{-1}$.

Thus, the ODE system becomes:

$$\begin{aligned} z \cdot \frac{d}{dt}\overline{Hm}(t) = k_{hm}^{+} \cdot (Im_0 + z \cdot \overline{Hm}_0 + M_0 - z \cdot \overline{Hm}(t) - M(t)) \\ - z \cdot (k_{hm}^{-} + k_m^{+}) \cdot \overline{Hm}(t) + k_m^{-} \cdot M(t), \end{aligned} \tag{2.29}$$

$$\frac{d}{dt}M(t) = k_m^{+} \cdot z \cdot \overline{Hm}(t) - k_m^{-} \cdot M(t). \tag{2.30}$$

With the same reasoning we obtain the other three models:

II $\qquad \mathrm{Im} \underset{k_{hm}^{-}}{\overset{k_{hm}^{+}}{\rightleftharpoons}} \mathrm{Hm} \xrightarrow{k_m^{+}} \mathrm{M}$

$$z \cdot \frac{d}{dt}\overline{Hm}(t) = k_{hm}^{+} \cdot (Im_0 + z \cdot \overline{Hm}_0 + M_0 - z \cdot \overline{Hm}(t) - M(t)) - (k_m^{+} + k_{hm}^{-}) \cdot z \cdot \overline{Hm}(t), \tag{2.31}$$

$$\frac{dM(t)}{dt} = k_m^{+} \cdot z \cdot \overline{Hm}(t). \tag{2.32}$$

III $\text{Im} \xrightarrow{k_{hm}^+} \text{Hm} \underset{k_m^-}{\overset{k_m^+}{\rightleftharpoons}} \text{M}$

$$z\cdot\frac{d}{dt}\overline{Hm}(t) = k_{hm}^+\cdot(Im_0+z\cdot\overline{Hm}_0+M_0-z\cdot\overline{Hm}(t)-M(t))-k_m^+\cdot z\cdot\overline{Hm}(t)+k_m^-\cdot M(t), \tag{2.33}$$

$$\frac{d}{dt}M(t) = k_m^+\cdot z\cdot\overline{Hm}(t) - k_m^-\cdot M(t). \tag{2.34}$$

IV $\text{Im} \xrightarrow{k_{hm}^+} \text{Hm} \xrightarrow{k_m^+} \text{M}$

$$z\cdot\frac{d}{dt}\overline{Hm}(t) = k_{hm}^+\cdot(Im_0+z\cdot\overline{Hm}_0+M_0-z\cdot\overline{Hm}(t)-M(t))-k_m^+\cdot z\cdot\overline{Hm}(t), \tag{2.35}$$

$$\frac{d}{dt}M(t) = k_m^+\cdot z\cdot\overline{Hm}(t). \tag{2.36}$$

In the case of the GFP model, the quantities of FP are expressed in green RFU and no rescaling factor is needed. To visualize results in the red and green scale on the same plot, the two quantities were normalized to be on the same scale.

2.5.2.2 Parameter estimation and model selection

The models defined above were calibrated using the experimental data of Fig 2.1C for RFP and Fig. 2.5 for GFP. Each model was fitted to the corresponding data and the R^2 for each measured protein species was calculated. Parameters k_{hm}^+, k_{hm}^-, k_m^+, k_m^- along with the initial quantity of Im (Im_0) and the RFU conversion factor z were estimated. The method of least-squares via the Python function `lmfit.minimize` was used to produce the best fit. For

Model	k_{hm}^{+}	k_{hm}^{-}	k_{m}^{+}	k_{m}^{-}	z	Im_0	R_m^2	R_{hm}^2	AIC
RFP I	4e-7	0.006	0.105	8e-10	1.01	7043	0.999	0.933	3528
RFP II	0.020	0.104	0.104	-	1.01	24480	0.999	0.933	3625
RFP III	0.020	-	0.104	7e-12	1.01	24480	0.999	0.933	3625
RFP IV	0.020	-	0.105	-	1.01	24476	0.999	0.913	3524
GFP	-	-	0.080	-	-	5222	0.972	-	401

Table 2.2: Parameter estimates for each model. The method of least-squares was used by minimizing the sum of squares of the residuals to obtain the best fit.

the calibration of the RFP models, the residuals produced from the red fluorescence curve were multiplied with the conversion factor z so that blue and red residuals are in the same scale. The estimated parameters for each models are presented in Table 2.2.

In parallel, for each RFP model variant the Akaike Information Criterion (AIC) was calculated [149]. We selected the model with the lowest AIC which is model IV, the model without backflows. For the selected model the superscript $+$ is dropped in the main text for convenience (*i.e.*, $k_m = k_m^+$ and $k_{hm} = k_{hm}^+$). An *a-posteriori* bootstrapping procedure was also performed to check for identifiability issues in the RFP model variants [150]. 1000 new datasets were created by randomly assigning the residuals obtained from the best fit to the data of Fig. 2.1C. The resulting datasets were used to fit each one of the four RFP maturation models. Fig. 2.17 demonstrates the parameter values estimated from the bootstrap datasets relative to the estimated values from the actual dataset (Table 2.2). The identifiability analysis shows that the parameters for the selected model variant can be determined with high precision from the data, contrary to the other model variants (Fig. 2.17).

Chapter 3

Inference of single-cell resource allocation profiles reveals heterogeneous adaptation of a bacterial population to changes in the environment

"One accurate measurement is worth a thousand expert opinions"

Grace Hopper

3.1 Context

In this chapter, I will present the conducted work on resource allocation of individual cells, along with the detailed results. To study resource allocation dynamically, we focused on the ribosomes and two metabolic proteins. Each protein was tagged with a fluorescent reporter on the chromosome of *E. coli*. I used a strain where the ribosomes were tagged with a GFP, constructed by Nils Giordano, former PhD student in the team. In the same strain I tagged the ArgG protein, implicated in the amino-acid metabolism with a RFP. In parallel, I created a control strain where the two proteins of interest were tagged with a YFP and CFP, respectively. Lastly, the third protein, Mdh, implicated in the energy metabolism, was tagged by Tristan Guillaumin, a M2 student I supervised during my PhD.

Numerous microfluidics experiments were conducted on a automated microscope setup using the three reporter strains, in collaboration with Irina

Mihalcescu et Mathilde Gateau (LIPhy). The obtained images were analyzed using a dedicated pipeline and the fluorescence and length measurements were obtained for each cell. In collaboration with Eugenio Cinquemani (Inria), a new method was developed to estimate the growth rate of individual cells from length measurements. With EC, the Kalman smoothing method developed in the context of the previous chapter was adapted to single-cell data. The dynamical model was also adapted so the resource allocation profiles can be reconstructed. The information on maturation of reporters we acquired in the previous chapter, along with the derived models, were used in this context to correct for the maturation of each of the four reporters used.

The experiments were conducted in the context of upshifts and downshifts from acetate to glucose, and back. Here, we will focus on the adaptation of individual bacteria during a nutrient upshift. As I will explain later, we did not have confidence in the estimates for the downshift so we did not further investigate the obtained dynamics. We also focused on the ribosomes since they control the cellular machinery and then we compared the obtained responses for the ribosomes with the two metabolic proteins, so as to have a more global overview of the adaptation.

The results that will be presented gave us new insights on bacterial adaptation in changing environments but also challenged some fundamental assumptions we make in the field. This chapter will be soon submitted for publication in a biological journal. The code in which we implemented all the methods will become available at the time of publication.

3.2 Introduction

Microbial growth involves the conversion of nutrients from the environment into biomass. The main component of biomass are proteins, which also play a major role in the synthesis of new biomass by functioning as enzymes in metabolism and by constituting the macromolecular machines of RNA polymerases and ribosomes responsible for gene expression. Microbial growth thus requires the coordinated investment of cellular resources in different categories of proteins. The basic principles of this coordination have been described by

so-called growth laws [25, 28, 31, 33]. Growth laws relate the growth rate of the cells to the (relative) abundance of different categories of proteins. In other words, they quantitatively describe the resource allocation strategies adopted by microbial cells.

The classical growth law advances a linear relation between growth rate and the (mass) fraction or concentration of proteins that are ribosomal proteins [33]. Ribosomes are probably the most important protein category for two reasons. First, they are responsible for the synthesis of all proteins in the cell. Second, they are themselves very costly to make: ribosomes constitute up to 25% of the total protein mass in *Escherichia coli* [36]. The ribosomal growth law has been shown to hold over a large range of growth rates and for a wide variety of microorganisms [33], and has been generalized to cases of reduced translation rates by antibiotic treatment [33] or different temperatures [151]. Simple mathematical models explain the growth law from basic hypotheses on cellular resource allocation under the assumption that cells have evolved to maximize their growth rate [33, 115, 119, 151, 152].

The growth law has been determined for populations of microorganisms in a state of balanced growth. In their natural environment, however, microorganisms are rarely in a state of balanced growth and have to cope with regular changes in nutrient availability [44]. Moreover, a large body of work has shown that individual cells within a microbial population do not behave identically, but display a range of phenotypic variability [54–56, 62, 64, 68].

Theoretical models have predicted how microorganisms dynamically adapt the synthesis of ribosomes during shifts from balanced growth on a poor carbon source to balanced growth on a rich carbon source (upshift) or the other way around (downshift) [118, 119, 153]. A recent proteomics study measured changes in the (relative) abundance of ribosomes and other proteins during such upshifts and downshifts [118]. However, the quantification of resource allocation on the single-cell level has not been attempted thus far.

In the work reported here, we addressed this question for upshift experiments from growth in a minimal medium with acetate to the same medium with glucose, in the case of the model organism *E. coli*. In particular, we were interested in quantifying dynamic resource allocation strategies for ribo-

somes, defined as the fraction of the total protein synthesis rate allocated to the synthesis of ribosomal proteins during the upshift, and in how these strategies vary over the individual cells of an isogenic population. To this end, we constructed appropriate chromosomal reporter systems to monitor expression changes of ribosomes. Moreover, we developed a method for the statistical inference of dynamic resource allocation strategies from single-cell, time-course fluorescence data acquired over extended periods of time (> 80 generations), corresponding to several upshifts from acetate to glucose.

We observed a complex adaptation trajectory on the population level, in which cells change their resource allocation strategy and their growth rate in a partly coupled and partly uncoupled manner. When analyzing this response on the level of individual cells, we found that it arises from lumping together (i) cells with a coupled response, where the allocation of resources to ribosome synthesis and the growth rate jointly and progressively increase until they reach their steady-state levels corresponding to balanced growth on glucose (slow adapters), and (ii) cells with an uncoupled response, where resource allocation immediately jumps to its steady-state level, followed by a gradual increase of the growth rate (fast adapters). The choice between coupled and uncoupled adaptation trajectories was found to be correlated with the pre-shift ribosome concentration. We also identified interesting similarities and differences, on the single-cell level, between the resource allocation strategies of ribosomes and representative enzymes in amino acid and energy metabolism.

In this study, we quantified for the first time, using a combination of fluorescent reporter genes and powerful statistical inference algorithms, dynamic resource allocation strategies of bacteria on the single-cell level. We notably found that individual *E. coli* cells, in an isogenic population under identical conditions, display very different trajectories of adapting the concentration of ribosomes and the growth rate during an upshift from a poor to a rich carbon source. The identification of slow and fast adapters in a population stepping up its growth rate after a sudden influx of rich nutrients does not only provide novel insights in the fundamental question of how bacteria respond to changes in their environment, but it is also potentially important for biomedical and biotechnological applications.

3.3 Results

3.3.1 Single-cell experiments to monitor changes in ribosome concentrations

In order to quantify the ribosome abundance in single *E. coli* cells, we tagged one of the ribosomal subunits with a green fluorescent protein (GFP). In particular, following previous work [81], we constructed on the chromosome of the *E. coli* strain BW25113 [154] a translational fusion of the *rpsB* gene, protein of the smal ribosomal subunit with the gene encoding the fast-folding GFPmut2 (*Materials and Methods*). The resulting strain, expressing the fusion protein RpsB::GFPmut2, was called RA. In order to be able to check for any specific effects of the chosen reporter, we also constructed a second strain (RA*) in which the same ribosomal subunit was tagged with a yellow fluorescent protein (YFP) instead of a GFP (*Materials and Methods*).

We verified for the two reporter strains, in batch experiments using a variety of different media, that the growth rates of the population are not significantly different from the corresponding growth rates of the wild-type strain (Fig. 3.9). Furthermore, we quantified in these experiments the fluorescence emitted by the cultures during steady-state exponential growth and normalized it by the optical density, yielding a quantity proportional to the ribosome concentration on the population level. When plotting this quantity as a function of the growth rate, for each of the different conditions, we found a linear relation in agreement with the well-known growth law for ribosomes ([33] and Fig. 3.8). These control experiments demonstrate that the constructed reporter strains provide a reliable read-out of ribosome concentrations over a broad range of growth conditions.

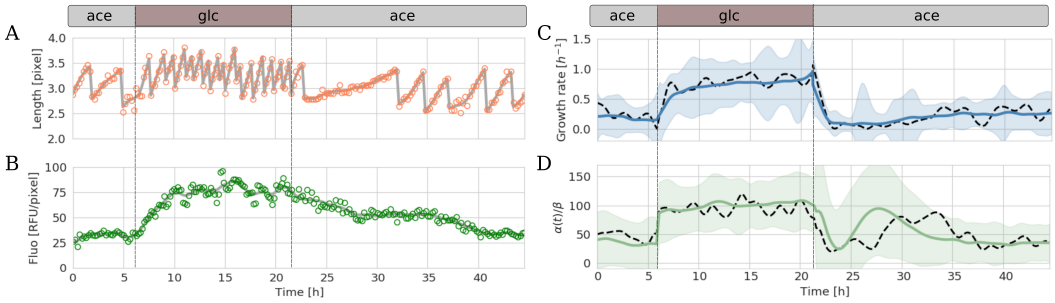


Figure 3.1: Measured and estimated quantities in mother machine experiments with a ribosomal reporter strain. A-B. Cell length (orange dots) and green fluorescence intensity (green dots) were quantified over time in a mother machine tracking individual bacteria of strain RA, carrying a fusion of the ribosomal subunit S2 and the green fluorescent protein GFPmut2. The experiment consisted in several consecutive upshifts and downshifts (vertical dashed lines) between minimal media with glucose or acetate. **C-D.** Cell length measurements were used to estimate growth rates and fluorescence intensity measurements to estimate resource allocation strategies, using two statistical inference methods (*Materials and methods*). The gray solid curves in panels A and B represent the fits of the single-cell data obtained from the two methods, whereas the black dashed curves in panels C and D represent the corresponding estimates of the growth rate $\mu(t)$ and the resource allocation strategy $\alpha(t)/\beta$ for this same (mother) cell. Blue, green and red solid curves represent the mean of the estimates over all cells considered in the experiment and confidence intervals are given as two times the standard deviation.

In order to quantify ribosome concentrations in single cells, we used an experimental setup in which cells of the RA or RA* reporter strain are grown in a mother machine and monitored using time-lapse fluorescence microscopy (*Materials and Methods*). Mother machines are microfluidics devices allowing cells trapped at the bottom of dead-end microchannels, so-called mother cells, to be followed over hundreds of generations in well-controlled conditions of growth [82]. We submitted cells in the mother machine to upshifts from a

poor growth medium (minimal medium with acetate) to a rich growth medium (minimal medium with glucose). After each upshift, the bacteria were left to grow for a sufficient number of generations (> 10) to reach a state of balanced growth in the new medium. Each experiment consisted of at least two upshifts, separated by corresponding downshifts from glucose to acetate (Fig. 3.10 and Tab. 3.2). During these experiments, we acquired phase contrast and green (yellow) fluorescence images at regular time-points (*Materials and Methods*). Using existing image analysis software [155], we segmented the (mother) cells and recorded for each cell at each measurement time the cell length and (average) fluorescence intensity (Fig. 3.11).

This experimental setup allowed us to monitor the accumulation of ribosomes in single cells over long periods of time with high sampling density. As illustrated in Fig. 3.1 for the RA strain, the experiments reproduce a number of well-known features of growth transitions in *E. coli*. After the upshift from acetate to glucose, the cells are seen to divide faster, corresponding to the higher growth rate on glucose, the preferred carbon source of *E. coli*, and attain a longer cell length before division [62]. The fluorescence intensities increase after the upshift and stabilize at a higher level, reflecting that the higher growth rate on glucose comes with a higher ribosome concentration [33]. Following the downshift, a lag phase is observed before the cells resume growth on acetate [156]. The observations for the second reporter strain RA* are comparable with those made for RA (Fig. 3.22).

3.3.2 Inference of dynamic resource allocation strategies

We exploited the data obtained from mother machine experiments with reporter strains to quantify (i) resources allocated to the production of ribosomes and (ii) changes in resource allocation after a nutrient upshift. To this end, we first formulated a simple model describing the dynamics of the ribosome concentration in a mother cell. The model is a variant of previously published models [119, 152] and derived from basic assumptions in Text S1. The model assumes that the ribosome concentration evolves continuously across cell divisions, which is a reasonable approximation given the high number of ribosomes

in *E. coli* cells growing on glucose or acetate (8000, 15000 [74, 157]). The dynamics of the ribosome concentration r as a function of time t is defined as a function of the time-varying growth rate $\mu(t)$, the protein degradation rate γ, the total protein concentration $1/\beta$, and the resource allocation strategy $\alpha(t)$. The total protein concentration is assumed constant, which is a good approximation for the conditions considered here [158], and is used as a proxy for the total biomass concentration in the cell.

$$\frac{d}{dt}r(t) = (\mu(t) + \gamma)\,(\frac{\alpha(t)}{\beta} - r(t)). \tag{3.1}$$

The model expresses that the change in ribosome concentration arises from an (im)balance between the rate of decay of ribosomes due to growth dilution and degradation ($(\mu(t) + \gamma)\, r(t)$) and the synthesis rate of new ribosomes as a fraction $\alpha(t)$ of the total protein synthesis rate $(\mu(t) + \gamma)/\beta$. In other words, resource allocation corresponds to the distribution of the total protein synthesis rate, the capacity to make new proteins, over different protein categories and $\alpha(t)$ is the fraction allocated to ribosomes.

How do the data provided by the mother machine experiments, that is, the time-lapse measurements of the length and (average) fluorescence intensities of the cell, relate to the above model? In brief, the cell length measurements allow us to infer an estimate of the time-varying growth rate $\mu(t)$, in units of 1/h, while the fluorescence intensities report on the time-varying ribosome concentration, expressed in relative fluorescence units (RFU) per pixel (RFU/pixel). Together with the measured degradation constant γ of the fusion protein, obtained in this and previous work ([159] and *Materials and methods*), this allows us to infer an estimate of $\alpha(t)/\beta$, in units of RFU/pixel, from the data. While the absolute value of this ratio is difficult to interpret, because the value of $1/\beta$ is not precisely known for our strain, relative changes in $\alpha(t)/\beta$ following nutrient upshifts will inform us about relative changes in $\alpha(t)$.

In order to estimate single-cell growth rates from cell length data, we developed a dedicated regularized least-squares method based on work from statistical signal processing ([124], *Materials and methods* and Text S2). Contrary to most estimation methods applied to microfluidics data, the method does

not assume that growth rates are constant between two cell divisions. This is critical for monitoring medium shifts, when bacteria change their growth rate dramatically within one generation (Fig. 3.17). The method uses regularization to cope with measurement noise in the data by penalizing rapid changes in growth rate. The regularization parameter was determined from the data by means of cross validation. Fig. 3.1A and C illustrate the estimation of the growth rate of a typical mother cell across an upshift and a downshift, and the corresponding fit to the length data. The mean growth rate over all mother cells considered in the experiment increases from 0.27 ± 0.02 h^{-1} on acetate to 0.79 ± 0.03 h^{-1} on glucose, in agreement with values measured in batch for the BW25113 wild-type strain [160].

For the estimation of single-cell resource allocation strategies, we developed another statistical inference method. We formulated the estimation of $\alpha(t)/\beta$ from the inferred growth-rate and the fluorescence measurements as a Bayesian estimation problem and solved this problem by means of Kalman smoothing [133, 134, 159]. The overall information flow of the procedure is summarized in Fig. 3.2 and more details can be found in the *Materials and methods* and Text S2. The method uses a variant of the model of Eq. 3.1, which accounts for maturation of the fluorescent reporter. Ignoring maturation may distort the inference of dynamic resource allocation strategies, especially outside balanced growth, after a nutrient upshift or downshift [159]. The maturation constants for the reporters used in this study were determined in targeted experiments, along with estimates of the degradation constants (*Materials and methods* and Tab. 3.3). Similar to the estimation of growth rates, the estimation of resource allocation strategies requires regularization parameters that were determined from the noise properties of the fluorescence intensity measurements for each experiment. As a side product, the inference method also returns an estimate of the total protein concentration, that is, the sum of the concentrations of the mature (observed) and immature (non-observed) protein.

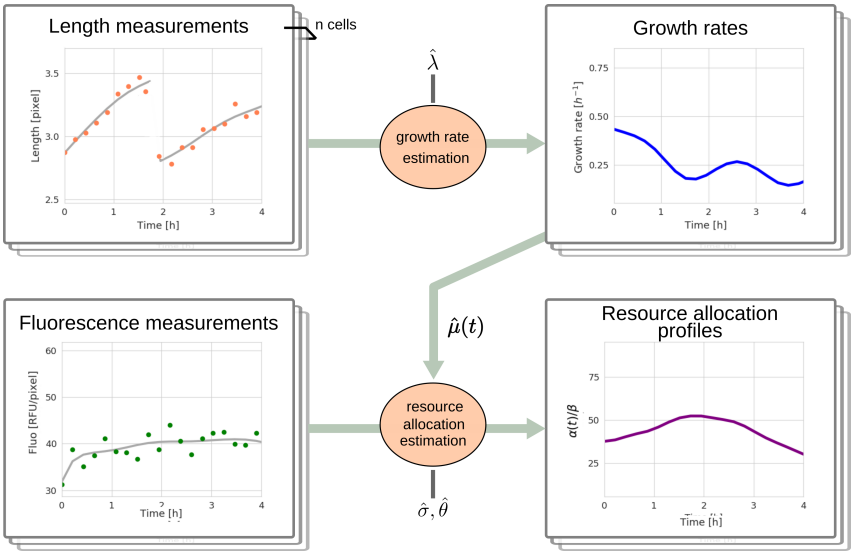


Figure 3.2: Inference procedures for estimating growth rates and resource allocation strategies from single-cell data. Time-lapse measurements of the length of mother cells are used as input by the growth-rate estimation method, returning estimates of the time-varying growth rate $\mu(t)$ that may vary within a generation. The growth rate estimates, along with time-lapse fluorescence intensity measurements, corresponding to intracellular concentrations of proteins tagged by a fluorescent reporter, are then used for the estimation of single-cell resource allocation strategies $\alpha(t)/\beta$. The two estimation methods rely on regularization to cope with measurement noise and provide smooth estimates, which requires values for the regularization parameters ($\hat{\lambda}$ and $\hat{\sigma}$, $\hat{\theta}$) which are estimated for each experiment. Details on the methods and the regularization parameters can be found in the *Materials and Methods* and in Text S2.

Fig. 3.1D shows the resource allocation strategy estimated for a single cell and Fig. 3.1C the corresponding fit to the fluorescence data for this cell. The figure also contains the mean resource allocation strategy estimated for all cells in the experiment. We observe that $\alpha(t)/\beta$ immediately increases

after the upshift from acetate to glucose and reaches a new, higher steady-state levels quite fast. The higher value of $\alpha(t)/\beta$ after the upshift means that a larger fraction of the protein synthesis rate is allocated to the ribosomes during fast growth on glucose, consistent with the observation that a higher growth rate requires a higher ribosome concentration. For the downshift, the dynamics of adaptation to acetate are slower and more complex, involving a transient undershoot lasting approximately five hours.

3.3.3 High variability of ribosomal resource allocation during balanced growth

During balanced growth, we expect the ribosome concentration to reach steady state, which implies $r(t) = \alpha(t)/\beta$ (Eq. 3.1). In other words, the steady-state ribosome concentration must correspond to the steady-state fraction of the constant total protein concentration. In order to verify this, we computed during balanced growth on glucose and acetate (more than 5 generations after an upshift or downshift), for every individual mother cell, the mean ribosome concentration and the mean of the resource allocation strategy over every generation. For the ribosome concentration, we used the total concentration estimated from the data instead of the directly observed fluorescence intensities. The results, consisting of more than 800 data points for individual generations of individual cells are shown in the scatter plot of Fig. 3.3A. As expected, the two quantities are correlated ($R^2 = 0.81$), confirming the direct relation between ribosome concentration and resource allocation at steady state.

The steady-state data also provide another straightforward test. The growth law for ribosomes, the linear increase of the ribosome concentration with the growth rate, has been extensively studied for populations of bacteria, grown in different media. Nothing is known, however, about the correlation of the ribosome concentration with the growth rate in individual cells grown in the same medium. While we might expect a linear relation in this case as well, the results of the test of another population-average growth law on the single-cell level show that some caution is needed. Taheri-Araghi *et al.* [62] found strong deviations from the expected exponential dependence of newborn cell

volume on growth rate for individual cells grown in the same medium. We first verified that the average of the growth rate and the ribosome concentration taken over individual generations of individual cells growing on acetate or glucose is consistent with the growth law established from population-level measurements (Fig. 3.3B). We then plotted the growth rate and the ribosome concentration for individual generations of individual cells in each of the two media (Fig. 3.3C). As can be seen, there is a weak positive dependence of ribosome concentration on growth rate for glucose and no positive dependence in the case of acetate. In both cases, there is a huge variability in growth rates and ribosome concentrations, such that individual cells can grow fast with low ribosome concentrations or slow with high ribosome concentrations.

How can this deviation from the population-average growth law be explained? In the framework of the model, the sum of the growth rate and the degradation rate is related to the time-varying total protein synthesis rate by $\mu(t) + \gamma = \beta\, v_{prot}(t)$ (Text S1). That is, for the total protein concentration to remain constant, the rate of newly synthesized proteins $v_{prot}(t)$ must match, at every time-point t, the sum of the rate of proteins decay by growth dilution or degradation ($(\mu(t) + \gamma)\,(1/\beta)$). Bearing in mind that the protein synthesis rate equals the product of ribosome concentration and time-varying ribosome activity $\sigma(t)$ [161], *i.e.* $v_{prot}(t) = r(t)\,\sigma(t)$, we find

$$r(t) = \frac{\mu(t) + \gamma}{\beta\,\sigma(t)}. \tag{3.2}$$

This relation provides a possible explanation for the observed lack of correlation between ribosome concentration and growth rate at steady state, by suggesting that there is a high variability of ribosome activity over individual cells and across generations. This variability of activity across single cells may be due to active regulation of ribosome availability or to passive regulation by the size of amino acid and ATP pools.

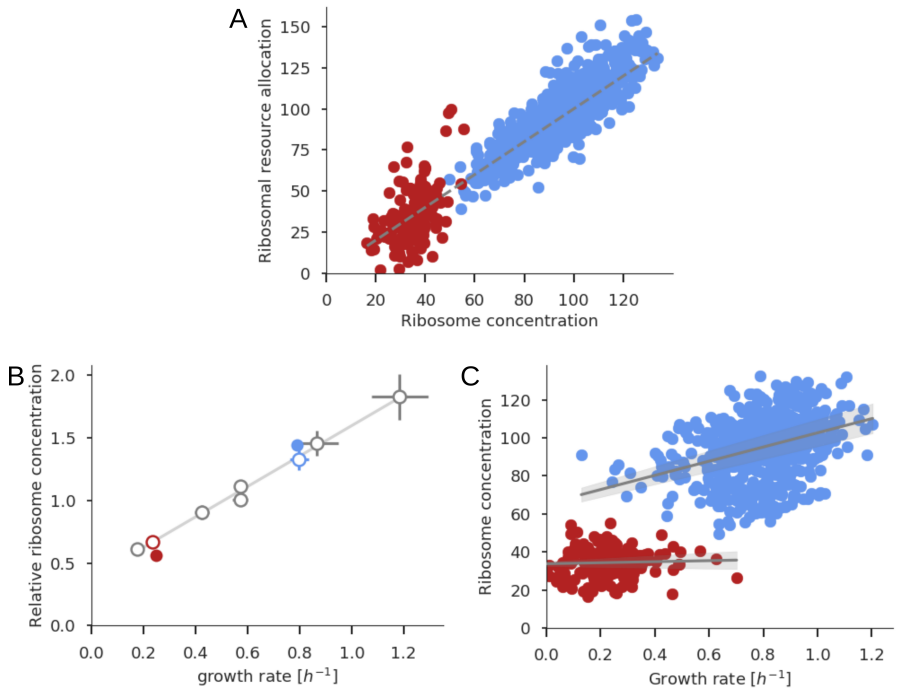


Figure 3.3: Single-cell resource allocation for ribosomes at steady state. A. Relation between ribosome concentration r and resource allocation strategy α/β for individual generations and individual mother cells during balanced growth on acetate (red) or on glucose (blue) in a mother machine. The experiment was carried out by means of the RA strain (Tab. 3.2). As expected by the model of Eq. 3.1, the two quantities are strongly correlated ($R^2 = 0.81$). **B.** Relation between the population average of the growth rates and total ribosome concentrations for the single cells growing on glucose or acetate in panel A (filled circles) and population-level measurements for the RA strain growing in batch on glucose, acetate, or other carbon sources (open circles) (Fig. 3.8). Data points for acetate are colored in red, for glucose in blue, and for other carbon sources in gray. In order to make the two data sets comparable, all ribosome concentrations were normalized by the mean of the (population average) of the ribosome concentrations on glucose and acetate. A line was fitted to the batch data and plotted as a visual aid. Confidence intervals are given by two times the standard error of the mean. The growth rates and the relative difference of the ribosome concentrations during growth on glucose and acetate are comparable between the two datasets. **C.** Relation between ribosome concentration and growth rate for individual generations of individual cells during balanced growth on glucose (blue) and acetate (red). A line was fitted for the two datasets and plotted in gray, along with the confidence bands that represent two times the standard deviation. We observe a weak dependence of the ribosome concentration and growth rate on the single-cell level (37.2 ± 3.9 for glucose and 2.7 ± 4.9 for acetate), contrary to the population-average growth law for ribosomes shown in panel B.

3.3.4 Heterogeneous adaptation dynamics of ribosomes after nutrient upshifts

The growth law for ribosomes states that, during balanced growth, the ribosome concentration increases linearly with the growth rate (Fig. 3.3A). Since at steady state $r = \alpha/\beta$, this linear dependency also holds for the resource allocation strategy. How do *E. coli* cells adapt their growth rate and resource allocation strategy following an upshift from acetate to glucose? One hypothesis would be that they gradually increase both the resources allocated to ribosomes and the growth rate, by following the linear relation of the growth law. Another hypothesis, proposed by Bremer and Dennis [153], is that after an upshift cells immediately set resource allocation to the steady-state value on glucose and then gradually adapt their growth rate. A recent kinetic model by Erickson *et al.* [118], which includes feedback regulation of α by the translational activity σ, made a very similar prediction. Yet another hypothesis is that the adaptation trajectory maximizes biomass accumulation, which was predicted to give rise to an oscillatory pattern for α [119]. The resource allocation strategies and growth rates estimated from the mother machine data can be exploited to reconstruct the adaptation trajectories after an upshift and test the above hypotheses.

We first computed the population average of the estimated adaptation dynamics of μ and α/β for an acetate-glucose upshift and plotted the trajectory in the $\mu - \alpha/\beta$ plane (Fig. 3.4A). As can be seen, the trajectory has two consecutive phases. In the first phrase, after the upshift, resource allocation jumps to a high value, which then remains approximately constant while growth rate gradually increases. In the second phase, μ and α/β jointly increase towards the steady-state value for balanced growth on glucose along the line of the growth law. The corresponding time-courses for μ and α/β are shown in Fig. 3.4D,G. A sharp increase in resource allocation occurs immediately after the upshift, but it takes more than 6 hours to reach the steady-state value for balanced growth on glucose. For the growth rate, the increase towards the steady-state value is more gradual over the whole adaptation period. We conclude that the adaptation trajectory after an upshift consists of phases of

uncoupled and coupled increase of resource allocation and growth rate, thus combining the second and first hypothesis, respectively.

This relation provides a possible explanation for the observed lack of correlation between ribosome concentration and growth rate at steady state, by suggesting that there is a high variability of ribosome activity over individual cells and across generations. This variability of activity across single cells may be due to active regulation of ribosome availability or to passive regulation by the size of amino acid and ATP pools.

We performed the same analysis for a glucose-acetate downshift in the same experiment. During the transition from glucose to acetate, we observed a complex pattern of adaptation, involving an overall decrease of μ and α/β towards their steady-state values for balanced growth on acetate, but also a transient dip in both growth rate and resource allocation reaching its minimum after 3-4 h (Fig. 3.4C,F,I). The undershoot of the growth rate corresponds to a lag phase that was observed in previous work [156]. The sharp drop in resource allocation directly after the downshift, however, arises from a strong decrease in fluorescence intensity (Fig. 3.1B) that is difficult to explain in terms of the biological processes included in the model. In particular, we found the ribosomal fusion protein to be very stable, with a half-life on the order of 37 h (Tab. 3.3), which is incompatible with the observed decrease in fluorescence intensity at a time when growth dilution is minimal because the cells are in lag phase. One possible explanation is a technical problem due to photobleaching. We established that the role of photobleaching is negligible during balanced growth on glucose and acetate (Fig. 3.15) but this does not exclude that it may lead to artifacts during a lag phase lasting almost 10 h, with minimal synthesis of new ribosomes. In the face of this uncertainty, we decided to restrict the analysis of the data to acetate-glucose upshifts.

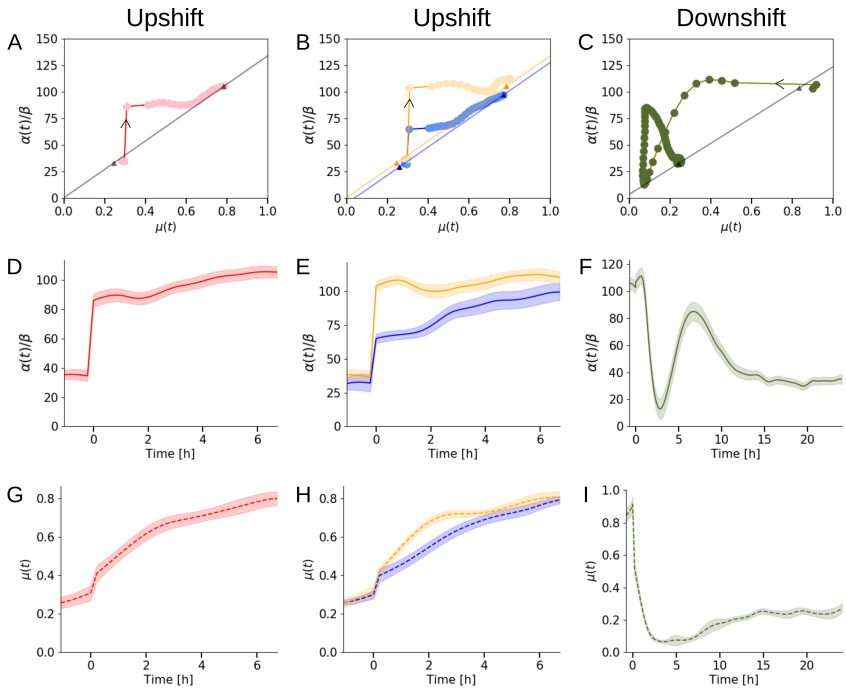


Figure 3.4: Adaptation dynamics of growth rate and ribosomal resource allocation after nutrient upshifts and downshifts. A. Population-averaged adaptation trajectory of the resource allocation strategy for ribosomes α/β and the growth rate μ after an acetate-glucose upshift applied to the RA strain growing in a mother machine. The resource allocation strategies and growth rates were estimated from the data using the inference methods of Fig. 3.2 and averaged over 127 cells. The arrows indicate increasing time after the upshift. The triangles denote the average values of α/β and μ during balanced growth on acetate (before the upshift) and glucose (after the upshift). The latter values were determined by computing for each cell the mean growth rate over a period of balanced growth (> 2 h), and then averaging these values over the individual cells. The black line through the population average before and after the upshift are shown as a visual aid. The trajectories show that the adaptation of resource allocation and growth rate are uncoupled in a first phase and coupled in a second. **B.** Clustering of the single-cell resource allocation strategies for ribosomes after the upshift using K-means (*Materials and Methods* and Fig. 3.27) reveals two distinct trajectories for fast adapters (orange, 70 cells) and slow adapters (blue, 57 cells). The orange and blue straight lines connect the balanced growth values (triangles) of the two populaitions before and after the shift. **C.** As in panel A, but for a glucose-acetate downshift carried out in the same experiment (125 cells). The population-averaged adaptation trajectory is complex with an undershoot that may be an artifact of the experimental conditions. **D-F.** Time-courses of the resource allocation strategies in panels A-C, respectively. Confidence intervals are given by two times the standard error of the mean. **G-I.** Time-courses of the growth rates in panels A-C.

Fig. 3.3B revealed an important cell-to-cell variability in resource allocation strategies and growth rates during balanced growth on glucose and acetate. This motivates an exploration of the variability underlying the population-averaged adaptation trajectory in Fig. 3.4A. We performed a clustering analysis to classify the resource allocation strategies of the individual cells in the first hours after the upshift (Fig. 3.26). We found that grouping the data into two clusters provided a statistically sound and biologically informative view of the variability in the data (*Materials and Methods* and Fig. 3.27). The adaptation trajectories for these two clusters are plotted in Fig. 3.4B. Interestingly, the two-phase trajectory of the population average turns out to be the superposition of the trajectories for the individual clusters. A first cluster of fast adapters immediately jumps to the resource allocation value for balanced growth on glucose and approximately remains at this value when increasing the growth rate (uncoupled adaptation of μ and α/β). A second cluster of slow adapters mostly increases growth rate and resource allocation simultaneously, along the line of the growth law, after an initial small jump in resource allocation (coupled adaptation of μ and α/β). The time-courses of the growth rates and resource allocation strategies for the two clusters confirm these observations (Fig. 3.4E,H). Note that the two clusters eventually reach the same growth rate, but that the cluster of fast adapters does so more rapidly than the cluster of slow adapters.

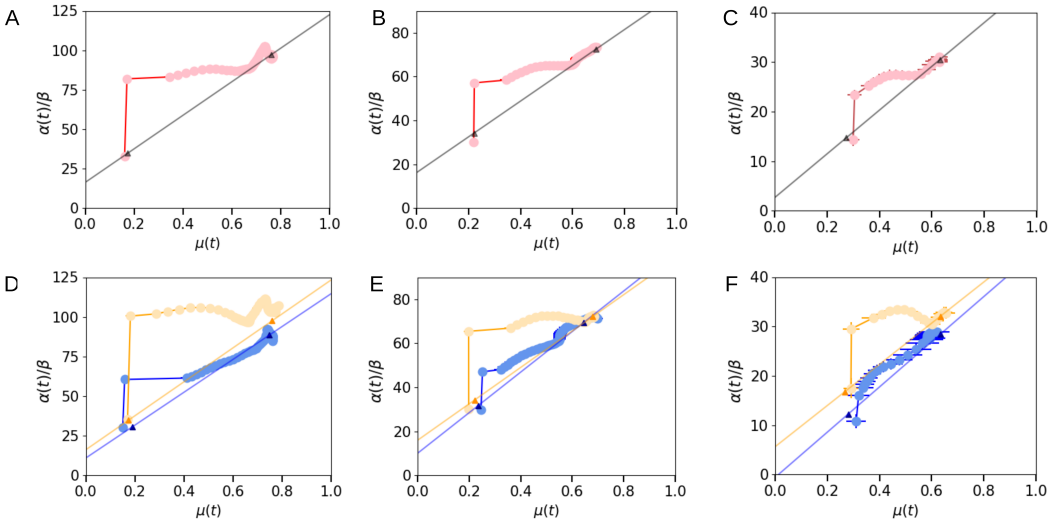


Figure 3.5: Reproducibility of the adaptation dynamics of growth rates and ribosomal resource allocation strategies. The analysis of the single-cell adaptation dynamics in Fig. 3.4 were repeated in three conditions. **A.** Population-averaged adaptation trajectory for a second acetate-glucose upshift from the same experiment (134 cells). **B.** As in panel A, but for an upshift in a replicate experiment in identical conditions (55 cells). **C.** As in panel A, but for an upshift in an experiment with the second strain RA* carrying the fusion protein RpsB::YFP (76 cells). A clustering analysis revealed two clusters for the replicate experiments as well (Fig. 3.27). **D.-F.** Adaptation trajectories for the individual clusters in the three experiments, corresponding to panels A-C. The time-courses for growth rates and ribosomal resource allocation strategies are shown in Figs 3.23-3.25. The plots follow the same graphical conventions as in Fig. 3.4.

This analysis of resource allocation strategies on the single-cell level is highly reproducible across experiments: the same patterns have been observed in a second upshift from the same experiment (Fig. 3.5A,D), in a replicate experiment in identical conditions (Fig. 3.5B,E), and in an experiment with

the second strain RA* carrying the fusion protein RpsB::YFP (Figs 3.5C,F). In two of the three replicates, the fast adapters reach their steady-state growth on glucose before the slow adapters (Figs 3.23-3.25).

What causes the population to divide in fast and slow adapters? A possible clue can be obtained by considering the allocation of resources to the ribosomes before the upshift (Fig. 3.4E). In this experiment and in the replicate experiments, the mean value of α/β is higher in the cluster of fast adapters than in the cluster of slow adapters, and by the steady-state relation $r = \alpha/\beta$ this also holds for the mean value of the ribosome concentration. That is, a higher ribosome concentration, and therefore potentially a higher ribosome reserve [161] before the upshift, may favor a rapid reallocation of resources to ribosomes when the medium is switched from a poor to a rich carbon source.

3.3.5 Correlation of adaptation dynamics of ribosomes and metabolic enzymes

The population-level analysis of resource allocation during balanced growth in different environments, and during the transition from one environment to another, has revealed that the investment of cellular resources in different protein categories is strongly correlated. For example, during growth in minimal medium, where *E. coli* cells have to produce the metabolic precursors necessary for protein synthesis, the concentration of enzymes in amino acid metabolism increases with the growth rate, in step with the concentration of ribosomes [118, 157]. On the other hand, the concentration of enzymes in energy metabolism, producing ATP and other energy co-factors, decreases with the growth rate [36, 39]. This reflects a strategy following which, at higher growth rates, costly but efficient respiration pathways are supplemented by cheaper but less efficient fermentation pathways [115]. These observations raise the question how single-cell resource allocation strategies for metabolic enzymes adapt after an acetate-glucose upshift, and how they compare with the observed adaptation dynamics for ribosomes in Fig. 3.4.

We therefore constructed, in the same way as for ribosomes, fluorescent reporter systems for an enzyme in amino acid metabolism, argininosuccinate

synthase (ArgG), involved in the biosynthesis of arginine [160, 162], and for an enzyme in energy metabolism, malate dehydrogenase (Mdh), reducing oxaloacetate to generate malate in the TCA cycle [163] (*Materials and methods*). The two enzymes were tagged with the red fluorescent protein (RFP) mScarlet-I, in the same strain which already carried a GFP tag for RpsB. The resulting strains, carrying a double fusion of RpsB::GFPmut2 and either ArgG::mScarlet-I or Mdh::mScarlet-I, are called RA and RE, respectively. The advantage of the double-fusion strains is that they allow the expression of ribosomes and metabolic enzymes to be compared within individual cells. As a further control, we also constructed a variant of the ArgG reporter in which the enzyme is fused with the cyan fluorescent protein (CFP) mCerulean, within the RA* strain carrying the fusion of RpsB with a YFP. Note that the experiments reported in the previous sections already used the RA and RA* strains, but without exploiting the RFP and CFP data. In the same way as for the RA and RA* strains, we checked that the growth rates of the RE strain and the wild-type control strain are the same (Fig. 3.9). Moreover, as a further validation of the strains, we verified that there exists a positive relation between the growth rate and the ArgG reporter concentration, approximated by the fluorescence intensity normalized by the optical density, as well as a negative relation between the growth rate and the Mdh reporter concentration (Fig. 3.8).

Like for the RA and RA* strains, we performed single-cell upshift experiments in a mother machine device with the RE strain, in the conditions of Fig. 3.1 (*Materials and methods*). We measured both green and red fluorescence during the experiments in order to reconstruct the resource allocation strategies for both the ribosomes and the metabolic enzymes in the same cell. The approach of Fig. 3.2 was used to infer single-cell resource allocation strategies as well as growth rates from the data. For the RFP fusions, we used a two-step model of protein maturation for the reconstruction of $\alpha(t)/\beta$ from the fluorescence intensities, instead of a one-step model, to accurately capture the more complex maturation kinetics of red fluorescent proteins ([159], see *Materials and methods* and Text S1). The measured cell lengths and fluorescence intensities quantified in the experiments with the RA, RE and RA* strains,

as well as the estimated growth rates and resource allocation strategies, are shown in Figs 3.20, 3.21, and 3.22, respectively. One direct observation is that the fluorescence intensity of the ArgG reporter and the corresponding value of $\alpha(t)/\beta$ increase after an acetate-glucose upshift. On the contrary, the fluorescence intensity and resource allocation for the Mdh reporter decrease after the upshift. In other words, the data show the expected positive and negative correlations of the accumulation of enzymes in amino acid and energy metabolism on the one hand, and ribosomes on the other.

Like for the ribosome data, we further analyzed the adaptation dynamics of the metabolic enzymes in the α/β - μ plane. In the case of Mdh, the population-averaged adaptation trajectory is surprisingly similar to the trajectory of the ribosomes, but in the opposite direction (Fig. 3.6A). That is, the immediate increase of the allocation of protein synthesis capacity to ribosomes after the upshift is accompanied by a parallel decrease in allocation of protein synthesis capacity to this enzyme in energy metabolism. This reflects that, the total protein synthesis capacity being limited, an increase of the available fraction for one protein category needs to be compensated for by a decrease of the fraction for another protein category. On a more detailed level, the adaptation trajectory of Mdh also differs from that of the ribosomes, in that a second, coupled adaptation phase follows the first, uncoupled phase. That is, the initial downward jump in α/β is enough to reach the steady-state resource allocation level for balanced growth on glucose. In the case of ArgG, on the contrary, the first uncoupled adaptation phase is entirely absent and the population-averaged trajectory only shows coupled adaptation of growth rate and resource allocation (Fig. 3.6B). These surprising differences suggest that, even though the expression levels of enzymes in amino acid metabolism and ribosomes are highly coordinated during balanced growth, they may be subject to different regulatory mechanisms responsible for their dynamic adaptation.

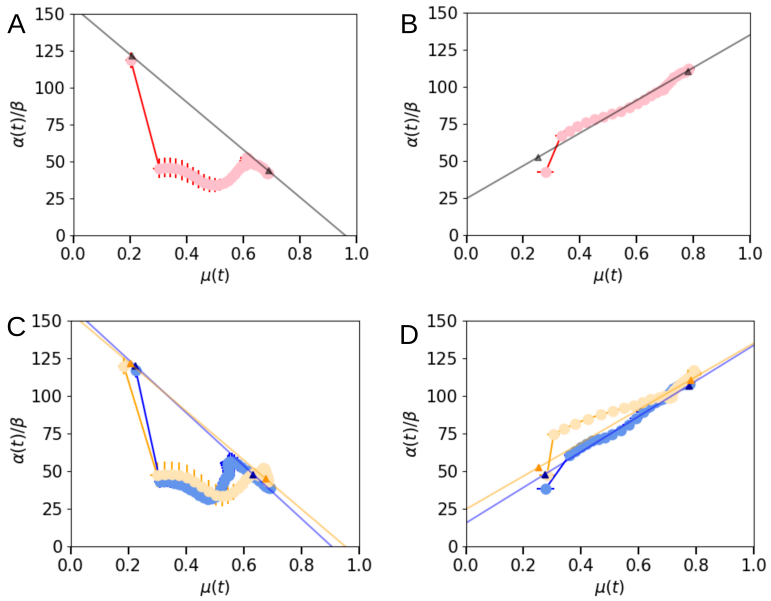


Figure 3.6: Adaptation dynamics of growth rate and metabolic resource allocation after nutrient upshifts. A. Population-averaged adaptation trajectory of the resource allocation strategy α/β for the enyzme Mdh in energy metabolism and the growth rate μ after an acetate-glucose upshift applied to the RE strain growing in a mother machine. The resource allocation strategies and growth rates were estimated from the data using the inference methods of Fig. 3.2 and averaged over 125 cells. The trajectory reveals uncoupled adaptation of resource allocation and growth rate, similar but in the opposite direction as for the ribosomes in Fig. 3.4A, but without a second coupled adaptation phase. **B.** Idem, but for the enzyme ArgG in amino acid metabolism, using the RA strain. The adaptation trajectory shows fully coupled adaptation of resource allocation and growth rate, contrary to what was observed for the ribosomes in Fig. 3.4A. **C.** The same data as in panel A, but for the clusters of fast and slow adaptors identified from the ribosomal data in Fig. 3.4C. There are no clear differences between the adaptation trajectories of the two clusters. **D.** Idem, but for the data in panel B. The plots follow the same graphical conventions as in Fig. 3.4.

We also verified whether the clusters of slow and fast adapters identified for the ribosomes (Fig. 3.4B) can be distinguished in the data for the metabolic enzymes. This is not evident, as shown in Fig. 3.6C-D. The adaptation trajectories of Mdh are very similar for the two clusters, with the same, initial downward jump in the value of α/β. In the case of ArgG, the adaptation dynamics is also qualitatively the same for the two clusters, but there is an interesting quantitative difference. The cluster of fast adapters consistently allocates more resources to this enzyme involved in the production of arginine, as would be expected for the higher protein synthesis demands in fast adapters.

3.4 Discussion

Growth laws empirically describe the resource allocation strategies adopted by microbial cells [25, 28, 33, 69]. These relationships have been defined for many species and conditions, yet their scope has been limited to steady-state growth. Even though a few studies have focused on the subject, we lack experimental data on the dynamic adaptation of resource allocation after a sudden change in nutrient availability or other changes in the environment [118, 119, 153]. Moreover, whereas the available data quantify adaptation on the population level, many phenomena have been described in which the individual bacteria exhibit a heterogeneous behavior [54–56, 62, 64, 68]. The question that remains unanswered is therefore how resource allocation strategies vary dynamically in time, across the cells in a population.

In this study, we addressed the above question by quantifying, for the first time, ribosomal resource allocation strategies on the single-cell level. We created fusions of a ribosomal subunit with different fluorescent reporters in *E. coli* and conducted time-lapse microfluidics experiments in which we recorded the dynamic adaptation of the abundance of the tagged ribosomes after an upshift from a poor to a rich carbon source (Fig. 3.1). Moreover, we developed robust statistical inference methods to reconstruct resource allocation strategies from the resulting fluorescence data, using a mathematical model taking into account protein synthesis, maturation, degradation, and growth dilution (Fig. 3.2). The methods are generic and applicable to the inference of resource

allocation strategies in other microorganisms and experimental scenarios.

We first explored the relationship of the growth rate and the allocation of resources to ribosomes during balanced growth on glucose or acetate (Fig. 3.3). Whereas the population-averaged data agree with the well-established growth law for ribosomes, which posits a linear dependence of the ribosome concentration on the growth rate across different media, we discovered that this relationship is weak or nonexistent in the case of individual cells growing in the same medium. The lack of correlation between growth rate and ribosome concentration indicates that ribosome activity is highly variable across the population of *E. coli* cells. This result is surprising in the light of the hypothesis, exploited in mathematical models [119, 164, 165], that bacteria have evolved to optimize their growth rate. One would expect growth-rate maximizing cells to tightly control their investment in costly ribosomal machinery and avoid wasting resources on inactive protein synthesis capacity. That we do not observe this seems to imply that the adopted resource allocation strategies are suboptimal from the perspective of growth-rate maximization [166]. They could be optimal with respect to another objective though. It has been suggested, for example, that bacterial cells maintain a reserve of ribosomes that can be exploited for rapid adaptation to environmental changes [161, 167].

Next, we analyzed the coordination between changes in growth rate and resource allocation during an upshift from glucose to acetate (Fig. 3.4). When aggregating the responses of the individual cells considered, we observed a complex adaptation trajectory with a partly coupled and partly uncoupled response of resource allocation and growth rate. A closer look at the data revealed that the population-averaged response consists of two different strategies on the single-cell level. Fast adapters immediately increase resource allocation to the steady-state level in the post-shift medium, after which the growth rate gradually reaches its steady-state level as well (uncoupled adaptation). Slow adapters, on the other hand, progressively and jointly increase the allocation of resources to ribosomes and growth rate, following a linear trajectory corresponding to the steady-state growth rate (coupled adaptation). Previous work concluded, on the basis of population-level measurements of the sharp increase of the ribosome synthesis rate directly after the upshift, that *E. coli* cells are

fast adapters [118, 153]. The single-cell measurements of this study reveal, however, that a sizable proportion of the cells are slow adapters, with a lower ribosome synthesis rate (Fig. 3.27). We found no experimental evidence for oscillatory adaptation trajectories predicted under the assumption that bacteria optimize the synthesis of new biomass (proteins) during the upshift [119].

Which regulatory mechanisms underlie the observed variability of ribosome activity before a nutrient upshift and the heterogeneous adaptation dynamics after the upshift? There exist a large body of work on the regulation of the synthesis of ribosomal proteins and ribosomal RNA, involving the nucleotide guanosine pentaphosphate (ppGpp) [24, 168], and the regulation of ribosomal activity by ribosome modulation factor (RMF) and other factors [169, 170]. Our data cannot directly answer this question, but the results of the analysis do pose constraints on possible answers. First, we showed that fast adapter cells have, on average, a slightly higher ribosome concentration before the upshift (Figs 3.4 and 3.5). This suggests that a higher ribosome reserve before the shift allows a higher proportion of ribosomes to be reallocated to the synthesis of new ribosomes after the shift, which is intuitively comprehensible and consistent with other perceived benefits of a proteome reserve in fluctuating environments [161]. Second, we observed that two enzymes, representative for amino acid and energy metabolism, respectively, exhibit resource allocation strategies that are correlated in interesting and sometimes unexpected ways with the strategies for ribosomes. For example, resource allocation for Mdh follows a trajectory opposite to the ribosomal trajectory, whereas the ArgG trajectory corresponds to the ribosomal trajectory of slow adapters but is very different from the trajectory of fast adapters. These observations need to be accounted for when proposing mechanisms for the coordination of the expression of ribosomal proteins and metabolic enzymes.

Why do fast and slow adapters occur in an isogenic bacterial population? The observed heterogeneity in ribosomal resource allocation may be due to unavoidable stochastic fluctuations in bacterial cells and have no ecological relevance. We believe this is unlikely: ribosomes are costly, abundant protein complexes, highly conserved in evolution, and responsible for the synthesis of proteins, the major component of biomass. Another possibility is that the

occurrence of fast and slow adapters confers higher fitness to an *E. coli* population in an environment characterized by rapid changes in nutrient availability. Examples of heterogeneous growth phenotypes abound in microbiology, such as the existence of persister cells, which grow slower but are resistant to antibiotics, and have been found to promote survival of bacterial populations in hostile environments [42, 55]. The present work opens new perspectives on the study of growth heterogeneity, by situating the source of variability on the level of ribosomes, the molecular machines at the center of cellular self-replication.

3.5 Materials and methods

3.5.1 Strain construction

All strains used in this study were derived from the *E. coli* K12 strain BW25113 [154] where the *fhuA* gene was deleted by homologous recombination to confer resistance to phage infections [171].

The expression of the three genes of interest (*rpsB*, *argG*, *mdh*) was studied with the help of translational fusions of a fluorescent reporter protein and the C-terminus of the genes, constructed by homologous recombination mediated by lambda Red [172]. A flexible linker (Fig. 3.7) connects the two proteins. The cloning procedure consisted in first amplifying a PCR fragment carrying 50 bp of sequence identity with the target insertion site. The PCR fragment, called (positive-negative selection) "cassette", contains two divergently transcribed genes: the gene coding for resistance to kanamycin, transcribed from a constitutive promoter, and the gene coding for the toxin CcdB, transcribed from the arabinose-inducible promoter pBAD [173]. This fragment was recombined into the target site using the pSIM plasmid that provides the lambda Red recombination functions [172]. The successful insertion was selected on LB containing glucose and kanamycin. Glucose prevents expression of the CcdB toxin and kanamycin selects for the presence of the cassette on the chromosome.

In a second recombination step, the cassette was removed by recombination of a PCR fragment containing the sequence to be integrated flanked by the

same 50 bp homology regions to the target site. Successful recombination is selected on LB containing 1% arabinose. Arabinose activates the pBAD promoter, and therefore the expression of the toxin CcdB, if the cassette is still present on the chromosome. A successful recombination will have removed the cassette and thus allow growth of the strain in the presence of arabinose. All constructions were verified by sequencing of the modified chromosomal regions. Primers are available upon request.

To visualize the ribosomes, we used a construction similar to the one described by Bakshi *et al.* [81] where the gene coding for the green fluorescent protein GFPmut2 [90] was fused to the C-terminus of *rpsB*, the gene coding for the S2 ribosomal subunit, via a flexible linker of three amino acids (LEI). Using the cloning strategy described above, we introduced a second reporter protein into this strain by fusing the gene coding for the red fluorescent protein (RFP) mScarlet-I [93] to either *argG*, coding for an enzyme involved in the biosynthesis of L-arginine, or *mdh*, coding for malate dehydrogenase, an enzyme of the Krebs cycle. This resulted in two strains where the ribosomes are tagged with GFP and an additional gene is tagged with RFP. We call the strain containing the *argG::mScarlet-I* fusion RA (for **R**ibosomes and **A**mino acids) and the other strain, containing the *mdh::mScarlet-I* fusion, RE (for **R**ibosomes and **E**nergy metabolism).

In parallel, we used the exact same procedure to construct a control strain with different fluorescent reporters. *rpsB* was tagged with the gene coding for yellow fluorescent protein (YFP) mVenus NB [93] and *argG* was fused to the gene coding for cyan fluorescent protein (CFP) mCerulean [93]. We call this strain RA*. All strains used in this study are detailed in Table 3.1 and graphically presented in Figure 3.7.

3.5.2 Growth media

For all experiments, we used M9 minimal medium [174] supplemented with vitamin B1 (5 mg/L) and mineral trace elements. Pre-cultures for the microfluidics experiments were prepared in the above medium supplemented with 2 g/L of acetate and 0.1 g/L of bovine serum albumin (BSA). The medium

used for the microfluidics experiments was supplemented with 2 g/L of acetate or glucose. The experimental conditions of the individual microfluidic experiments are detailed in Tab. 3.2.

For strain validation in batch experiments, minimal medium was supplemented with 2 g/L of one of the following carbon sources: D-glucose, acetate, D-xylose, pyruvate, glycerol, D-fructose, D-galactose and maltose. Furthermore, 2 g/L of casamino acids were added into D-glucose and glycerol in separate conditions. For pad experiments, the medium was supplemented with 2 g/L of acetate before adding agarose.

3.5.3 Batch growth experiments

All strains were streaked from the freezer stock on LA plates several days prior to the experiments. Overnight cultures were grown by inoculating a single colony in liquid minimal medium supplemented with specified carbon sources. Bacterial growth for all experiments occurred at 37°C and liquid cultures were shaken at 200 rpm.

Overnight cultures of the wild-type and reporter strains were diluted into a 96-well microplate at an OD600 of 0.02 where each well contained the same medium as used for the preculture. 2 mm glass beads were added in each well to improve oxygenation. The bacteria were left to grow for up to 24 hours at 37°C in a Tecan Infinite 200M plate reader, while monitoring the absorbance and green (yellow) and red (cyan) fluorescence.

Outliers in the absorbance and fluorescence curves due to the light reflection by the beads were filtered using WellInverter [147]. The fluorescence background was corrected by subtracting the fluorescence emitted by the wild-type strain in the same conditions. The absorbance curves were corrected by subtracting the background of a well containing medium but no bacteria. The growth rate during balanced growth of the population was estimated directly from the absorbance curves, using the method described by Zulkower *et al.* [130]. In order to obtain a proportional estimate of the reporter protein concentration during balanced growth, the corrected fluorescence curves were divided by the corrected absorbance curves.

3.5.4 Microfluidics experiments

Fabrication of microfluidic devices. Several master-molds were created using standard fabrication techniques [82]. In order to obtain PDMS mother machines, we mixed a curing agent and polymer base at a 1:10 ratio (Sylgard Elastomer Kit 184). The mix was poured into the mold and trapped air bubbles were removed in a vacuum chamber. The resulting devices were cured for a minimum of 2 hours at 65°C. Devices were then removed from the master-mold and a 0.75 mm punch was used to create the inlet and the outlet. Before use, the devices were treated with pentane and acetone and were left to dry overnight. A plasma cleaner (PDC-32G Harrick Plasma) operated at high intensity for 40 seconds was used for surface activation of the devices and previously cleaned microscope slides. Each device was placed on a microscope slide and then put at 65°C for 10 minutes. 50 mg/mL of BSA was injected into the sealed devices for surface passivation and the devices were then incubated at 37°C for 1 hour.

Device loading and medium shifts. Precultures were performed in the same way as for the batch growth experiments described above. The optical density of strains growing in minimal medium with acetate was monitored and the cells were harvested during balanced growth (after at least 6 generations). The cells were then washed with fresh medium containing 50 mg/mL of BSA, concentrated around 100-fold and loaded into the device. The individual cells growing in the microchannels were then left overnight in acetate to reach balanced growth before starting image acquisition.

An Elveflow pressure controller (OB1 MK3) equipped with a microfluidic flow sensor (0 to 50 µL/min) was used to assure a constant flow delivery of 20 µL/min to the device. The bottles of medium to be changed at the time of medium switches were prepared and put at 37°C many hours before the switch. After the switch, 2 to 3 minutes were required for the new medium to reach the device. The corresponding dead-volume and dead-time were taken into account in all our analyses. Figure 3.10 illustrates the scenario for a typical microfluidics experiment.

Image acquisition. A Zeiss Axio Observer Z1 inverted microscope equipped with a halogen lamp controlled by a Vincent D1 shutter and a motorized xy-stage was used to perform the experiments. The focus was maintained by Zeiss Definite Focus. Images were recorded using a Zeiss Plan-APOCHROMAT 63x/1.40 oil objective and a Digital CMOS camera (Hamamatsu Orca-Flash 4.0). A constant temperature of 37°C was ensured by a Peltier-equipped box with temperature sensors. The setup was controlled by the Micromanager software [175] and images were recorded using the Multi-Dimensional Acquisition (MDA) feature. A phase contrast image (70 ms exposure) was recorded along with two fluorescence images (green-red or yellow-cyan) every 13 minutes for multiple positions in the device. To obtain fluorescence images, a Zeiss Colibri 7 was used equipped with 7 distinct LEDS, all with fixed intensity at 6%. For each color, the following filters, purchased from Chroma, were used: CFP (reference 49001), GFP (49002), YFP (49003), RFP (49008).

3.5.5 Pad experiments

Agarose pads were prepared by melting medium supplemented with appropriate carbon sources and 1.5% of low-melt agarose. 1 mL of hot mix was placed between two 2x2 cm microscope cover slips and let to solidify for 1 hour in a Petri dish. The upper cover slip was then removed and the pad was cut into four distinct pieces to accommodate multiple strains. Overnight cultures of two reporter strains and the wild-type strain were diluted to an OD600 of 0.1, and 2 μL of each culture was spotted onto a distinct section of the pad. The pad was left to dry for 15 minutes and a microscope slide was then placed on top of the pad. The pad was sealed using paraffin and transferred to the microscope. The same image acquisition setup and parameters were used as described for the mother machine experiments.

In total, two pad experiments were conducted in order to quantify the autofluorescence of our reporter strains as compared to the autofluorescence of the wild-type strain. The obtained images were analyzed in ImageJ and 70 cells were manually segmented using the phase-contrast channel. The fluorescence intensity was then extracted for each segmented cell, fluorescence channel and

frame. For each cell, the fluorescence intensities were averaged across all frames to obtain more robust estimates of the auto-fluorescence.

3.5.6 Image processing

Pre-processing and segmentation For image pre-processing and segmentation, we used the mother machine segmentation software tool BACMMAN [155]. Raw images were sorted and imported into the software. After a quick pre-processing step during which the images were stabilized and de-noised, the microchannels and individual bacteria were segmented on the green (yellow) fluorescence channel. For each experiment, empty microchannels and microchannels presenting double-loading were discarded from the analysis, along with a few cells that stopped growing at various phases of the experiment. The segmentation was performed using the BacteriaFluo algorithm. In our study, we included around 300 mother cells from 4 different types of experiment (Tab. 3.2). After segmentation, the mother cells and first generation daughters were visually checked and the few segmentation errors ($< 5\%$) were manually corrected.

Post-processing For each segmented cell, we obtained the mean fluorescence intensity for each channel [RFU/pixel], the length of each cell [pixel] and various other statistics containing information on position and lineage. These results were imported into Python where several post-processing steps were carried out. The camera noise and background were corrected for each channel by subtracting the base value we obtained with a closed shutter.

The autofluorescence was evaluated using pad experiments, as described above. For all four channels (green, red, yellow, cyan), the auto-fluorescence of the wild-type strain was found to be around zero, after background correction, and negligible in comparison with the fluorescence emitted by the reporter strains (Fig. 3.14). We therefore applied no autofluorescence correction to the measured fluorescence intensities.

To detect any reporter photobleaching effects in our experiments, we performed two microfluidics control experiments with reporter strains growing at

steady state in minimal medium with acetate (Tab. 3.2). The same image acquisition parameters as described in the *Microfluidics experiments* section were used. The images thus obtained were segmented and the fluorescence intensity for each of the four channels extracted (Fig. 3.15). Each curve was fitted by a straight line and the obtained slope coefficients were found to be indistinguishable from zero for all four channels, so no photobleaching correction was applied in the measurement models.

3.5.7 Model definition and calibration

To reconstruct the resource allocation profiles from single-cell fluorescence data, we developed mechanistic models that take into account fluorescent protein maturation, degradation and growth dilution.

GFP has first-order maturation kinetics [93], so the dynamics of the RpsB::GFPmut2 fusions in the RA and RE strains can be described by the following simple model (Text S1 and [159]):

$$\frac{d}{dt} r_{im}(t) = (\mu(t) + \gamma) \left(\frac{\alpha(t)}{\beta} - r_{im}(t) \right) - k_{mat}\, r_{im}(t), \tag{3.3}$$

$$\frac{d}{dt} r_{m}(t) = k_{mat}\, r_{im}(t) - (\mu(t) + \gamma)\, r_{m}(t), \tag{3.4}$$

r_{im} and r_m refer to the concentrations of ribosomes tagged with immature and mature GFP, respectively, in units RFU/pixel. Maturation is characterized by the constant k_{mat} [1/h]. Protein degradation is considered to be identical for both protein species, as determined by the degradation constant γ [1/h]. $\alpha(t)/\beta$ is the (scaled) dynamic resource allocation strategy for ribosomes in the cell that can be reconstructed from single-cell fluorescence data using the methods described below. Note that the sum of Eqs 3.3-3.4 results in Eq. 3.1 in the main text, describing the dynamics of the total (mature and immature) ribosome concentration.

The maturation kinetics of RFP are more complex [93], requiring an addi-

tional equation to capture the dynamics of the ArgG::mScarlet-I fusion:

$$\frac{d}{dt}a_{im}(t) = (\mu(t)+\gamma)\left(\frac{\alpha(t)}{\beta} - a_{im}(t)\right) - k_{mat}\, a_{im}(t), \tag{3.5}$$

$$\frac{d}{dt}a_{hm}(t) = k_{mat}\, a_{im}(t) - k'_{mat}\, a_{hm}(t) - (\mu(t)+\gamma)\, a_{hm}(t), \tag{3.6}$$

$$\frac{d}{dt}a_{m}(t) = k'_{mat}\, a_{hm}(t) - (\mu(t)+\gamma)\, a_{m}(t). \tag{3.7}$$

where a_{im}, a_{hm} and a_m refer to the concentrations of enzymes tagged with immature, half-mature and mature RFP, respectively, in units RFU/pixel. Two maturation constants, k_{mat} and k'_{mat} [1/h] , account for the two-step maturation process of RFP. As above, the sum of the three equations yields the total (mature, half-mature, immature) enzyme concentration. The assumptions underlying the model and the derivation of the equations can be found in Text S1.

The maturation parameters for each fluorescent reporter were determined experimentally, as described before [159], in experiments where an antibiotic (Chloramphenicol) was added to a growing culture to stop translation and the ensuing fluorescence was measured. The maturation models were then fitted to the data to estimate the constants k_{mat} and k'_{mat}. Parameters for GFPmut2 and mScarlet-I were taken from our previous work [159], whereas the same calibration experiments were repeated for mVenus and mCerulean in this work (Figure 3.16).

The degradation parameters for each protein were estimated using data from the same experiment. All parameter values are summarized in Tab. 3.3.

3.5.8 Growth-rate estimation

Growth rates of mother cells over the entire duration of an experiment were inferred from measured cell lengths by means of a custom-made estimation method. The method is based on the following model of growth rate $\mu(t)$:

$$\frac{d}{dt}L(t) = \mu(t)\, L(t), \tag{3.8}$$

where $L(t)$ is the time-varying length of a mother cell, expressed in pixel units. Cell length is a good approximation of cell size in rod-shaped bacteria like *E. coli* [25, 82]. Contrary to many growth-rate estimation methods used for microfluidics data, we did not assume that growth rates are constant between two consecutive cell divisions. However, growth rates were assumed constant between two consecutive measurement times, which is reasonable given the high sampling density (once every 13 min, corresponding to 4-30 length measurements per generation).

As explained in Text S2, the above assumptions result in an estimation problem defined for a piecewise-linear model with as many equations as cell length measurements. Because the solutions of this problem are underdetermined by the (noisy) experimental data, we employed a regularized least squares method [124]. Regularization involves a cost function penalizing rapid changes in growth rate, multiplied with a regularization parameter $\lambda > 0$. We fixed an appropriate value for λ by means of cross-validation [126]. In particular, we estimated an optimal value $\hat{\lambda}^c$ for all mother cells c considered in an experiment. The median of these estimates was used as our best choice for λ, which makes the latter robust to measurement outliers caused by a variety of experimental errors, such as an occasional loss of focus.

The results of the growth-rate estimation method were validated in two different ways. First, we assessed the performance of the method by applying it to synthetic cell length data generated by means of Eq. 3.8 for a given time-course growth-rate profile. To the cell lengths thus generated, we added Gaussian noise in agreement with the experimentally observed distribution. From these synthetic data, the method succeeded in robustly reconstructing the known growth-rate profile (Fig. 3.18). Second, we compared the performance of the method with a baseline approach consisting in the fit of a linear curve to log-transformed length data between two consecutive cell divisions, corresponding to the assumption of a constant growth rate during each generation time. During exponential steady-state growth, when the growth rate is approximately constant, the two methods give comparable results. However, when the assumption is not satisfied, especially during transitions between growth phases, our method gives more plausible estimates (Fig. 3.17).

3.5.9 Inference of resource allocation strategies

Kalman smoothing method. For each mother cell, the unknown resource allocation strategy $\alpha(t)$ was inferred from noisy fluorescence intensity measurements $y_k = r_m(t) + \epsilon_k$, obtained at time-points t_k, $k = 1, \ldots, n$. The measurement errors ϵ_k form a sequence of independent random variables with mean zero and variance s_k^2. We assumed that all measurements obtained in a specific growth phase in an experiment, that is, in a specific growth medium, have the same variance. The variance was estimated from the fluorescence data when the cells had been in balanced growth in the medium for more than 6 generations, and $\alpha(t)$ can be assumed stationary. In particular, we determined the variance from the observed standard deviation around a constant mean in the corresponding time interval. In order to increase the robustness of the estimate, we selected the median of the values for the individual mother cells.

Estimation of the unknown resource allocation profile demands the formulation of a probabilistic prior on $\alpha(t)$. This prior was expressed in the form of a stochastic differential equation

$$d\alpha(t) = -\theta \cdot \alpha(t)dt + \sigma \cdot dW(t), \tag{3.9}$$

where $W(t)$ is the standard Wiener (white noise) process, and σ and θ ¿ 0 are parameters defining the (magnitude and time-scale of) of fluctuations in the resource allocation strategy [159]. We assume that $\alpha(0)$ has a zero-mean Gaussian distribution with variance $\sigma^2/(2\theta)$, which implies that the process is stationary. The parameters σ and θ play the role of regularization parameters that ensure that the inference of $\alpha(t)$ is robust to measurement noise. We denote the optimal values for these parameters by $\hat{\sigma}$ and $\hat{\theta}$ (see below).

With Eq. 3.9, the problem of calculating an optimal estimate $\hat{\alpha}(t)$ for $\alpha(t)$, given fluorescence intensity measurements $y_1, \ldots, y_n$, becomes the Bayesian problem of calculating the conditional expectation

$$\hat{\alpha}(t) = \mathbb{E}[\alpha(t) \mid y_1, \ldots, y_n]. \tag{3.10}$$

This reconstruction is optimal in the sense that it minimizes the variance of the estimation error at any time t [142]. In order to solve the above problem,

the measurement model and the stochastic prior were combined with the maturation models of the green, yellow and cyan reporter proteins (Eqs 3.3-3.4) or the red reporter protein (Eqs 3.5-3.7), giving rise to the following linear stochastic differential equation model:

$$\begin{aligned} d\begin{bmatrix} x(t) \\ \alpha(t) \end{bmatrix} &= \begin{bmatrix} F(t) & G(t) \\ 0 & -\hat{\theta} \end{bmatrix} \begin{bmatrix} x(t) \\ \alpha(t) \end{bmatrix} dt + \begin{bmatrix} 0 \\ \hat{\sigma} \end{bmatrix} dW(t), \\ y_k &= [C0] \begin{bmatrix} x(t_k) \\ \alpha(t_k) \end{bmatrix} + \varepsilon_k. \end{aligned} \tag{3.11}$$

The vector x represents the augmented state of the system, equal to $[r_{im}, r_m, \alpha]$ in the case of the maturation model of Eqs 3.3-3.4, and $[a_{im}, a_{hm}, a_m, \alpha]$ in the case of Eqs 3.5-3.7. Matrices $F(t)$ and $G(t)$ depend on the kinetic parameters in the maturation models, whose values are known (Tab. 3.3), as well as the estimated growth-rate profile $\hat{\mu}(t)$, which was derived as explained above. More precisely, we have

$$\begin{aligned} F(t) &= \begin{bmatrix} -(\hat{\mu}(t) + \gamma + k_{mat}) & 0 \\ k_{mat} & -(\hat{\mu}(t) + \gamma) \end{bmatrix}, \\ G(t) &= \begin{bmatrix} (\hat{\mu}(t) + \gamma)/\beta \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \end{aligned} \tag{3.12}$$

for the green, yellow and cyan reporter proteins (Eqs 3.3-3.4), and

$$\begin{aligned} F(t) &= \begin{bmatrix} -(\hat{\mu}(t) + \gamma + k_{mat}) & 0 & 0 \\ k_{mat} & -(\hat{\mu}(t) + \gamma + k'_{mat}) & 0 \\ 0 & k'_{mat} & -(\hat{\mu}(t) + \gamma) \end{bmatrix}, \\ G(t) &= \begin{bmatrix} (\hat{\mu}(t) + \gamma)/\beta \\ 0 \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \end{aligned} \tag{3.13}$$

the red reporter protein (Eqs 3.5-3.7).

Recasting the system in the above form allowed the use of standard Kalman filtering and smoothing methods [132] to compute $\hat{\alpha}(t)$ from the fluorescence intensity measurements. Text S2 explains how Kalman filtering and smoothing were instantiated for this specific problem. Instead of solving the estimation problem for the entire measurement time-series, we treated the growth phases in different media separately, like for growth-rate estimation. At the medium switching times separating the growth phases, the resource allocation profile is expected to rapidly change over a short time interval, so that a smooth reconstruction of $\alpha(t)$ would be artifactual.

The computational complexity of the whole procedure is linear in the number of data points n, and is mostly determined by the inversion of the (small) matrices in Eq. 3.11. For a single mother cell with fluorescence time-series of $n = 300$ points (our case), the Python implementation using function `odeint` of the `scipy.integrate` module for numerical integration, requires 10 seconds. If the Gaussian assumptions are violated, the estimates computed by this procedure have the interpretation of optimal estimates (minimal error variance) in the class of linear functions of the data [142]. Thanks to this, the method is robust to moderate deviations from Gaussianity.

Automatic tuning of regularization parameters. An optimal choice of σ and θ was determined by solving the maximum likelihood problem

$$(\hat{\theta}, \hat{\sigma}) = \arg \max_{(\theta,\sigma)} f_{\theta,\sigma}(y_1, \ldots, y_n), \tag{3.14}$$

where $f_{\theta,\sigma}(\cdot)$ is the probability density function for the observed data under the stochastic prior of Eq. 3.9 [159]. In practice, for any value of (θ, σ), evaluation of the above likelihood can be performed by Kalman filtering, in a similar way as for the estimation problem above. The details are given in Text S2.

The computational efficiency of Kalman filtering is important for the automatic tuning of the regularization parameters, because the optimization problem requires a large number of evaluations of the right-hand side of Eq. 3.14. Optimization was carried out numerically by the Python function `minimize` of the `scipy.optimize` (File S1). In a typical experiment with 100 mother cells

and $n = 300$, the total computation time of the tuning procedure was around 2 hours.

For every mother cell c, we computed optimal regularization parameters $(\hat{\theta}^c, \hat{\sigma}^c)$ in the way described above. The final choice of the regularization parameters consisted in choosing the median of the values over all c. The latter value was then used for the estimation of the resource allocation strategy from the time-series data for every individual mother cell.

Validation of the method. The method was validated in two different ways. First, we generated synthetic fluorescence data using the model of Eq. 3.11, for a given resource allocation profile and noise characteristics corresponding to those of the real data. The Kalman smoothing estimation method was able to accurately reconstruct $\alpha(t)$ (Fig. 3.19). Second, we visually inspected the fit of the fluorescence intensities predicted by the model from the inferred resource allocation strategy and the actually observed fluorescence intensities for a large number of cells (Fig. 3.1).

3.5.10 Clustering of resource allocation strategies

In order to discover shared adaptation patterns across individual cells considered in an experiment, we clustered the reconstructed resource allocation strategies for ribosomes by means of a K-means clustering algorithm [176] implemented in Python (`sklearn` package [177]). Since all cells reach approximately the same steady-state levels of $\alpha(t)/\beta$, we only clustered the data during the first two hours after the upshift from acetate to glucose (Fig. 3.26). In order to select the number of clusters in the k-means algorithm, we relied on the elbow method [178, 179]. The optimal number of clusters suggested by this method is 2 or 3 (Fig. 3.27). We decided to retain two clusters for the analysis, because they correspond to two clearly distinguishable patterns in the fluorescence intensities (Fig. 3.27). Moreover, as shown in Fig. 3.28, the addition of a third cluster does not lead to distinct growth rate and resource allocation patterns.

3.6 Supplementary Information on Chapter 3

3.6.1 Supplementary Figures

Short name	Long name	Genetic modifications
RA	AP030	*argG::LEI::mScarlet-I//rpsB::LEI::GFPmut2//*Δ*fhuA*
RE	TG008	*mdh::LEI::mScarlet-I//rpsB::LEI::GFPmut2//*Δ*fhuA*
RA*	AP040	*argG::LEI::mCerulean-ME//rpsB::LEI::mVenus//*Δ*fhuA*
WT	CP050	Δ*fhuA*

Table 3.1: Strains used in this study. All strains were derived from the *E. coli* K-12 wild-type strain BW25113 [154]. RA stands for **R**ibosomes and **A**mino acid metabolism, and RE for **R**ibosomes and **E**nergy metabolism.

Condition	Switches: Medium switches	Strain	Cells
ace-glc	ace 24h → glc 16h → ace 24h → glc 16h	RA	134
ace-glc	ace 41h → glc 16h → ace 24h → glc 16h	RE	55
ace-glc	ace 24h→ glc 16h→ ace 24h→ glc 16h	RA*	76
ace	ace 12h	RE	30

Table 3.2: Experimental conditions of microfluidic experiments. The following abbreviations were used: ace - acetate; glc - glucose and arg - L-arginine. Each experimental condition was the subject of at least two independent microfluidic experiments. The column "Cells" presents the total number of mother cells retained for analysis in each experimental condition.

Protein	k_{mat} [min^{-1}]	k'_{mat} [min^{-1}]	γ [min^{-1}]	Source
GFPmut2	0.09 ± 0.004	-	0.00030	[159]
mScarlet-I	0.02 ± 10^{-4}	0.105 ± 0.001	0.00029	[159]
mVenus ME	0.12 ± 0.01	-	0.00025	This study
mCerulean	0.06 ± 0.01	-	0.00031	This study

Table 3.3: Values of the parameters in the dynamic models used for estimating resource allocation strategies. The degradation parameters (γ) and maturation parameters (k_{mat} and k'_{mat}) characterize the dynamics of the four fluorescent proteins used in this study, according to Eqs 3.3-3.4 and Eqs 3.5-3.7 in the main text. The values were determined as described in the *Materials and methods*.

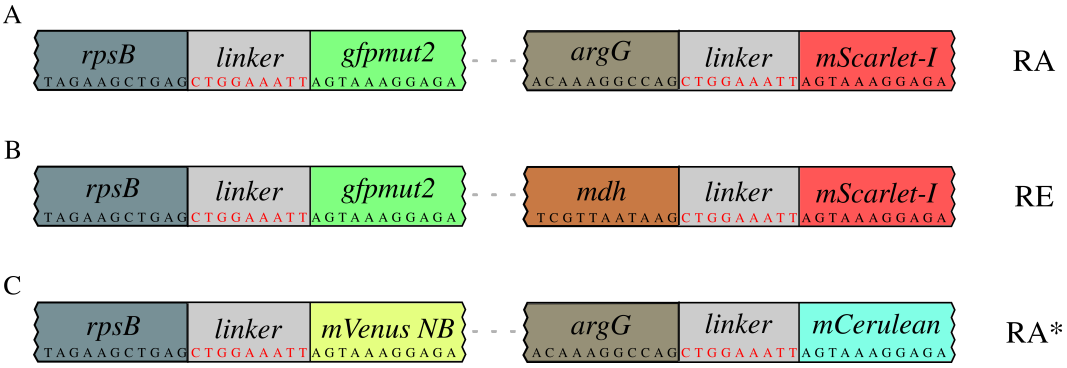


Figure 3.7: Reporter strains used in this study. The three reporter strains were constructed using lambda Red-mediated homologous recombination via the intermediate of a selection "cassette", as described in the *Materials and methods* section. The reporter genes (*gfpmut2*, *mScarlet-I*, *mVenus NB*, *mCerulean*) were fused with one of the target genes (*rpsB*, *argG*, *mdh*) by means of a flexible linker of three amino acids. The genotype of each strain is also detailed in Table 3.1.

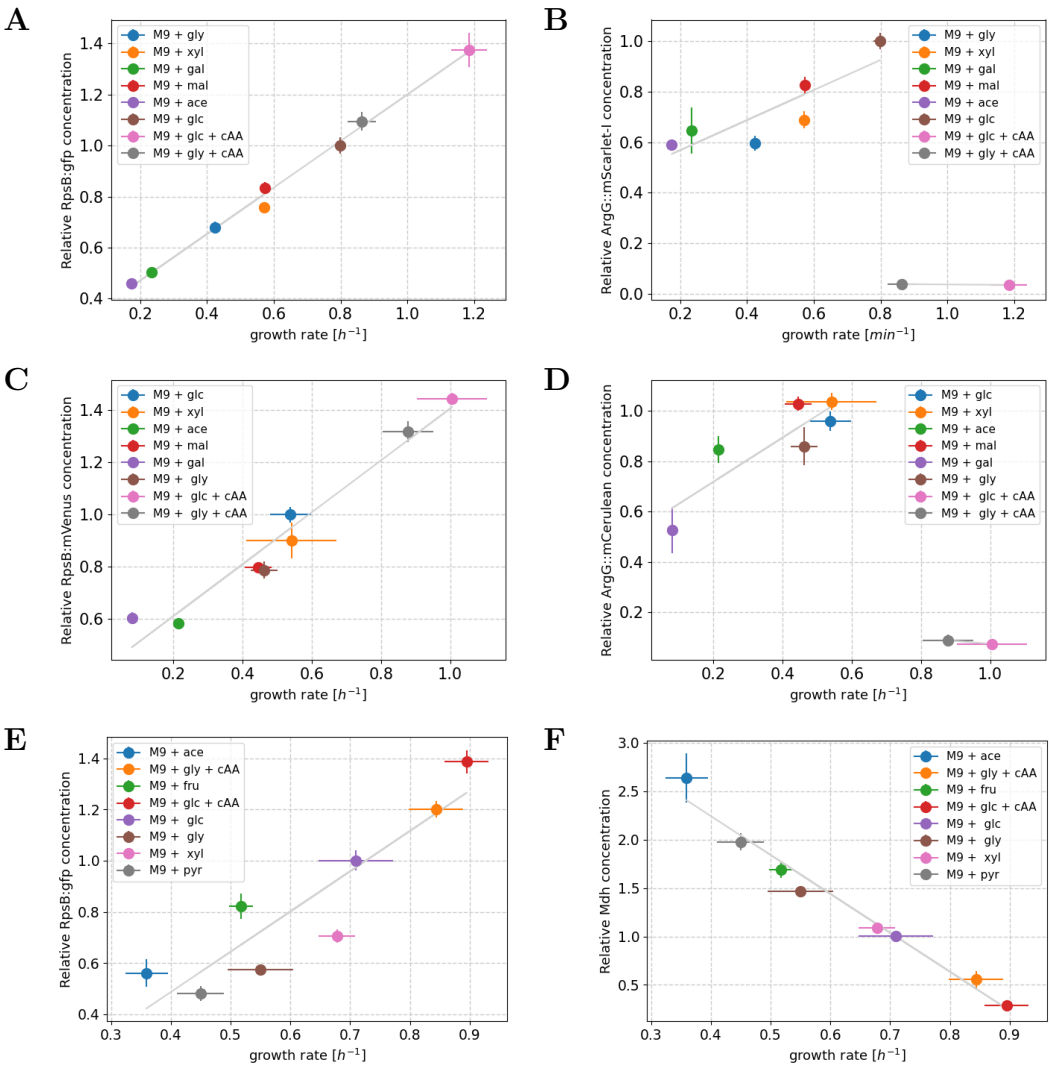


Figure 3.8: Validation of growth rate dependence of protein concentrations in reporter strains. Strains RA (**A-B**), RA* (**C-D**), and RE (**E-F**) were grown in M9 medium supplemented with a variety of carbon sources, as detailed in each figure legend (glc - glucose, ace - acetate, gly - glycerol, fru - fructose, xyl - xylose, pyr - pyruvate, mal - maltose, gal - galactose, cAA - casamino acids). Absorbance, green fluorescence (yellow) and red fluorescence (cyan) were monitored during batch growth in a microplate reader, as described in the *Materials and methods*. After outlier filtering and background correction, the growth rates and reporter concentrations were estimated from the absorbance and fluorescence curves, respectively, using the method described by Zulkower *et al.* [130].

Figure 3.8 (Cont.): The reporter concentrations were then normalized by the concentration obtained in the glucose condition to obtain the relative reporter concentration. Each point is the mean of 5 replicates and confidence intervals are given by two times the standard error of the mean. Straight lines (grey) were used to fit the data and serve as visual aids for interpretation. The observed relationships between growth rate and reporter concentration follow previously published results. The RpsB concentration follows the robust linear relationship between ribosomal content and growth rate, first described by Neidhardt *et al.* [31, 33]. The ArgG concentration also linearly increases with the growth rate, except when amino acids are added to the medium and there is no need for arginine biosynthesis [160]. The Mdh concentration linearly decreases with the growth rate. Both ArgG and Mdh follow the same patterns as previously observed in proteomics data [36, 157] and agree with predictions of resource allocation models for their respective protein categories [33, 36].

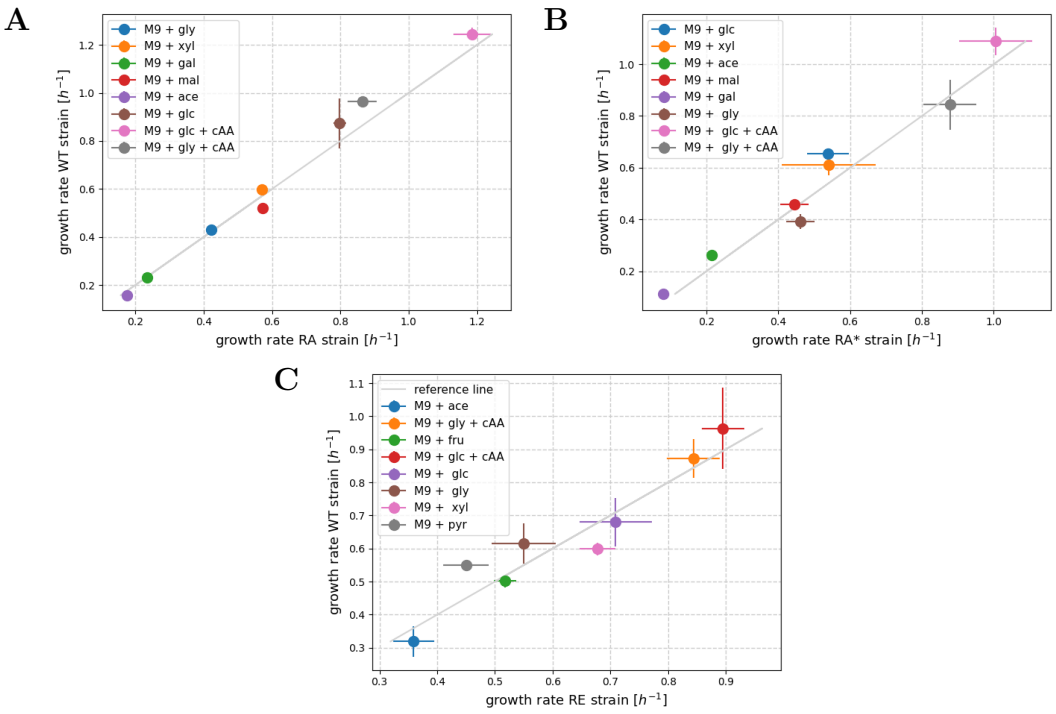


Figure 3.9: Comparison of the growth rate of wild-type and reporter strains in multiple conditions. Strains RA (**A**), RA* (**B**), RE (**C**) and WT were grown in a microplate reader in M9 minimal medium supplemented with several carbon sources while monitoring the absorbance, as described in the *Materials and methods*. The same abbreviations as in Fig. 3.8 were used. After outlier filtering and background correction, the growth rate during exponential growth was estimated from the absorbance curves using the method described by Zulkower *et al.* [147]. Each point is the mean of 5 replicates and confidence intervals are given by two times the standard error of the mean. The scatter plots show the growth rates of the wild-type strain compared with the growth rate of each of the three reporter strains over the different conditions (diagonal in gray). The three reporter strains and the wild-type strain grow very similarly ($R^2_{RA} = 0.99$, $R^2_{RA^*} = 0.96$ and $R^2_{RE} = 0.91$). Therefore, no growth defect is observed after tagging the three genes of interest with fluorescent reporters.

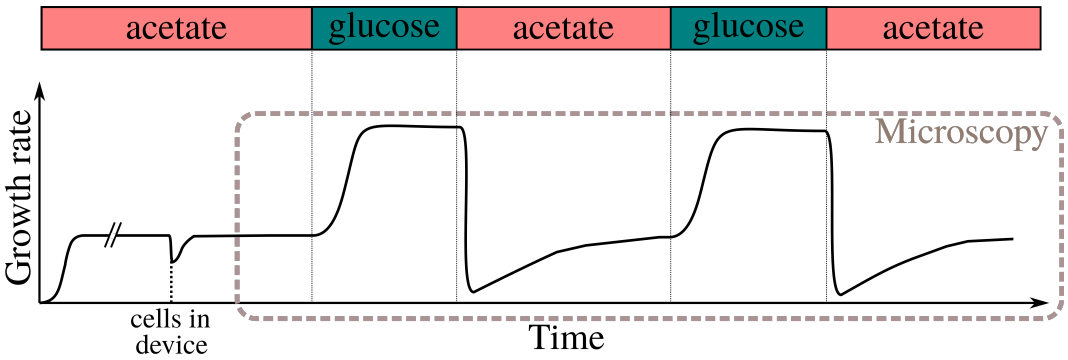


Figure 3.10: Schematic outline of a typical microfluidic experiment. Pre-cultures growing exponentially in M9 minimal medium supplemented with acetate were prepared and injected into a mother machine. A constant flow of 20 µL/min of fresh acetate medium was supplied to the growing bacteria in the device. The cell length and the green (yellow) and red (cyan) fluorescence intensity were monitored over time using microscopy (*Materials and methods*). The bacteria were submitted to a sequence of an upshift to glucose followed by a downshift to acetate, twice. After each shift, the bacteria were allowed to adapt to the new environment for at least eight generations. The exact scenarios of all experiments are detailed in Table 3.2.

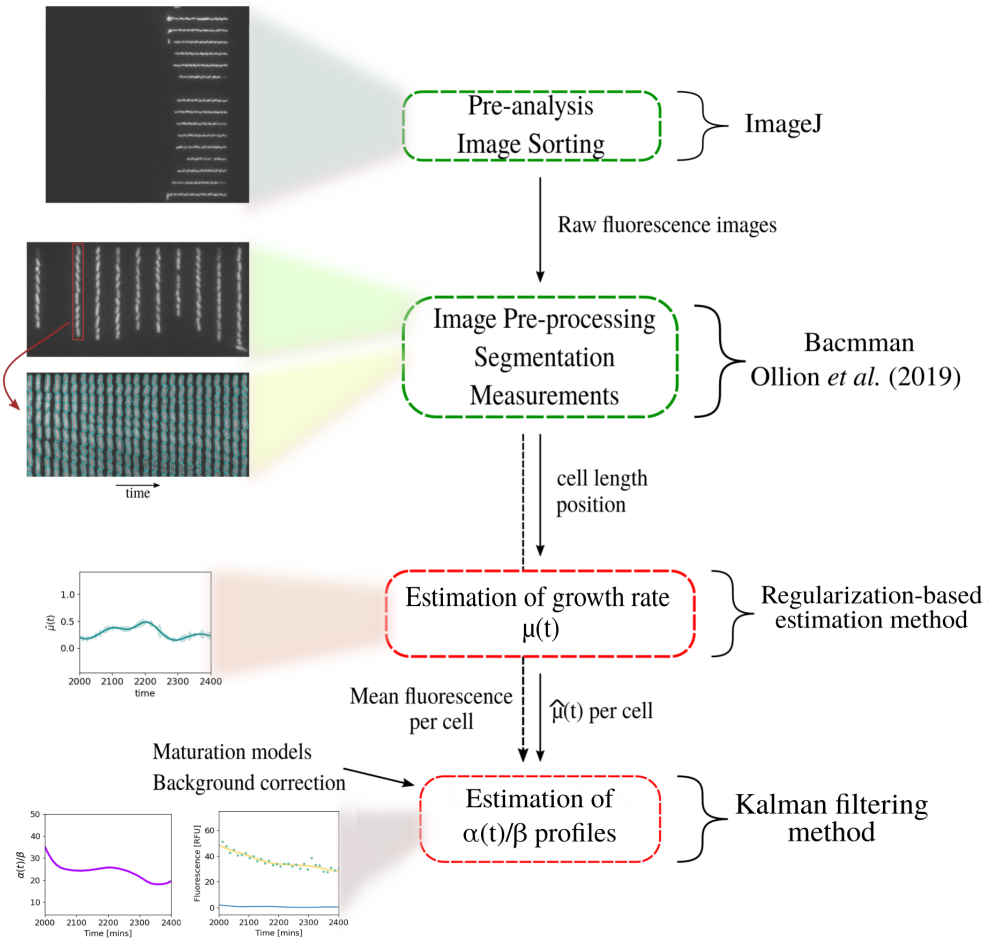


Figure 3.11: Data analysis pipeline implemented for this study. Raw fluorescence images are imported into ImageJ for image sorting. The sorted images are then imported into BACMMAN [155] for image pre-processing, and segmentation of microchannels and bacteria on the brightest fluorescence channel. After manual checking and correction of the segmentation, cell length and mean fluorescence intensities of all channels are determined for each cell. The cell length is used to estimate growth rate using a regularization-based fitting method in the log domain. The growth-rate estimates are used along with the background-corrected mean fluorescence intensity for each cell by a Kalman smoothing method that can reconstruct the time-varying resource allocation strategies. Maturation models are used in the method to correct for fluorescent reporter dynamics. The figure shows an example image or plot for each step of the procedure. Green boxes represent tools developed in previous studies, red boxes represent methods developed for this study. More information on each in the procedure can be found in the *Materials and methods*.

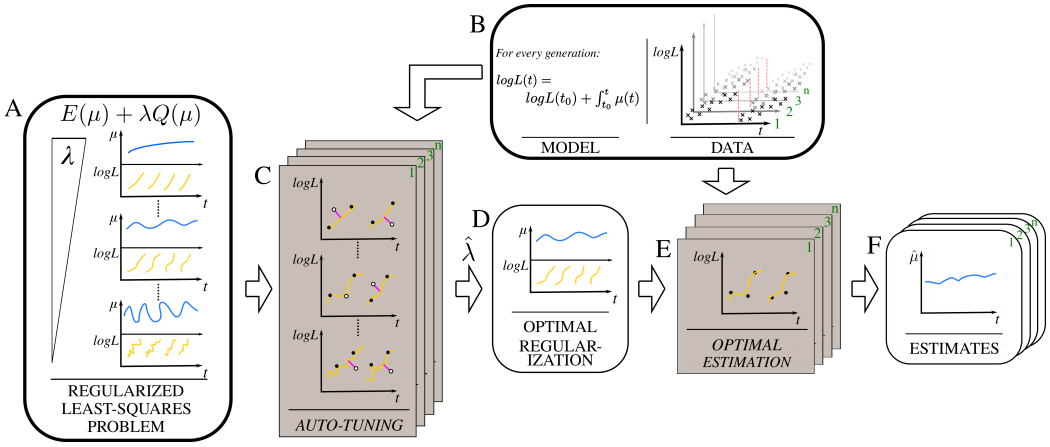


Figure 3.12: Schematic outline of the method used for the estimation of growth rate from cell length data. (**A**) Definition of the regularized least-squares problem. $Q(\mu)$ represents a cost function penalizing large fluctuation in the growth rate and λ is the associated regularization parameter. Larger values of λ penalize faster fluctuations in the growth rate. (**B**) Noisy length measurements in the log domain between divisions are used in a piecewise-linear model to obtain growth rate estimates that may change within a generation. (**C**) The optimal regularization parameter is chosen for each cell via a cross-validation method and (**D**) is used to obtain the globally optimal regularization parameter over all cells. (**E**) The optimal regularization parameter along with the defined model is used to fit the data and (**F**) obtain the sought growth rate estimates for each cell. Rectangular boxes represent procedures. Rounded boxes represent inputs, outputs and results of the method.

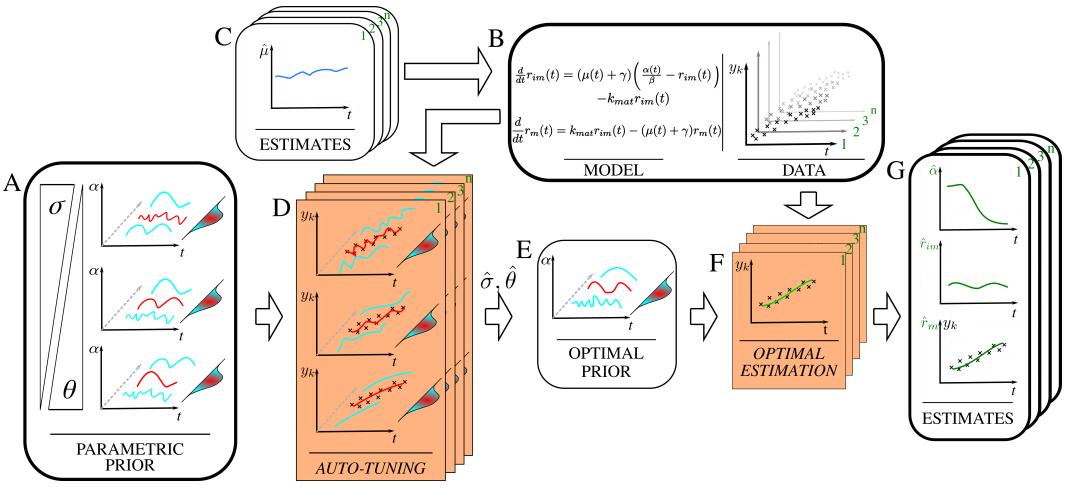


Figure 3.13: Schematic outline of the Bayesian approach for estimating single-cell resource allocation profiles. (**A**) Probabilistic priors on α. A larger σ (smaller θ) assigns higher probability to faster fluctuations in the profile. (**B**) The kinetic model, the fluorescence data of individual cells (**C**), along with their corresponding estimated growth rate (Fig. 3.12) are used in (**D**) the auto-tuning step so that the best values of σ and θ are selected by comparing the fluorescence profiles predicted by the model with the data via a maximum-likelihood approach. (**E**) The optimal prior selected is (**F-G**) used with the single-cell fluorescence intensity measurements for the optimal estimation of individual resource allocation profiles and predicted fluorescence outputs. Rectangular boxes represent procedures. Rounded boxes represent inputs, outputs and results of the method.

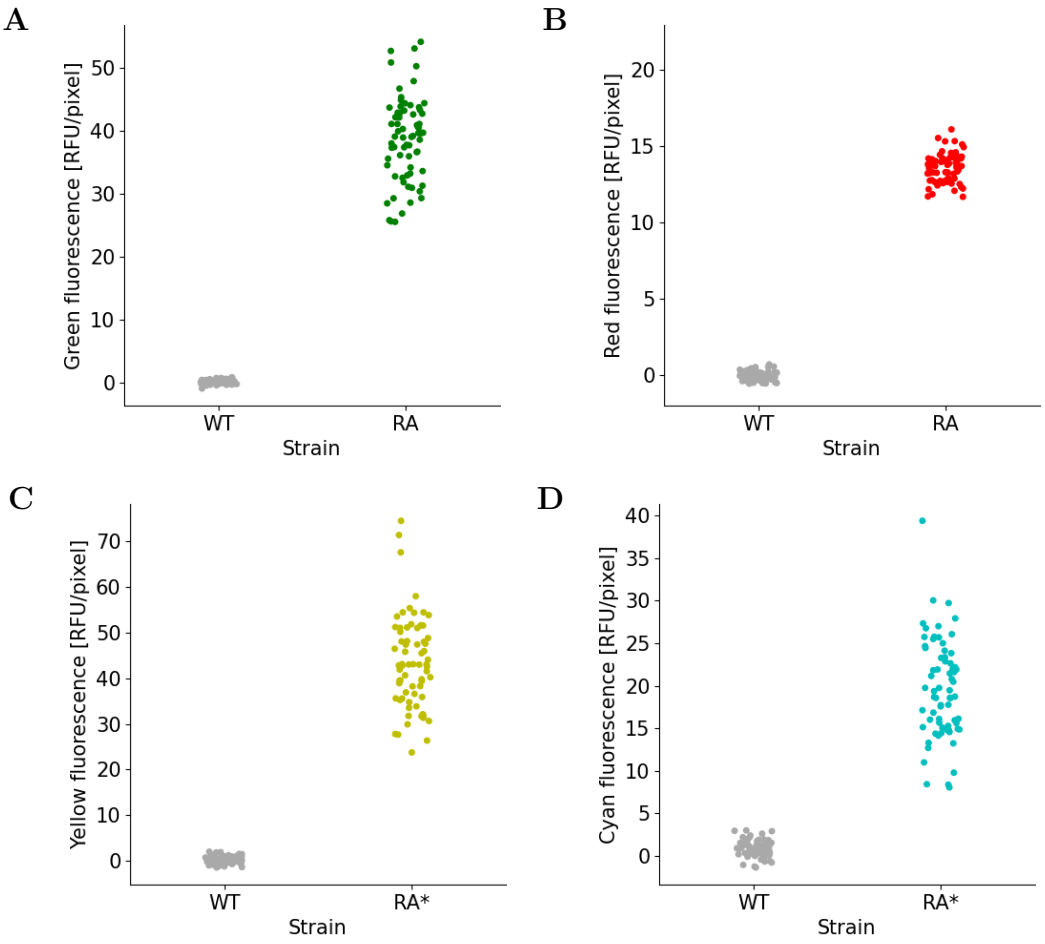


Figure 3.14: Quantification of the effect of photobleaching for the fluorescent reporters used in this study. Strains RA (**A**-**B**), RA* (**C**-**D**) and WT were inoculated on an agar pad containing M9 supplemented with 2 g/L of acetate and left to grow overnight, as described in the *Materials and methods* section. Images of green (**A**) and red (**B**) fluorescence were acquired for strain RA, and yellow (**C**) and cyan (**D**) fluorescence for strain RA*. Images from the WT strain were obtained for all channels in parallel. Image acquisition parameters were used as described in the *Materials and method* section. For each channel and strain, 70 cells were segmented using ImageJ and the mean fluorescence intensity [RFU/pixel] was determined for each cell over a sequence of frames. For each image, the background was removed by subtracting the fluorescence intensity of the pad background. The mean fluorescence intensity of the wild-type strain is 30-4000-fold lower than the fluorescence of the modified strains, and therefore negligible for all channels.

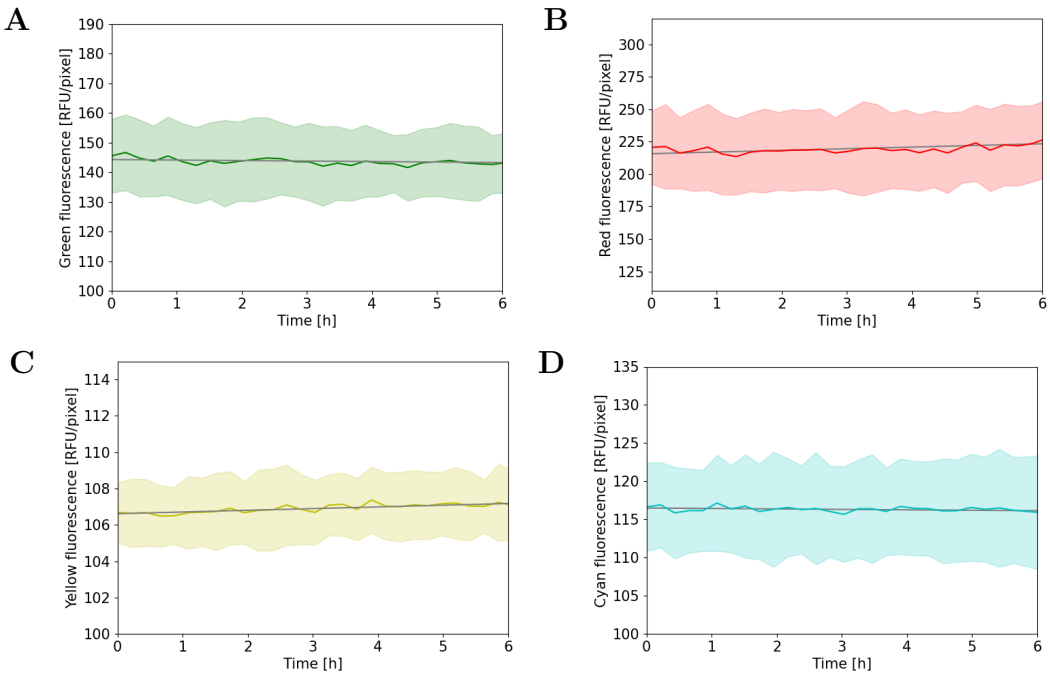


Figure 3.15: Quantification of the autofluorescence of the reporter strains and the wild-type strain. Strains RA (**A**-**B**), RA* (**C**-**D**) and WT were inoculated on an agar pad containing M9 supplemented with 2 g/L of acetate and left to grow overnight, as described in the *Materials and methods*. Images of green (**A**) and red (**B**) fluorescence were acquired for strain RA, and yellow (**C**) and cyan (**D**) fluorescence for strain RA*. Images from the WT strain were obtained for all channels in parallel. Image acquisition parameters were used as described in the *Materials and methods*. For each channel and strain, 70 cells were segmented using ImageJ and the mean fluorescence intensity [RFU/pixel] was determined for each cell over a sequence of frames. For each image, the background was removed by subtracting the fluorescence intensity of the pad background. The mean fluorescence intensity of the wild-type strain is 30-4000-fold lower than the fluorescence of the modified strains, and therefore negligible for all channels.

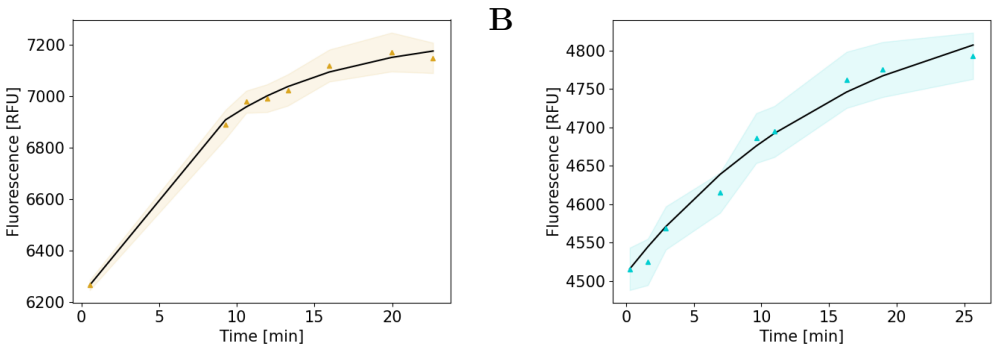


Figure 3.16: Estimation of maturation parameters for mVenus NB and mCerulean. Strain RA* was grown in M9 medium supplemented with glucose. At time zero Chloramphenicol was added to the medium to stop translation. Yellow (A) and cyan (B) fluorescence were monitored. A one-step maturation model (Eqs 3.3-3.4 in the main text) was used to fit each curve [159] and estimate the maturation parameters for mVenus NB and mCerulean (best fit - black line). Each graph shows the mean of 4 replicates, and confidence intervals are given by two times the standard error of the mean. The R^2 for yellow and cyan curve fits are 0.98 and 0.99 respectively and the estimated parameter values can be found in Tab. 3.3.

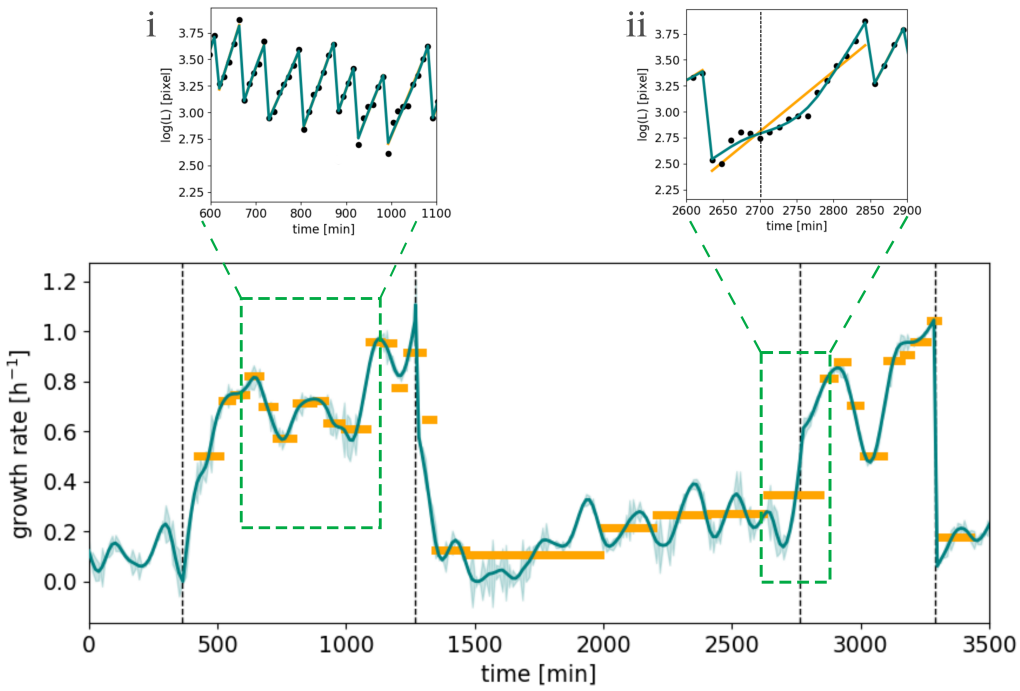


Figure 3.17: Comparison of growth-rate estimation methods. The length measurements of one mother cell were used to estimate the growth rate by means of the regularization method described in the *Materials and methods* (blue curves), for an ace-glc experiment (Tab. 3.2) with multiple growth transitions (dotted lines). The blue shading represents the standard errors of the estimates. This method was compared with a simple fit of an exponential curve to the measured cell lengths between divisions (yellow lines) on the same data. The fits obtained from our regularization method (pink) and the exponential fit (green) on the logarithm of length data (blue dots) were compared for two distinct phases of the experiment (**i - ii**). During steady-state exponential growth on (**i**), the two methods give similar estimates as shown by the superimposed fits in the inserted panel **i**. After switches between growth phases (**ii**), our method, which does not assume that growth rate is constant over a generation, produces more accurate fits, and therefore provides more reliable growth-rate estimates.

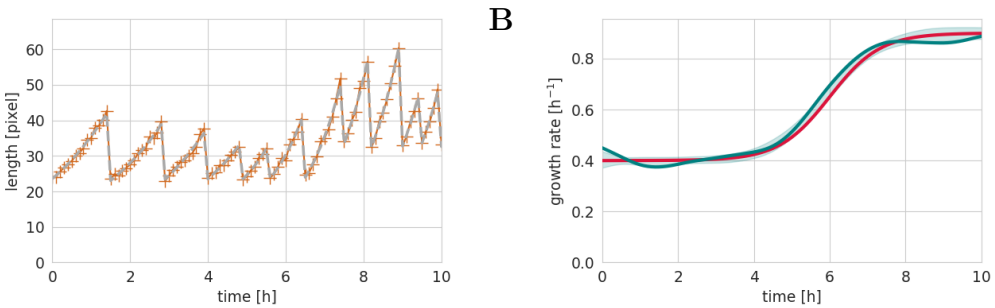


Figure 3.18: Performance of growth-rate estimation method on synthetic data. A sigmoid-like function was defined to resemble a transition from low to high growth rate. A series of cell division points, along with cell lengths after division with some added white noise were defined to resemble a transition from a poor to a rich carbon source, as observed in our microfluidics experiments. Using the model for cellular growth described in the *Materials and methods* of the main text (Eq. 3.29), and the defined cell division points, a series of length measurements was generated with added white noise $\sim$N(0, 1.4) (orange points). **B** The growth-rate estimation described in the *Materials and methods* and Text S2 was used to provide regularized estimates of the growth rate. The corresponding fit to the length data is shown in panel A (gray curve) and the resulting growth rate curve in panel B (blue curve). The shaded blue area represents two times the standard error of the mean of growth-rate estimates obtained from 100 generated datasets. The method is shown to provide growth-rate estimates that are close to the true growth rate used as input for the simulations.

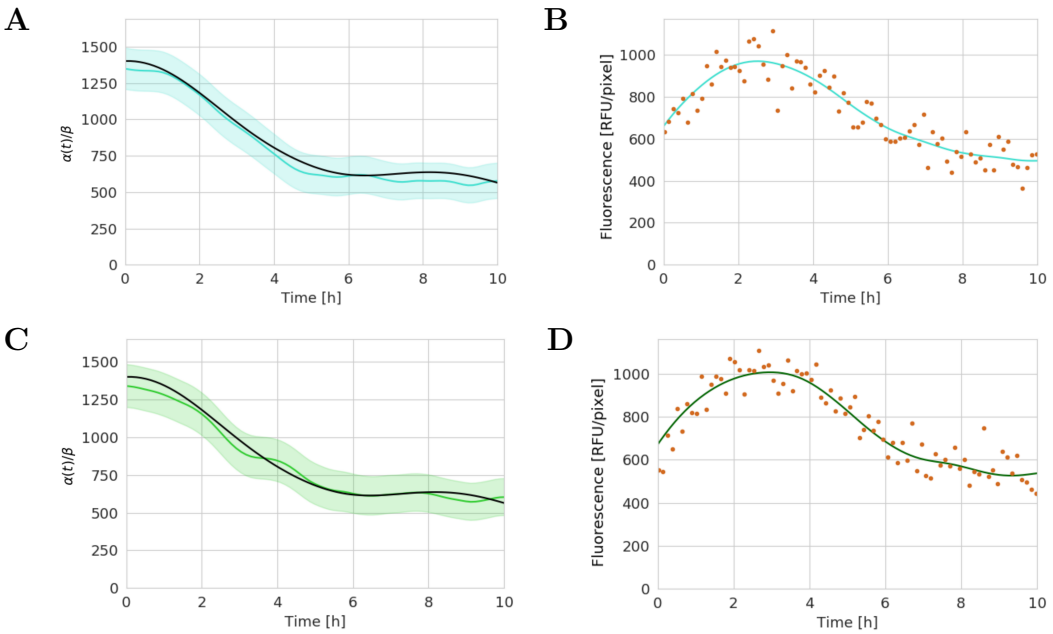


Figure 3.19: Performance of Kalman smoothing algorithm on synthetic data. (**A**) A sinus-like promoter activity $pa(t) = (\mu(t) + \gamma) \cdot \alpha(t)/\beta$ was generated and was divided by the sum of the simulated growth rate (Figure 3.18B, red curve) and degradation constant γ to obtain the simulated resource allocation strategy (black curve). This curve was compared with the estimated resource allocation profile (blue curve) obtained by applying the Kalman smoothing method (*Materials and methods* and Text S2) to simulated data. The shaded blue area represents the confidence interval of the estimates produced by the method (Text S2). (**B**) Synthetic fluorescence intensity data generated by means of Eqs 3.3-3.4 in the main text and the resource allocation profile in panel A, with added white noise of amplitude equivalent to that observed in the single-cell fluorescence measurements ($\sim$N(0, 80), orange points). The blue curve is the predicted fluorescence intensity predicted from the estimated resource allocation strategy. As can be seen in panels A and B, the algorithm is able to robustly reconstruct the resource allocation profile from the synthetic data (**C** - **D**). To verify that the method is able to reconstruct the resource allocation strategy using estimates of the growth rate, $\hat{\mu}(t)$, rather than the true growth rate, $\mu(t)$, the resource allocation profile was reconstructed using synthetic data generated from the same sinus-like promoter activity as above, but divided by the growth rate estimates $\hat{\mu}(t)$ of Figure 3.18B (blue curve). The use of growth-rate estimates instead of the true growth rate does affect the quality of the reconstruction of the resource allocation strategies and the correspondence of the predicted and observed fluorescence intensities.

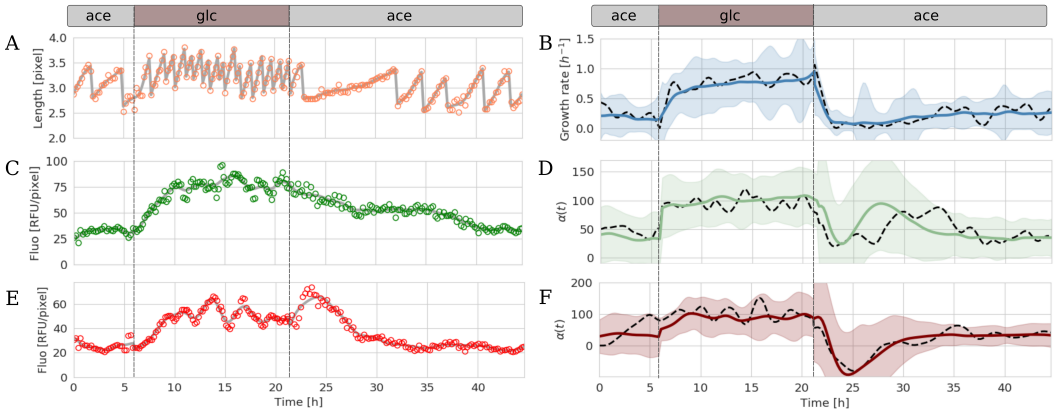


Figure 3.20: Measured and estimated quantities in mother machine experiments with the RA reporter strain (RpsB and ArgG) A-C-E. Cell length (orange dots), green (green dots) and red (red dots) fluorescence intensity were quantified over time in a mother machine tracking individual bacteria of strain RA, carrying a fusion of the ribosomal subunit S2 and GFPmut2 and a fusions of ArgG and mScarlet-I. The experiment consisted in several consecutive upshifts and downshifts (vertical dashed lines) between minimal media with glucose or acetate. **B-D-F.** Cell length measurements were used to estimate growth rates and fluorescence intensity measurements to estimate resource allocation strategies, using two statistical inference methods (*Materials and methods*). The gray solid curves in panels A, C and E represent the fits of the single-cell data obtained from the two methods, whereas the black dashed curves in panels B, D and F represent the corresponding estimates of the growth rate $\mu(t)$ and the resource allocation strategies $\alpha(t)/\beta$ for this same (mother) cell. Blue, green and red solid curves represent the mean of the estimates over all cells considered in the experiment and confidence intervals are given as two times the standard deviation.

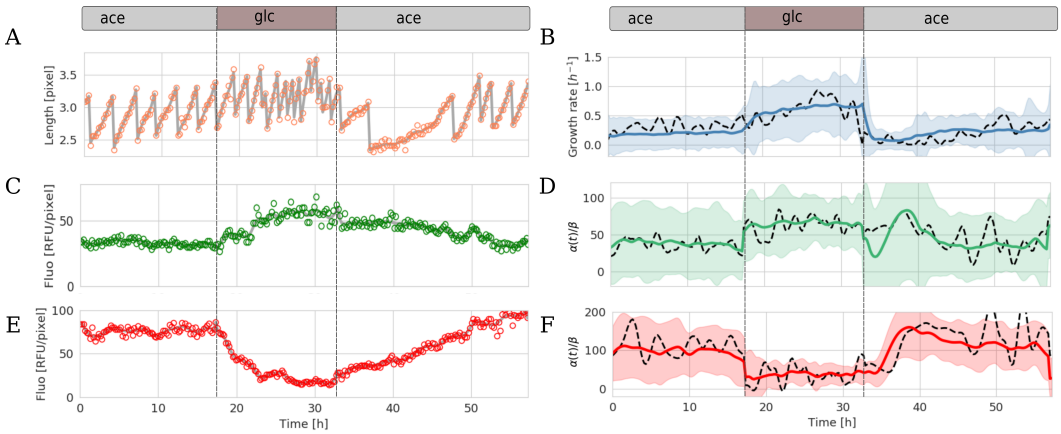


Figure 3.21: Measured and estimated quantities in mother machine experiments with the RE reporter strain (RpsB and Mdh). A-C-E. Cell length (orange dots), green (green dots) and red (red dots) fluorescence intensity were quantified over time in a mother machine tracking individual bacteria of strain RE, carrying a fusion of the ribosomal subunit S2 and GFP-mut2 and a fusions of Mdh and mScarlet-I. The experiment consisted in several consecutive upshifts and downshifts (vertical dashed lines) between minimal media with glucose or acetate. **B-D-F.** Cell length measurements were used to estimate growth rates and fluorescence intensity measurements to estimate resource allocation strategies, using two statistical inference methods (*Materials and methods*). The gray solid curves in panels A, C and E represent the fits of the single-cell data obtained from the two methods, whereas the black dashed curves in panels B, D and F represent the corresponding estimates of the growth rate $\mu(t)$ and the resource allocation strategies $\alpha(t)/\beta$ for this same (mother) cell. Blue, green and red solid curves represent the mean of the estimates over all cells considered in the experiment and confidence intervals are given as two times the standard deviation.

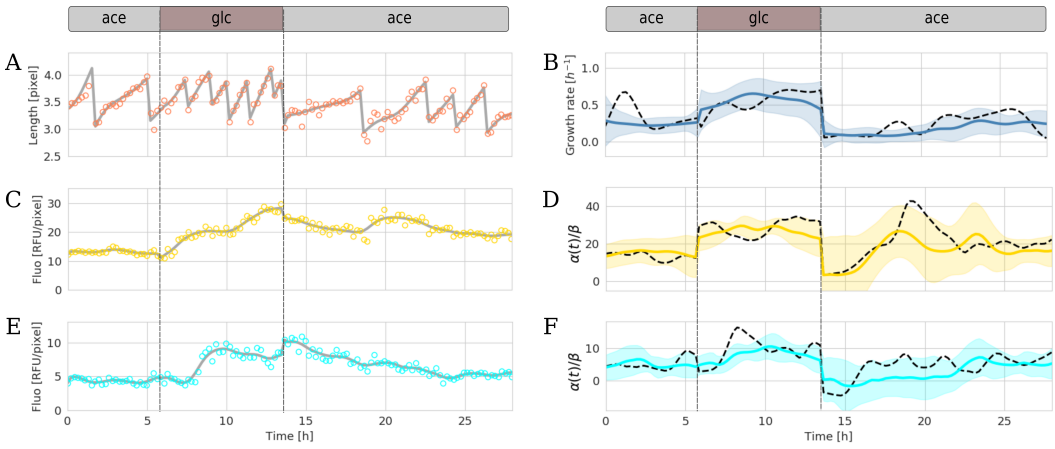


Figure 3.22: Measured and estimated quantities in mother machine experiments with the RA* reporter strain (RpsB and ArgG). A-C-E. Cell length (orange dots), yellow (yellow dots) and cyan (cyan dots) fluorescence intensity were quantified over time in a mother machine tracking individual bacteria of strain RA*, carrying a fusion of the ribosomal subunit S2 and mVenus NB and a fusions of ArgG and mCerulean. The experiment consisted in several consecutive upshifts and downshifts (vertical dashed lines) between minimal media with glucose or acetate. **B-D-F.** Cell length measurements were used to estimate growth rates and fluorescence intensity measurements to estimate resource allocation strategies, using two statistical inference methods (*Materials and methods*). The gray solid curves in panels A, C and E represent the fits of the single-cell data obtained from the two methods, whereas the black dashed curves in panels B, D and F represent the corresponding estimates of the growth rate $\mu(t)$ and the resource allocation strategies $\alpha(t)/\beta$ for this same (mother) cell. Blue, yellow and cyan solid curves represent the mean of the estimates over all cells considered in the experiment and confidence intervals are given as two times the standard deviation.

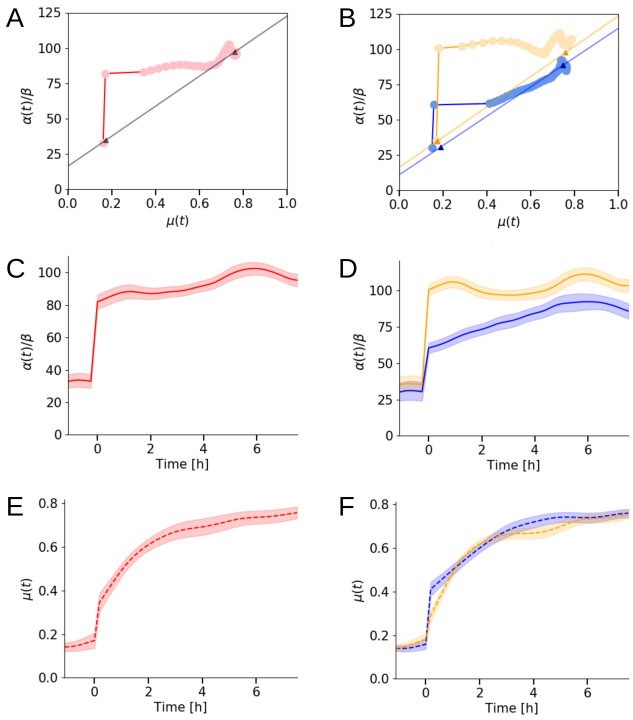


Figure 3.23: Adaptation dynamics of growth rate and ribosomal resource allocation after second nutrient upshift in reference experiment. A. Population-averaged adaptation trajectory of the resource allocation strategy for ribosomes α/β and the growth rate μ for another acetate-glucose upshift applied to the RA strain growing in a mother machine. The resource allocation strategies and growth rates were estimated from the data using the inference methods of Fig. 3.2 and averaged over 127 cells. The arrows indicate increasing time after the upshift. The triangles denote the average values of α/β and μ during balanced growth on acetate (before the upshift) and glucose (after the upshift). The latter values were determined by computing for each cell the mean growth rate over a period of balanced growth (> 2 h), and then averaging these values over the individual cells. The black line through the population average before and after the upshift are shown as a visual aid. The trajectories show that the adaptation of resource allocation and growth rate are uncoupled in a first phase and coupled in a second. **B.** Clustering of the single-cell resource allocation strategies for ribosomes after the upshift using K-means (*Materials and Methods* and Fig. 3.27) reveals two distinct trajectories for fast adapters (orange, 81 cells) and slow adapters (blue, 46 cells). The orange and blue straight lines connect the balanced growth values (triangles) of the two populaitions before and after the shift. **D-F.** Time-courses of the resource allocation strategies in panels C-D, respectively. Confidence intervals are given by two times the standard error of the mean. **E-F.** Time-courses of the growth rates in panels A-B.

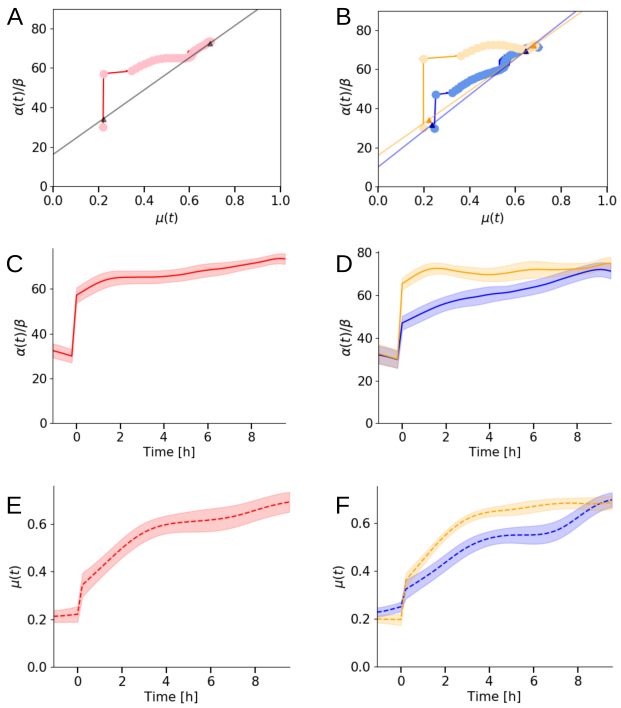


Figure 3.24: Adaptation dynamics of growth rate and ribosomal resource allocation after a nutrient upshift in an independent replicate of the reference experiment. A. Population-averaged adaptation trajectory of the resource allocation strategy for ribosomes α/β and the growth rate μ for an acetate-glucose upshift applied to the RA strain growing in a mother machine in an independent replicate of the reference experiment. The resource allocation strategies and growth rates were estimated from the data using the inference methods of Fig. 3.2 and averaged over 55 cells. The arrows indicate increasing time after the upshift. The triangles denote the average values of α/β and μ during balanced growth on acetate (before the upshift) and glucose (after the upshift). The latter values were determined by computing for each cell the mean growth rate over a period of balanced growth (> 2 h), and then averaging these values over the individual cells. The black line through the population average before and after the upshift are shown as a visual aid. The trajectories show that the adaptation of resource allocation and growth rate are uncoupled in a first phase and coupled in a second. **B.** Clustering of the single-cell resource allocation strategies for ribosomes after the upshift using K-means (*Materials and Methods* and Fig. 3.27) reveals two distinct trajectories for fast adapters (orange, 21 cells) and slow adapters (blue, 24 cells). The orange and blue straight lines connect the balanced growth values (triangles) of the two populaitions before and after the shift. **D-F.** Time-courses of the resource allocation strategies in panels C-D, respectively. Confidence intervals are given by two times the standard error of the mean. **E-F.** Time-courses of the growth rates in panels A-B.

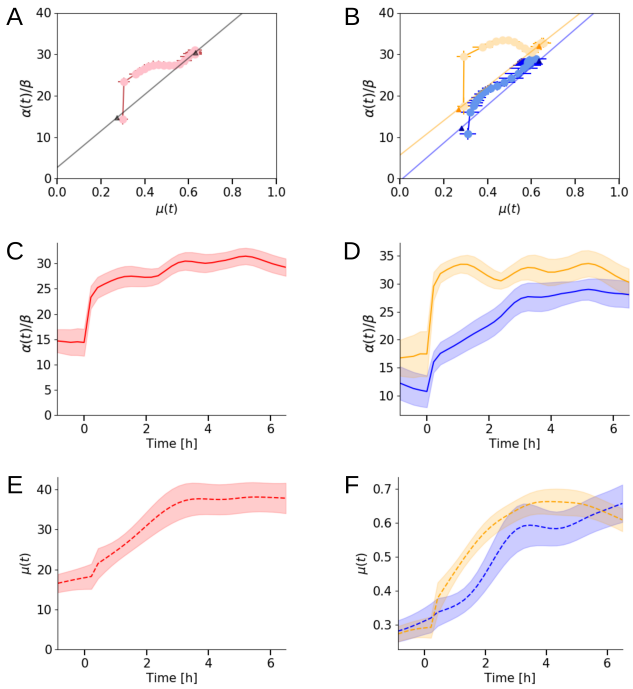


Figure 3.25: Adaptation dynamics of growth rate and ribosomal resource allocation after a nutrient upshift in a replicate reference experiment using the alternative reporter strain. A. Population-averaged adaptation trajectory of the resource allocation strategy for ribosomes α/β and the growth rate μ for an acetate-glucose upshift applied to the RA* strain growing in a mother machine in an independent replicate of the reference experiment with an alternative reporter strain. The resource allocation strategies and growth rates were estimated from the data using the inference methods of Fig. 3.2 and averaged over 76 cells. The arrows indicate increasing time after the upshift. The triangles denote the average values of α/β and μ during balanced growth on acetate (before the upshift) and glucose (after the upshift). The latter values were determined by computing for each cell the mean growth rate over a period of balanced growth (> 2 h), and then averaging these values over the individual cells. The black line through the population average before and after the upshift are shown as a visual aid. The trajectories show that the adaptation of resource allocation and growth rate are uncoupled in a first phase and coupled in a second. **B.** Clustering of the single-cell resource allocation strategies for ribosomes after the upshift using K-means (*Materials and Methods* and Fig. 3.27) reveals two distinct trajectories for fast adapters (orange, 40 cells) and slow adapters (blue, 36 cells). The orange and blue straight lines connect the balanced growth values (triangles) of the two populaitions before and after the shift. **D-F.** Time-courses of the resource allocation strategies in panels C-D, respectively. Confidence intervals are given by two times the standard error of the mean. **E-F.** Time-courses of the growth rates in panels A-B.

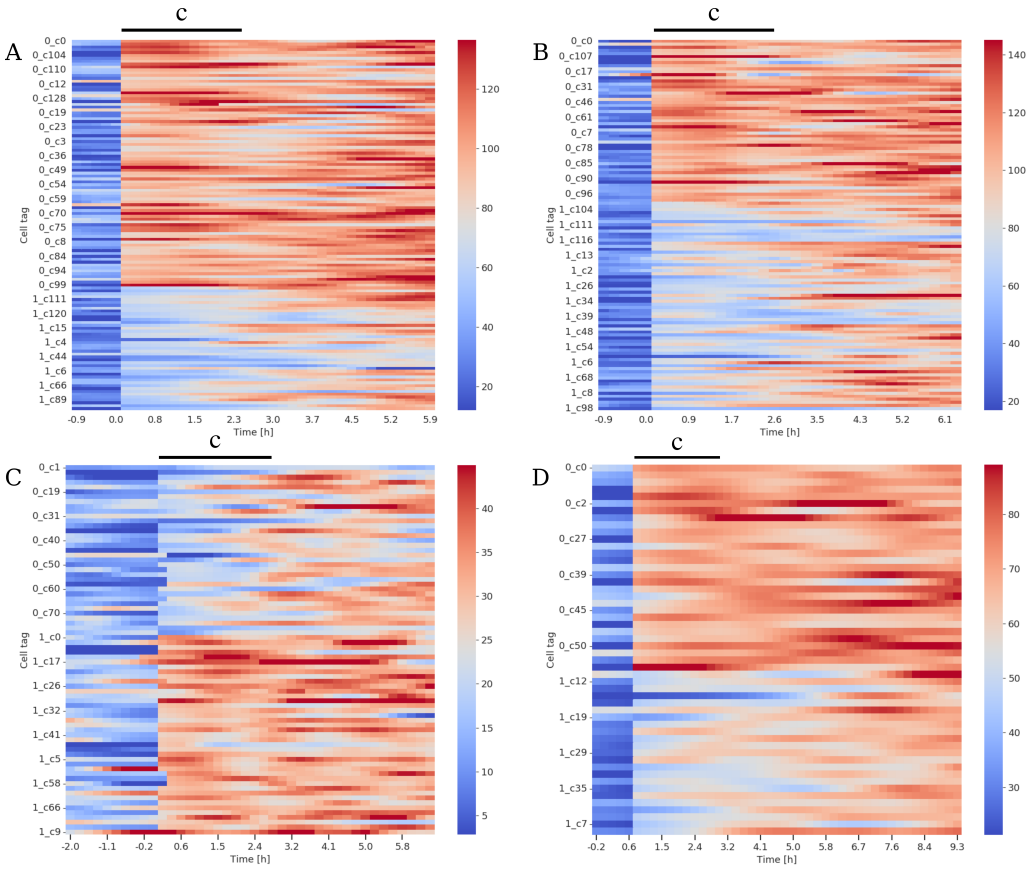


Figure 3.26: Heatmaps of clustered single-cell resource allocation strategies after a nutrient upshift. Time-varying resource allocation strategies after an acetate-glucose upshift, inferred from the fluorescence data using the method in Fig. 2, were clustered by means of a k-means algorithm (*Materials and methods*). Each heatmap corresponds to one upshift in one mother machine experiment: **A.** Fig. 3.4, **B.** Fig. 3.23, **C.** Fig. 3.25, **D.** Fig. 3.24. The strategies were clustered over the initial time-interval after the upshift, indicated by the black lines and letter c. Each cell contains a tag composed of a cluster index (0 or 1, for two clusters) and a cell number indicating its position in the mother machine. Time 0 indicates the time of the upshift.

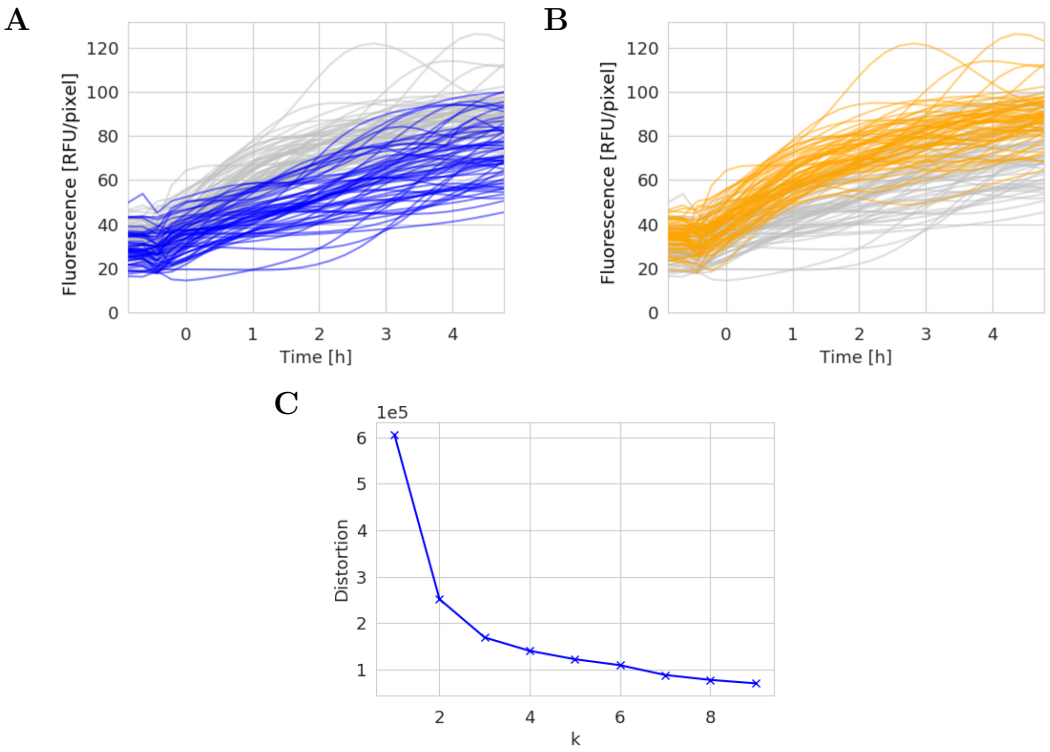


Figure 3.27: Determination of number of clusters used by k-means algorithm. In order to evaluate the variability of the adaptation dynamics of ribosomes, we applied the k-means clustering algorithm to the inferred resource allocation strategies after an acetate-glucose upshift (*Materials and methods* and Fig. 3.26). We decided on the number of clusters (k) appropriate for the data by relying on the so-called Elbow method, often used in the context of k-means clustering [178]. As a metric for the information added by each cluster, we used cluster distortions [180]. The optimal number of clusters is given by the "elbow" in the curve displaying the distortion as a function of k, where adding another cluster significantly decreases the marginal gain in information. In our case, the optimal number of clusters is 2 or 3 (C). We chose to distinguish two clusters because they correspond to two clearly distinguishable patterns in the fluorescence intensities (**A-B**). We show in Fig. 3.28 that adding a third cluster is not more informative from a biological point of view.

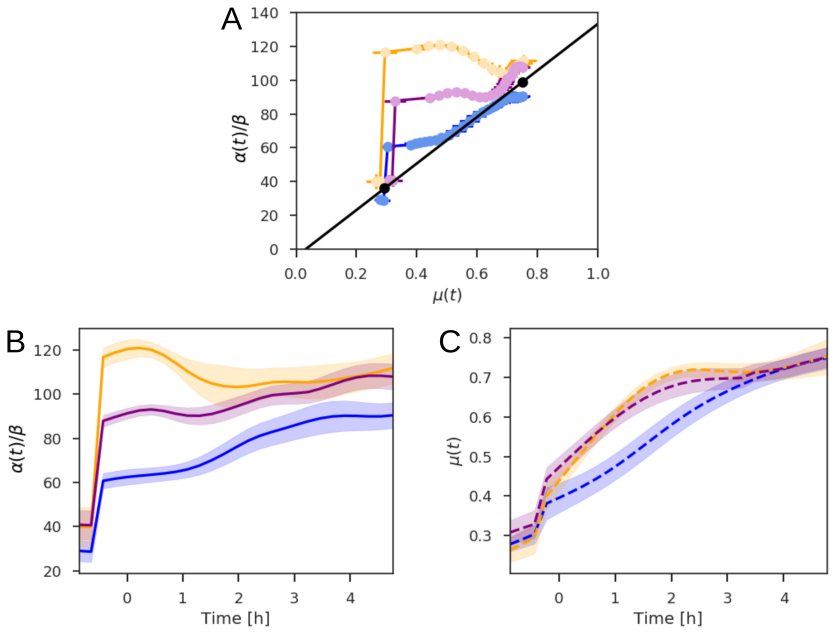


Figure 3.28: Adaptation dynamics of growth rates and ribosomal resource allocation strategies in the case of three clusters (**A**) Adaptation trajectory of ribosomes after an acetate-glucose upshift for the same data presented in Fig. 3.4 of the main text but for three instead of two clusters. K-means clustering algorithm was used, as described in the *Materials and methods* (**B-C**) Corresponding time-courses of the mean resource allocation strategies (panel B) and mean growth rates (panel C) for the three clusters. Confidence intervals are given by two times the standard error of the mean.The yellow and purple cluster have the same resource allocation values for balanced growth on glucose and acetate, and the same growth rate during the upshift. The distinction of three rather than two clusters therefore does not add much information from a biological point of view.

3.6.2 Text S1: Model definition

This supplementary text describes how the dynamical models used for the inference of the ribosomal resource allocation strategy in the main text can be derived from basic biological assumptions. We proceed in two steps. In the first step, we relate resource allocation strategies to the total concentration of ribosomes and other classes of proteins. In the second step, we couple this biological model with a measurement model, in order to relate the resource allocation strategies to the observed fluorescence in single-cell time-lapse microscopy experiments. While we develop the models for the case of ribosomes, the same reasoning applies to other protein categories, in particular enzymes in amino acid and energy metabolism.

3.6.2.1 Resource allocation model

Let $\mathrm{Vol}(t)$ [L] be the volume of the growing bacterial population at time t [h]. We assume that the volume is proportional to the biomass, that is, that the biomass density is constant, an assumption justified by steady-state data over a large range of growth rates in *E. coli* [158]. Moreover, we approximate the biomass by the protein mass, its major component [181, 182]. Accordingly, we define

$$\mathrm{Vol}(t) = \beta\,(R(t) + P(t)), \tag{3.15}$$

where R [g] is the total mass of ribosomes in the growing bacterial population and P [g] the total mass of other proteins. $1/\beta$ [g/L] is the biomass density or, equivalently in this case, the total protein concentration.

Let $V_{prot}(t)$ be the total protein synthesis rate in the growing bacterial population, with units g/h. We now introduce a resource allocation strategy $\alpha(t)$ that attributes a time-varying fraction (between 0 and 1) of V_{prot} to the synthesis of ribosomal proteins. By definition, $1-\alpha(t)$ is the resource allocation strategy for the other proteins.

The following system of differential equations describes the dynamics of the protein masses:

$$\frac{d}{dt}R(t) = \alpha(t)\, V_{prot}(t) - \gamma\, R(t), \tag{3.16}$$

$$\frac{d}{dt}P(t) = (1-\alpha(t))\, V_{prot}(t) - \gamma\, P(t), \tag{3.17}$$

where γ [1/h] is a degradation constant. Most half-lives of proteins are long (>10 h) [183, 184] and we assume that they are similar across the different protein categories considered. We complete the model with the dynamics of the volume of the growing bacterial population:

$$\frac{d}{dt}\mathrm{Vol}(t) = \mu(t)\,\mathrm{Vol}(t). \tag{3.18}$$

The variables in the above model are extensive quantities, computed over the entire growing population. For our purpose, it is more convenient to define models based on intensive quantities, that is, to normalize the protein masses by the population volume. Accordingly, we define protein concentrations $r = R/\mathrm{Vol}$ and $p = P/\mathrm{Vol}$, with units g/L, and the specific protein synthesis rate $v_{prot} = V_{prot}/\mathrm{Vol}$, with unit g/L/h.

The above definitions allow, first of all, to rewrite the growth rate as the net (specific) protein synthesis rate. Using the definition of the growth rate (Eq. 3.18) and the proportionality of biomass and volume (Eq. 3.15), we find that

$$\begin{aligned}\mu(t) &= \frac{\beta\left(\frac{d}{dt}R(t) + \frac{d}{dt}P(t)\right)}{\mathrm{Vol}(t)} = \beta\left(\frac{V_{prot}(t)}{\mathrm{Vol}(t)} - \gamma\frac{R(t)+P(t)}{\mathrm{Vol}(t)}\right)\\ &= \beta\, v_{prot}(t) - \gamma,\end{aligned} \tag{3.19}$$

that is, the growth rate equal the total protein synthesis rate relative to the total protein concentration $1/\beta$, minus the protein degradation rate.

Second, using Eqs 3.16-3.17, as well as the definition of the growth rate, the dynamics of the protein concentrations can be written as follows:

$$\frac{d}{dt}r(t) = \frac{\frac{d}{dt}R(t)\,\mathrm{Vol}(t) - R(t)\,\frac{d}{dt}\mathrm{Vol}(t)}{\mathrm{Vol}(t)^2} = \frac{\frac{d}{dt}R(t)}{\mathrm{Vol}(t)} - \mu(t)\,\frac{R(t)}{\mathrm{Vol}(t)}$$
$$= \alpha(t)\,v_{prot}(t) - (\mu(t)+\gamma)\,r(t), \tag{3.20}$$
$$\frac{d}{dt}p(t) = \frac{\frac{d}{dt}P(t)\,\mathrm{Vol}(t) - P(t)\,\frac{d}{dt}\mathrm{Vol}(t)}{\mathrm{Vol}(t)^2} = \frac{\frac{d}{dt}P(t)}{\mathrm{Vol}(t)} - \mu(t)\,\frac{P(t)}{\mathrm{Vol}(t)}$$
$$= (1-\alpha(t))\,v_{prot}(t) - (\mu(t)+\gamma)\,p(t). \tag{3.21}$$

Note that when defining the protein dynamics on the concentration level, protein synthesis is balanced by degradation and a term accounting for growth dilution [99]. The model can be further simplified by taking into account the proportionality of the protein synthesis rate with the sum of the growth and degradation rates (Eq. 3.19):

$$\frac{d}{dt}r(t) = (\mu(t)+\gamma)\left(\frac{\alpha(t)}{\beta} - r(t)\right), \tag{3.22}$$
$$\frac{d}{dt}p(t) = (\mu(t)+\gamma)\left(\frac{(1-\alpha(t))}{\beta} - p(t)\right). \tag{3.23}$$

The interest of this model lies in that it relates the resource allocation strategies to the concentrations of the different protein categories. In particular, when knowing the degradation rate and observing the time-varying growth rate and ribosomal protein concentration, we can estimate α/β. In general, no reliable value for β will be available, so that we cannot compute an absolute value for the resource allocation strategy $\alpha(t)$. When we are only interested in changes in $\alpha(t)$, however, an estimate of the scaled strategy α/β is sufficient.

3.6.2.2 Measurement model for the inference of resource allocation strategies

The observation of the ribosome concentration is indirect, involving the fusion of a fluorescent reporter protein with a representative ribosomal subunit, RpsB.

The reporter protein has its own dynamics, in particular due to maturation of the fluorophore and photobleaching [93], which requires the resource allocation model of Eqs 3.22-3.23 to be coupled with a dedicated measurement model. In this section, we extend the model to account for maturation effects. The effects of photobleaching were found to be negligible in our conditions (*Materials and methods*), so we ignored these in the analysis. While we focus on the case of fluorescent tags of ribosomes, the approach is generic and applies to other protein categories as well, in particular enzymes in amino acid metabolism and energy metabolism. The measurement models are based on previous work on the maturation of fluorescent proteins [93, 159].

In the case of green fluorescent protein (GFP), the maturation dynamics adds a second equation to the model:

$$\frac{d}{dt}r_{im}(t) = (\mu(t)+\gamma)\left(\frac{\alpha(t)}{\beta} - r_{im}(t)\right) - k_{mat}\, r_{im}(t), \tag{3.24}$$

$$\frac{d}{dt}r_{m}(t) = k_{mat}\, r_{im}(t) - (\mu(t)+\gamma)\, r_{m}(t), \tag{3.25}$$

where r_{im} and r_m refer to the concentrations of ribosomes tagged with immature and mature GFP, respectively [g/L]. The total ribosome concentration is given by the sum of the concentrations of mature and immature fusion proteins $(r_{im} + r_m)$. Maturation is modeled as a first-order reaction with a constant k_{mat} [1/h]. The stability of ribosomes with both the mature and the immature fluorescent tags is assumed to be the same, as determined by the degradation constant γ [1/h].

The units of the ribosome concentrations in the above model are taken to be gram per liter. The observations of fluorescent proteins in the fluorescence microscopy images are in different units: namely Relative Fluorescence Units (RFU) per pixel (*Materials and methods*). We make the assumption that the two scales are linearly related. Multiplying the concentration variables r_{im} and r_m, as well as the total protein concentration $1/\beta$, by a scaling factor θ [RFU/pixel L/g] changes the concentration units, but not the structure of the model of Eqs 3.24-3.25. In order to avoid unnecessarily complex notation,

we keep the same variable names in the original and rescaled models (and distinguish the latter by specifying the concentration units).

The model of Eqs 3.24-3.25 also applies to yellow fluorescent proteins (YFP) and blue fluorescent proteins (CFP), which have maturation kinetics similar to GFP [93]. In the case of red fluorescent proteins (RFP), the maturation kinetics are more complicated [93]. Instead of the direct conversion of an immature (colorless) protein to a mature (green, yellow, blue) protein, the immature is transformed into a mature red protein *via* the formation of a blue intermediate absorbing at around 400 nm [139, 140]. This gives rise to an additional equation in the maturation model [159], resulting in

$$\frac{d}{dt}r_{im}(t) = (\mu(t)+\gamma)\left(\frac{\alpha(t)}{\beta} - r_{im}(t)\right) - k_{mat}\, r_{im}(t), \tag{3.26}$$

$$\frac{d}{dt}r_{hm}(t) = k_{mat}\, r_{im}(t) - k'_{mat}\, r_{hm}(t) - (\mu(t)+\gamma)\, r_{hm}(t), \tag{3.27}$$

$$\frac{d}{dt}r_{m}(t) = k'_{mat}\, r_{hm}(t) - (\mu(t)+\gamma)\, r_{m}(t). \tag{3.28}$$

We now distinguish between ribosomes with immature, half-mature, and mature fluorescent tags, with concentrations r_{im}, r_{hm}, and r_m, respectively. The model has two first-order maturation reactions, with constants k_{mat} and k'_{mat}, accounting for the transformation of immature proteins into half-mature (blue) proteins, and half-mature proteins into mature (red) proteins, respectively.

As explained in the *Materials and methods*, we have carried out targeted calibration experiments in this and previous work [159] to determine the values of the maturation and degradation constants. The results are shown in Tab. 3.3. The models of Eqs 3.24-3.25 and Eqs 3.26-3.28 were used for the inference of the resource allocation strategies as described in the main text.

3.6.3 Text S2: Estimation

This supplementary text describes the methods developed for the inference of a time-varying resource allocation profile from single-cell time-course measurements of fluorescent reporter intensities and cell lengths. The method relies on the models presented in the *Materials and methods* and Text S1, where the dynamics of reporter abundance $r_m(t)$ depends on the unknown resource allocation profile $\alpha(t)$ and the growth rate $\mu(t)$. Our approach relies on estimating $\mu(t)$ from cell length data first, and then using this estimate in Eqs 3.3-3.4 or Eqs 3.5-3.7 in the main text to reconstruct the profile of α from fluorescence data. Our methods apply to so-called mother cells sitting at the dead-end of a mother machine channel and observed over a large number of consecutive generations. Before describing the estimation methods, in a first section we discuss the relation between the single-cell measurements and the quantities $r_m(t)$ and $\mu(t)$, and establish some notation. Leveraging regularization [123, 124], the estimation methods are discussed in the next two sections, with subsections discussing the automated choice of regularization parameters and the application of the methods to data from experiments where the growth medium is periodically switched.

3.6.3.1 Relation between single-cell models and experimental observations

After image processing (*Materials and methods*), we have a series of length measurements (a sawtooth-type profile because of cell growth and divisions, see Fig. 3.1A-B) along with the mean fluorescence intensities over the observed cell area. For $i = 1, \ldots, m$, let $t_{i,1}, t_{i,2}, \ldots, t_{i,n_i}$ be the n_i measurement times in-between two subsequent cell divisions, with $t_{i,j} < t_{i,j+1}$ and $t_{i,n_i} < t_{i+1,1}$ for all relevant i and j. This partitioning (corresponding to $m-1$ divisions) was obtained from the cell length measurement sequence by the segmentation procedure of BACMMAN (*Materials and methods*. We assume that $n_i \geq 2$ for all i, that is, that we have at least two length measurements per generation. We denote with $n = n_1 + \ldots + n_m$ the total number of measurement times for one mother cell. Correspondingly, let $L_{i,j}$ and $f_{i,j}$ be the cell length and the

mean fluorescence intensity of the cell, respectively, as at time $t_{i,j}$, obtained from post-processing the image analysis results (*Materials and methods*). To relate these measurements to the model quantities $\mu(t)$ and $r_m(t)$, we make the following assumptions:

1. $L_{i,j}$ is a noisy observation of the actual cell length $L(t)$ at time $t = t_{i,j}$, where, in-between divisions, the dynamics of $L(t)$ relate to growth rate by the model
$$\frac{d}{dt}L(t) = \mu(t)\, L(t), \tag{3.29}$$
2. The mean intensity $f_{i,j}$ is a noisy measurement of the reporter abundance $r_m(t)$ at time $t = t_{i,j}$.

Note that no claim is made in Assumption 1 concerning the cell length before (L_{i,n_i}) and after ($L_{i+1,1}$) a division. Besides possible deviations from a symmetric division, this is because the precise division time is generally not known (it occurs between two measurement times). Also note that Assumption 2 is (up to a scaling factor) in accordance with the interpretation of $r_m(t)$ as an intensive variable, that is, the reporter concentration within the given (growing) cell at time t, as per model Eqs 3.3-3.4 or Eqs 3.5-3.7 in the *Materials and methods*. Further mathematical characterization of Assumptions 1–2 is made in the appropriate sections below.

3.6.3.2 Growth rate estimation from cell-length data

The method developed below (see Fig. 3.12 as a reference) works on the logarithm of cell length. That is, we consider the data $\tilde{\ell}_{i,j} = \log L_{i,j}$, which are noisy measurements of $\ell(t) = \log L(t)$. For a constant growth rate μ, in-between divisions, $\ell(t)$ sits on a straight line. We instead let $\mu(t)$ change over time even in-between divisions. In particular, we assume that the growth rate between two consecutive measurements is well approximated by a constant, that is, for some $\mu_{i,j}$, $\mu(t) \simeq \mu_{i,j}$, for all $t \in [t_{i,j}, t_{i,j+1})$. For every i and some initial value ℓ_i, using the notation $\delta_{i,j} = t_{i,j+1} - t_{i,j}$ we can now construct the

piecewise-linear model

$$\begin{aligned}
\ell(t_{i,1}) &= \nu_i, \\
\ell(t_{i,2}) &= \mu_{i,1}\delta_{i,1} + \ell(t_{i,1}) = \mu_{i,1}\delta_{i,1} + \nu_i, \\
\ell(t_{i,3}) &= \mu_{i,2}\delta_{i,2} + \ell(t_{i,2}) = \mu_{i,2}\delta_{i,2} + \mu_{i,1}\delta_{i,1} + \nu_i, \\
&\vdots \\
\ell(t_{i,n_i}) &= \mu_{i,n_i-1}\delta_{i,n_i-1} + \ell(t_{i,n_i-1}) = \mu_{i,n_i-1}\delta_{i,n_i-1} + \ldots + \mu_{i,1}\delta_{i,1} + \nu_i.
\end{aligned} \tag{3.30}$$

Note that $\mu_{i,j}$ (that is, the growth rate) is allowed to change from one measurement to the next. Growth rate variations in-between measurements could also be considered (at the price of a more complex inference problem), but this is not justified in our case given the high sampling density.

For all i and j, the model unknowns to be inferred from the data $\tilde{\ell}_{i,j}$ are the variable growth rates, $\mu_{i,j}$, and (of less interest here) the cell lengths at the first measurement times after divisions, ν_i. As such, this is an underdetermined problem. To cope with this and with measurement noise, we resort to regularized least squares, a well-established method in the statistical literature [124]. Let us arrange the model parameters into vectors $\underline{\nu} = [\nu_1, \ldots, \nu_m]^T$ and $\underline{\mu} = [\underline{\mu}_1^T, \ldots, \underline{\mu}_m^T]^T$, where $\underline{\mu}_i = [\mu_{i,1}, \ldots, \mu_{i,n_i}]^T$, with $i = 1, \ldots, m$ (for every i, entry μ_{i,n_i}, which does not appear in Eq. 3.30, should be understood as the growth rate after the last measurement before cell division). Note in particular that the entries of $\underline{\mu}$ are arranged according to time order. We seek parameters $\underline{\nu}$ and $\underline{\mu} \geq 0$ that minimize

$$(1-\lambda) \cdot \sum_{i,j} \left(\tilde{\ell}_{i,j} - \Delta_{i,j}^T \cdot \underline{\mu}_i - \nu_i\right)^2 + \lambda \cdot Q(\underline{\mu}), \tag{3.31}$$

where, for $\Delta_{i,j}^T = [\delta_{i,1}, \ldots, \delta_{i,j-1}, 0, \ldots, 0]$ of size n_i, with $j = 1, \ldots, n_i$ and $i = 1, \ldots, m$, the term $\Delta_{i,j}^T \cdot \underline{\mu}_i + \nu_i$ denotes the predictions of $\ell(t_{i,j})$ obtained from the model of Eq. 3.30 for putative values of $\underline{\nu}$ and $\underline{\mu}$. Term $Q(\underline{\mu}) \geq 0$ denotes a suitable cost function chosen so as to penalize unrealistic fluctuations in the growth rate time-series $\underline{\mu}$, and $\lambda \in (0, 1)$ is a regularization parameter that trades off perfect data fit ($\lambda \simeq 0$) with penalization of fluctuating

solutions ($\lambda > 0$). The discussion of the automated choice of λ from data is postponed to a dedicated section below. Well-established choices of Q correspond to penalizing the sum of squares of the dth (discrete-time) derivative of $\underline{\mu}$ (regarded as a time series), with customary choices $d = 1$ or 2 [124]. Discrete differentiation is expressed by a matrix operation $D_d \cdot \underline{\mu}$, where D_d is an $(n - d) \times n$ matrix. For $d = 1$ (our choice), the jth row of D_1 has the only nonzero entries 1 and -1 in positions j and $j + 1$.

Importantly, for such choice of Q, minimization of Eq. 3.31 is a linear least squares problem with bound constraints. For $\underline{\tilde{\ell}} = [\tilde{\ell}_1^T, \ldots, \tilde{\ell}_m^T]^T$, where $\tilde{\ell}_i = [\tilde{\ell}_{i,1}, \ldots, \tilde{\ell}_{i,n_i}]^T$ with $i = 1, \ldots, m$, the estimates sought, denoted by $\underline{\hat{\nu}}$ and $\underline{\hat{\mu}}$, are given by

$$(\underline{\hat{\nu}}, \underline{\hat{\mu}}) = \arg \min_{\underline{\nu}, \underline{\mu} \geq 0} \left\| \begin{bmatrix} \sqrt{(1-\lambda)} \cdot \underline{\tilde{\ell}} \\ 0 \end{bmatrix} - \begin{bmatrix} \sqrt{(1-\lambda)} \cdot \Delta & \sqrt{(1-\lambda)} \cdot L \\ \sqrt{\lambda} \cdot D_1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \underline{\mu} \\ \underline{\nu} \end{bmatrix} \right\|^2 \tag{3.32}$$

where Δ is a block-diagonal matrix with blocks defined by $\Delta_i = [\Delta_{i,1}, \ldots, \Delta_{i,n_i}]^T$, $i = 1, \ldots, m$, and L is a block-diagonal matrix with blocks defined by the size-n_i column vectors $L_i = [1, \ldots, 1]^T$, $i = 1, \ldots, m$. Thanks to regularization, the solution of this problem is well-defined, and we solve it via the Python function `lsq_linear` of the `scipy.optimize` module. For a fixed value of λ and a given dataset $\underline{\tilde{\ell}}$ with 300 entries (our case), optimization takes less than 10^{-2} seconds.

Automatic tuning of regularization parameters

To fix an appropriate value for λ directly from the data, we used a cross-validation approach. The rationale behind cross-validation is to choose λ such that the resulting estimates are predictive, that is, estimates of $\underline{\mu}$ and $\underline{\nu}$ predict well data that have not been used for their calculation [126]. Among several possible alternatives, we choose a procedure where for every mother cell c, the corresponding data sequence $\tilde{\ell}$ is partitioned in two data sets, one for estimation and one for validation. This is expressed by the partitioning of indices (i, j) into two index sets, I^E and I^V. In particular we choose $I^E = \{(i, j) : \ j \text{ odd}\}$ and $I^V = \{(i, j) : \ j \text{ even}\}$, but we keep the description below general. For simplicity we avoid appending a mother cell index to the following

developments, until this becomes necessary.

For any putative value of $\lambda \in (0,1)$, consider the problem of estimating growth rates from estimation data $\{\tilde{\ell}_{i,j} : (i,j) \in I^E\}$ by the procedure above. For any i, let $j_1, j_2, \ldots$ be the increasing sequence of indices j such that $(i,j) \in I^E$. The generic relation in Eq. 3.30 now becomes

$$\ell(t_{i,j_h}) = \nu_i + \sum_{k=1,\ldots,h-1} \mu_{i,j_k}(t_{i,j_{k+1}} - t_{i,j_k}). \tag{3.33}$$

Overall, this reduces the growth rate unknowns to the set $\{\mu_{i,j} : (i,j) \in I^E\}$, which we arrange in vector form as $\underline{\mu}^E$ (according to increasing time order). Provided corresponding adaptation of matrices Δ, L and D_1, for any value of λ, the solution of Eq. 3.32 yields estimates $\hat{\underline{\mu}}^E(\lambda)$ and $\hat{\underline{\nu}}^E(\lambda)$. In accordance with Eq. 3.33, this allows one to define estimates $\hat{\ell}(t|\lambda)$ of ℓ as a piecewise-linear function of t. Precisely, for any i and $t \in [t_{i,1}, t_{i+1,1})$,

$$\hat{\ell}^E(t|\lambda) = \hat{\nu}_i^E(\lambda) + \left(\sum_{k=1,\ldots,h-1} \hat{\mu}_{i,j_k}^E(\lambda) \cdot (t_{j_{k+1}} - t_{j_k}) \right) + \hat{\mu}_{i,j_h}^E(\lambda) \cdot (t - t_{j_h}), \tag{3.34}$$

where j_h is the largest index j such that $(i,j) \in I^E$ and $t_{i,j_h} \leq t$. We then define the error in the prediction of validation data $\{\tilde{\ell}_{i,j} : (i,j) \in I^V\}$ as

$$e(\lambda) = \sum_{(i,j) \in I^V} \left(\tilde{\ell}_{i,j} - \hat{\ell}^E(t_{i,j}|\lambda)\right)^2. \tag{3.35}$$

Armed with these definitions, separately for every mother cell c in a given experiment, we calculate the best value of λ as

$$\hat{\lambda}^c = \arg\min_{\lambda} e^c(\lambda), \tag{3.36}$$

where the superscript in e^c recalls the dependence of Eq. 3.35 on the mother cell c. This yields a set of estimates $\hat{\lambda}^c$, one per mother cell c. Finally, we establish our choice of λ by taking the median of the $\hat{\lambda}^c$ values. Taking the median makes our final choice $\hat{\lambda}$ robust to minor residual uncertainties associated with the pipeline. In practice, we solve Eq. 3.36 with the Python function

`minimize` of the `scipy.optimize` module. To avoid numerical optimization being trapped into local minima, we restrict optimization to a reasonable interval of λ determined by visual inspection. The execution of the whole procedure over 100 cells and $n = 300$ data-points (our case) takes around 5 minutes.

Application growth medium switches

The regularization parameter $\hat{\lambda}$ determined as explained above was inserted into Eq. 3.32 to estimate growth rate profiles $\hat{\mu}^c$ for the various mother cells from the corresponding length sequences $\tilde{\ell}^c$. However, regularization in Eq. 3.32 implicitly assumes that no prior information exists on where the most relevant changes in the growth rate time-series $\underline{\mu}$ occur. For cells exposed to growth medium switches, instead, non-smooth growth rate changes are expected to occur after the switching times. This suggests a slight modification of length the definition of the regularization term. If $t_{i,j}$ is the first measurement time after a medium switch, we redefine D_1 by simply removing the rows corresponding to indexes (i, j) and $(i, j+1)$. As a consequence, a non-smooth change in growth rate right after the medium switch is not penalized. With this definition and the $\hat{\lambda}$ from the previous section, the calculation of estimates $\underline{\hat{\mu}}^c$ (and $\underline{\hat{\nu}}^c$) from $\underline{\tilde{\ell}}^c$ via the optimization Eq. 3.32 is otherwise unchanged.

3.6.3.3 Estimation of resource allocation profiles from fluorescence data

The description below refers to the estimation problem for a single mother cell c. For simplicity, we omit dependence on c from the notation, until this becomes needed. Estimation of the unknown resource allocation profile $\alpha(t)$ relies on the model of Eqs 3.3-3.4 in the main text. This is an ODE model for the concentrations of immature and mature reporter proteins ($r_{im}(t)$ and $r_m(t)$) in a single cell. The model applies to the green, yellow and cyan reporter proteins. The method can be straightforwardly generalized to the case of red proteins, described by the model of Eqs 3.5-3.7 in the main text.

As explained in Sec. 3.6.3.1, the analysis of single-cell microfluidics data yields noisy measurements of $r_m(t)$ at times $t_{i,j}$. Because the evolution of r_m is

assumed to be smooth (no jumps following cell division), *i.e.,* index i plays no special role, we simplify the indexing of data and consider that measurements are taken at (increasing) times t_k, with $k = 1, \ldots, n$. More precisely, we assume that noisy measurements y_k obey

$$y_k = r_m(t) + \varepsilon_k, \tag{3.37}$$

where the measurement errors ε_k form a sequence of independent random variables with mean zero and variance s_k^2. We assume that the variance can be estimated from the data (*Materials and methods*).

Let $x(t) = [r_{im}(t), r_m(t)]^T$. The model of Eqs 3.3-3.4 can be written in the form

$$\dot{x}(t) = F(t)x(t) + G(t)\alpha(t), \tag{3.38}$$

where $\alpha(t)$ is the unknown profile to be inferred from $y_1, \ldots, y_n$. Matrices $F(t)$ and $G(t)$ depend on parameters γ and k_{mat} whose values are known (Tab. 3.3), β, which only appears as a rescaling of $\alpha(t)$, and the growth rate profile $\mu(t)$. The parameter β cannot be determined from the data, so we drop it from the model and bear in mind that estimation of $\alpha(t)$ below is in fact estimation of $\alpha(t)/\beta$. For $\mu(t)$ known, Eq. 3.38 is a linear state-space model. To leverage this property, we pretend that $\mu(t)$ is known. In practice $\mu(t)$ will be replaced by estimates $\hat{\mu}(t)$ obtained by interpolation of the vector estimate $\underline{\hat{\mu}}$ of Sec. 3.6.3.2.

Reconstruction of $\alpha(t)$ from sampled, noisy measurements $y_1, ..., y_n$ is an ill-posed problem. To obtain relevant, robust estimates we resort again to regularization. Here, the standard, linear state-space form of model 3.37–3.38 prompts us to pose estimation as a Bayesian problem, and solve it with effective tools based on Kalman filtering [132]. Similar to [159], we introduce a probabilistic prior on α in the form of the Ornstein-Uhlenbeck process

$$d\alpha(t) = -\theta \cdot \alpha(t)dt + \sigma \cdot dW(t), \tag{3.39}$$

where $W(t)$ is a standard Wiener process. We additionally assume that $\alpha(0)$ has a zero-mean Gaussian distribution with variance $\sigma^2/(2\theta)$, which implies that the process is stationary. In simple words, this prior introduces information about the profile sought by assigning different probabilities to different

profiles, depending on parameters θ and σ. Large (smaller) values of the ratio $\sigma/\sqrt{\theta}$ favor larger (smaller) fluctuations of α. Large (resp. smaller) values of θ favor faster (slower) fluctuations. This information is quantitatively conveyed by the autocovariance function of the process [141], which has the form

$$\mathbb{E}[\alpha(t)\alpha(\tau)] = \sigma^2/(2\theta) \cdot \exp(-\theta \cdot |t - \tau|), \tag{3.40}$$

for any two times t, τ, where the first factor is the process variance and the second factor shows the exponentially decaying "memory" of the process. We postpone the illustration of the automated choice of θ and σ, which play the role of regularization parameters, to Section 3.6.3.3 below.

Given the probabilistic prior, the problem of calculating an optimal estimate $\hat{\alpha}(t)$ for $\alpha(t)$ given data $y_1, \ldots, y_n$ becomes the Bayesian problem of calculating the conditional expectation

$$\hat{\alpha}(t) = \mathbb{E}[\alpha(t)|y_1, \ldots, y_n]. \tag{3.41}$$

This reconstruction is optimal in the sense that it minimizes the variance of the estimation error at any time t [141]. To compute the right-hand side of Eq. 3.41, for the sake of simplicity, we restrict attention to estimates $\hat{\alpha}(t)$ at times t_k, and leverage linearity properties of Eqs 3.37–3.39 to recast the problem into the form of an efficient computation known as Kalman smoothing [142]. Define the augmented state $\xi(t) = [x(t)^T \ \alpha(t)]^T$. From the previous relations one can write the linear stochastic state-space model

$$\begin{aligned} d\xi(t) &= \begin{bmatrix} F(t) & G(t) \\ 0 & -\theta \end{bmatrix} \xi(t)dt + \begin{bmatrix} 0 \\ \sigma \end{bmatrix} dW(t), \\ y_k &= C\xi(t_k) + \varepsilon_k, \end{aligned} \tag{3.42}$$

where the zeros denote null matrix blocks of appropriate dimensions, and $C = [0, 1, 0]$. For generic indices define

$$\hat{\xi}_{k|h} = \mathbb{E}[\xi(t_k)|y_1, \ldots, y_h], \qquad P_{k|h} = \mathbb{E}[\big(\xi(t_k) - \hat{\xi}_{k|h}\big)\big(\xi(t_k) - \hat{\xi}_{k|h}\big)^T],$$

where $\hat{\xi}_{k|0}$ and $P_{k|0}$ are in particular the prior mean and covariance matrix of $\xi(t_k)$.

The Kalman filter is an forward iteration to calculate $\hat{\xi}_{k|k}$ (and $\hat{\xi}_{k|k-1}$) along with $P_{k|k}$ (and $P_{k|k-1}$) for k from 1 to n. On top of this, the Kalman smoother is obtained by further running a backward iteration that calculates $\hat{\xi}_{k|n}$ along with $P_{k|n}$ for k from n down to 1. This provides estimates of all entries of ξ. In particular, an estimate $\hat{\alpha}(t)$ at time t_k and the corresponding estimation error variance is then obtained by the equations

$$\hat{\alpha}(t_k) = S\hat{\xi}_{k|k}, \qquad \mathbb{E}[\big(\hat{\alpha}(t_k) - \alpha(t_k)\big)^2] = SP_{k|n}S^T, \tag{3.43}$$

with $S = [0, 0, 1]$. As the terminology suggests, $\hat{\alpha}$ will display smooth dynamics, which agrees with expectations except at times where the culture medium is switched. The adaptations of the approach to cope with this are described in Sec. 3.6.3.3 below.

The Kalman filtering equations are standard. For every k, two steps are performed. The prediction step consists in calculating $\hat{\xi}_{k+1|k}$ and $P_{k+1|k}$ from $\hat{\xi}_{k|k}$ and $P_{k|k}$ by the solution of the system dynamics given by 3.42 and a corresponding Riccati equation from time t_k to t_{k+1}. The update step consists in integrating the information from the new measurement y_{k+1} to get $\hat{\xi}_{k+1|k+1}$ and $P_{k+1|k+1}$ from $\hat{\xi}_{k+1|k}$ and $P_{k+1|k}$ via simple algebraic operations involving Eq. 3.37. This iteration is performed for $k = 1, \ldots, n-1$, starting from $\hat{\xi}_{1|0}$ and $P_{1|0}$ (at $k = 1$, the prediction step is skipped). The less known smoothing iteration is also standard, and consists purely in algebraic calculations on data and quantities calculated in the forward filtering iteration. See detailed equations in *e.g* [134, 142]. The choice of initial conditions $\hat{\xi}_{1|0}$ and $P_{1|0}$ (the prior on $\xi(t_1)$) is determined in part by the prior on α. We set $\hat{\xi}_{1|0} = [0, 0, 0]^T$ and $P_{1|0} = \mathrm{diag}\big(M, M, \sigma^2/(2\theta)\big)$, with M large (10^6), corresponding to virtual lack of information on r_{im} and r_m prior to the experiments.

The computational complexity of the whole procedure is linear in the number of data points n, and is mostly determined by the inversion of small (3×3) matrices. For a single mother cell data series with $n = 300$ points (our case), our Python implementation using function `odeint` of the `scipy.integrate` module for numerical integration, takes 10 seconds. If the Gaussian assumptions are violated, the estimates computed by this procedure have the interpretation of optimal estimates (minimal error variance) in the class of linear

functions of the data [142]. Thanks to this, the method is robust to moderate deviations from Gaussianity. Computational efficiency of the method is especially important for the automated tuning of parameters θ and σ, as will be explained next.

Automatic tuning of regularization parameters

To fix the prior parameters we follow a maximum-likelihood approach. That is, we seek parameters θ and σ that maximize

$$f_{\theta,\sigma}(y_1,\ldots,y_n) = f_{\theta,\sigma}(y_1)\cdot\prod_{k=2}^{n} f_{\theta,\sigma}(y_k|y_1,\ldots,y_{k-1}),$$

where the notation indicates probability densities under candidate values of θ and σ, and the factorization of the joint density of the data is a consequence of Bayes' law. Maximization of the likelihood is equivalent to minimization of the negative log-likelihood. Under Gaussianity of the prior and of measurement noise, from the equation above and Eq. 3.37, the negative log-likelihood is, up to constant additive terms, given by

$$\mathscr{L}(\theta,\sigma|y_1,\ldots,y_n) = \sum_{k=1}^{n}\log\Lambda_k(\theta,\sigma) + \frac{1}{2}\left[\frac{y_k-\hat{y}_{k|k-1}(\theta,\sigma)}{\Lambda_k(\theta,\sigma)}\right]^2,$$

where $\hat{y}_{k|k-1}(\theta,\sigma)$ and $\Lambda_k(\theta,\sigma)$ are respectively the mean and variance of the conditional distribution $f_{\theta,\sigma}(y_k|y_1,\ldots,y_{k-1})$, with $k=2,\ldots,n$, while $\hat{y}_{1|0}(\theta,\sigma)$ and $\Lambda_1(\theta,\sigma)$ are those of $f_{\theta,\sigma}(y_1)$. It follows from the Kalman filtering theory that $\hat{y}_{k|k-1}(\theta,\sigma) = C\hat{\xi}_{k|k-1}(\theta,\sigma)$ and $\Lambda_k(\theta,\sigma) = CP_{k|k-1}(\theta,\sigma)C^T + s_k^2$, with $k=1,\ldots,n$, where $\hat{\xi}_{k|k-1}(\theta,\sigma)$ and $P_{k|k-1}(\theta,\sigma)$ are the one-step state predictions and covariance matrices obtained in the Kalman filtering iteration of Sec. 3.6.3.3 run for the given values of θ and σ [142].

By the machinery above, for every mother cell c, we obtain estimates of θ and σ by the optimization

$$(\hat{\theta}^c,\hat{\sigma}^c) = \arg\min_{\theta>0,\sigma>0}\mathscr{L}(\theta,\sigma|y_1^c,\ldots,y_n^c),$$

where $y_1^c, \ldots, y_n^c$ is the fluorescence time-series for mother cell c. In practice, optimization is solved numerically by the Python function `minimize` of the `scipy.optimize` module. In this optimization, objective function $\mathscr{L}(\theta, \sigma | y_1^c, \ldots, y_n^c)$ is evaluated at every explored value of θ and σ by running the corresponding Kalman filter iteration of Sec. 3.6.3.3 on data $y_1^c, \ldots, y_n^c$. In total, in a typical experiment with 100 cells and for $n = 300$, the computation takes around 2 hours.

Finally, our choice for parameters θ and σ is defined by the median of the values $\hat{\theta}^c$ and $\hat{\sigma}^c$ obtained over all mother cells c. This single choice, $\hat{\theta}$ and $\hat{\sigma}$, is plugged into the method of Sec. 3.6.3.3, which is then applied separately to all cells c in an experiment to get estimates $\hat{\alpha}^c$ of the resource allocation profile from the corresponding fluorescence time-series $y_1^c, \ldots, y_n^c$.

Application to growth medium switches

For the estimation of resource allocation profiles over switches of the culture medium, considerations similar to Section 3.6.3.2 apply. At switching times, resource allocation may change over a short time interval. Smooth reconstruction of α at these times may therefore be artifactual. In essence, this is a consequence of the stationarity of the prior autocovariance function Eq. 3.40, which stipulates that the dependence between a "current" value $\alpha(t)$ and its "future" evolution $\alpha(\tau)$ $(\tau > t)$ is only a function of the time-lag $\tau - t$, irrespective of the specific value of t. Whereas this uniform behavior is relevant for a given medium, the relation between $\alpha(t)$ and $\alpha(\tau)$ is expected to be looser across a medium switch due to changed conditions.

To cope with this, rather than introducing arbitrary (and technically cumbersome) local changes in the prior on α, we separate out the reconstruction problem over different growth conditions. The procedure applies equally to all mother cells c in the analysis, and is the following.

Let q_h, with $h = 1, \ldots, M-1$, be the (increasing sequence of) indices such that t_{q_h} are the measurement times right before (or equal to) the time of a medium switch. M is the number of medium switches over the whole experimental horizon. Note that these times are unrelated with cell division times. Also let $q_M = n$. With the parameters λ and θ calibrated as per

Sec. 3.6.3.3:

- Starting from the prior on $\xi(t_1)$, $\hat{\xi}_{1|0}$ and $P_{1|0}$, defined in Sec. 3.6.3.3, run the Kalman filering-smoothing procedure of Sec. 3.6.3.3 over indices $(1, \ldots, q_1)$. This yields estimates $\hat{\xi}_{1|q_1}, \hat{\xi}_{2|q_1}, \ldots, \hat{\xi}_{q_1|q_1}$ and estimation error covariance matrices $P_{1|q_1}, P_{2|q_1}, \ldots, P_{q_1|q_1}$.

- For $h = 1, \ldots, M-1$: define $\hat{\xi}_{q_h+1|q_h} \triangleq \hat{\xi}_{1|0}$ and $P_{q_h+1|q_h} \triangleq P_{1|0}$ as the prior on $\xi(t_{q_h+1})$, and run the Kalman filering-smoothing procedure of Sec. 3.6.3.3 over indices $(q_h + 1, \ldots, q_{h+1})$. This yields estimates $\hat{\xi}_{q_h+1|q_{h+1}}, \hat{\xi}_{q_h+2|q_{h+1}}, \ldots, \hat{\xi}_{q_{h+1}|q_{h+1}}$ and estimation error covariance matrices $P_{q_h+1|q_h+1}, P_{q_h+2|q_h+1}, \ldots, P_{q_{h+1}|q_{h+1}}$.

- Return the whole sequence of estimates

  $$(\hat{\xi}_{1|q_1}, \ldots, \hat{\xi}_{q_1|q_1}), \ldots, (\hat{\xi}_{q_h+1|q_{h+1}}, \ldots, \hat{\xi}_{q_{h+1}|q_{h+1}}), \ldots, (\hat{\xi}_{q_{M-1}+1|q_M}, \ldots, \hat{\xi}_{q_M|q_M})$$

  and corresponding estimation error covariance matrices

  $$(P_{1|q_1}, \ldots, P_{q_1|q_1}), \ldots, (P_{q_h+1|q_{h+1}}, \ldots, P_{q_{h+1}|q_{h+1}}), \ldots, (P_{q_{M-1}+1|q_M}, \ldots, P_{q_M|q_M}).$$

Crucially, in every period h, only measurements from the same period are used (which breaks in particular the dependence from the future dynamics, exposed to sudden changes). A demonstration of the effectiveness of this approach was given in the validation studies of the approach (*Materials and methods* and Fig. 3.19).

Chapter 4

Discussion

"Time has two aspects. There is the arrow, the running river, without which there is no change, no progress, or direction, or creation. And there is the circle or the cycle, without which there is chaos, meaningless succession of instants, a world without clocks or seasons or promises."

Ursula K. Le Guin

4.1 Concluding remarks

Microbiology is rapidly evolving into a transdisciplinary field. Now, the aim is not only to describe, but also to quantify, predict and control [25]. These tasks require the integration of physical measurement techniques, mathematical modeling and computer science for simulating and predicting the system behavior. In this thesis, we provide new tools for quantification and prediction and we set these tools to work on a general problem of biological systems: the question of how to allocate the available resources to different cellular functions. For the first time, the quantification of bacterial resource allocation at the single cell level was possible. To accomplish this, a substantial amount of work went into overcoming technical problems and standardizing our experimental practices. In this chapter, I will summarize our main contributions, discuss the challenges that emerged throughout this project and present some ideas for continuing this work.

A large number of techniques for quantifying biological properties in real time in a living cell rely on measuring spectroscopic signals, most of the time

the fluorescence of reporter proteins. In dynamical conditions, the raw fluorescence of reporters does not reflect the concentration of these proteins in the cell since they have to mature before becoming fluorescent. In chapter 2, we developed and calibrated mathematical models that can correct for the maturation kinetics of fluorescent reporters. We also developed an inference algorithm to reconstruct underlying signals, here, promoter activities, from fluorescence data. We tested the maturation correction and reconstruction on fluorescence data in batch experiments, using bacterial strains specifically constructed for this purpose. The developed inference method was able to robustly reconstruct the underlying promoter activities. Our experiments and analyses clearly show that maturation can introduce bias in fluorescence data. Therefore, maturation models should be used to correct for this phenomenon.

We constructed maturation models for each fluorescent protein and we specifically made the distinction between GFPs and RFPs. These two types of proteins mature in distinct ways, where RFPs possess a blue intermediate on the path to complete maturation. Consequently, our model for maturation of GFPs involved one step, whereas a two-step model is needed for RFP.

In order to assess whether a two-step model gives more reliable reconstructions for RFPs than a one-step model, we also calibrated a one-step model for our RFP. We found that, based on the same data, the reconstructed promoter activities were not significantly different using the one-step or the two-step model. This comparison was later repeated on single-cell data from the same RFP. We wanted to see whether more subtle dynamics, as the ones observed in microfluidics experiments would be differently reconstructed using the two models. In this case, we did find differences, especially during the transition from glucose to acetate where peaks that were due to maturation were interpreted by the method as an increase in resource allocation with the one-step but not with the two-step model (data not shown).

Our advice would be to use a two-step model for modeling the maturation of RFPs, whenever possible. However, for simplicity, one can use a calibrated one-step model to capture simple dynamics (in batch, for example). For single-cell data and for complex gene expression patterns at the population-level, we advise the use of a two-step model. Lastly, when working with a new reporter

protein, we advise to always verify the maturation mechanism of the protein because some new variants with spectra in other colors also mature in a more complex way (some cyan proteins, for example [93]).

The main scientific question of the thesis was: how do bacteria dynamically allocate resources during growth transitions. For the reconstruction of resource allocation, the underlying dynamical model was fairly simple, with no unknown parameters to estimate. The quantity we reconstruct is $\alpha(t)/\beta$, which represents the fraction of the resources allocated to the production of the proteins of interest (ribosomes, ArgG or Mdh) among all cellular proteins. The only assumption we made is that the total protein concentration, $1/\beta$, is constant and does not change with the growth rate. This assumption is generally accepted for resource allocation models and supported by experimental evidence [158].

In chapter 3, we conducted time-lapse single-cell experiments during transitions from acetate to glucose and back, with strains where the chromosomal genes coding for ribosomes and two metabolic proteins were tagged with fluorescent reporters. The fluorescence and length of individual cells were monitored as a function of time. The obtained measurements for each cell, extracted using a custom-made image analysis pipeline, were used to estimate the growth rates and resource allocation profiles of the proteins of interest, at the single-cell level. The obtained estimates for individual cells allowed us to examine the adaptation strategies of the ribosomes during steady-state growth and during growth transitions.

Unexpectedly, we found that during steady-state growth, even though the growth law is maintained at the population level, this is not the case for the single-cell level, where we observe a large variability in growth rates and ribosomal concentrations (Fig. 3.4). We observe only a weak correlation between growth rate and ribosome concentration. Mechanistically, this implies that a cell can grow at largely different rates with the same number of ribosomes.

We next examined the resource allocation profile in dynamical conditions and found some unexpected, but very reproducible behaviors. During an upshift from acetate to glucose, the observed adaptation strategy of the population strongly deviates from the steady-state growth law, with a more rapid

adaptation of resource allocation followed by a slower adaptation of the growth rate. Furthermore, if we look at the trajectories of individual cells, we observe again a huge variability in the response. Our analysis decomposes the variability into two main classes: the strategy of one part of the population greatly deviates from the "classical" growth law, while the other part follows almost perfectly the steady-state growth law.

The trajectories of resource allocation for the ribosomes were compared with the ones obtained for two metabolic proteins. Surprisingly, the ribosomes and ArgG, two proteins that have a comparable adaptation regime at steady-state (higher concentration at higher growth rate), have very different adaptation regimes dynamically. Whereas resource allocation to the ribosomes increases almost immediately after the upshift, resource allocation and growth rate increase progressively and simultaneously for ArgG (compare Figs 3.4A and 3.6B). Inversely, the ribosomes and Mdh, two proteins that had opposite adaptation patterns at steady-state (the concentration of Mdh decreases when growth rate increases), are characterized by very similar adaptations dynamically. Of course, the signs of the adaptation are the opposite, but the strategy consists in very rapidly changing the resource allocation parameter to the final value, followed be a slow adaptation of the growth rate (Figs 3.4A and 3.6A).

The patterns from the upshift were extremely interesting, at the population-level and for single cells. We also analyzed the dynamical adaptation pattern for a nutrient downshift 3.4CFI. The adaptation pattern is very complex, showing a transient undershoot of resource allocation and growth rate. The pattern is even more surprising for ArgG where the reconstructed resource allocation profiles at this point show negative values of the resource allocation profile, which is not biologically possible (Fig. 3.20D). Such negative values arise from a relatively rapid decrease in fluorescence upon transfer to acetate. In these conditions, the cells show a long lag-phase with very little growth (Fig. 3.1) where the cells do not produce new proteins. Since we continue the observation, it is possible that a significant fraction of the reporter proteins are photobleached. Even though we estimated that photobleaching was negligible in normal growth conditions (Fig. 3.15), we can not exclude that photobleaching becomes important for "non-growing" cells. We were therefore unable to

exploit the downshift data correctly. To do that, a time-varying photobleaching parameter needs to be introduced in the models. In order to estimate this parameter, we would need some additional single-cell calibration experiments. We thus decided to focus the interpretation of our experimental results on the very reproducible and reliable data of the upshift experiments.

To prove the robustness of our conclusions, we repeated the key experiments and swapped reporter genes in order to exclude any effect of the reporter system. We have managed to obtain reproducible results from three independent experiments, two with the strain RA and one with the control strain RA*. For the second replicate experiment, many cells needed to be discarded for various reasons. Therefore, the number of cells left in the analysis was not as high as we would hope for. I plan to add more cells in the analysis (around 70) after this thesis is submitted.

4.2 Perspectives

This work has provided new insights into resource allocation in bacteria but also brought a fresh perspective into the growth laws in general. Here, I would like to discuss questions and ideas that were generated through this project and which we did not have the time to explore. Furthermore, we realized that the generated microfluidics data are extremely rich and can be used to study other phenomena in bacteria. Thus I will also propose some ideas about possible extensions of our current analyses.

One of the most striking observation that was recurrent throughout the interpretation of the data, was the fact that the dynamic behavior of proteins to nutrient changes is very different from the corresponding steady-state behavior. This was true for all three proteins included in this study. This same observation holds for the comparison of single-cell and population data: the dynamical response of single cells often differs significantly from the pattern of the population average. Growth laws have been established on the population level and for steady-state growth. However, in order to expand bacterial physiology, it is time to generalize these growth laws to scenarios that are closer to the natural environment of bacteria, which rarely grow in steady

state conditions. With some further theoretical work, these growth laws can be transformed into "control laws" [25].

Several models have theoretically predicted the resource allocation strategies in dynamical conditions. Our data can be compared with these predictions. For example, our results for individual bacteria, both in steady-state and in dynamic conditions, do not correspond to the predictions of the self-replicator model [119]. This model is based on the assumptions that bacteria maximize biomass accumulation in all conditions. For example, in dynamical conditions, the self-replicator model predicts an on-off strategy (Fig. 1.3). This behavior is not reflected in our data.

In steady state conditions, in the scenario of biomass maximization, one would expect a more pronounced dependency between ribosomal resource allocation and growth rate, in which a higher ribosomal content would yield a higher growth rate, until reaching a ribosomal concentration that is extremely elevated and thus a trade-off for other cellular components necessary for growth is made [33, 119]. This observation was again not visible in our results. Growth rate maximization has been used as an optimization criterion in bacteria [119, 164, 165]. This hypothesis is challenged by our data. We thus have two alternative hypotheses: (i) bacteria allocate their resources in a sub-optimal way [166] or (ii) something else has been optimized through evolution, for example the thermodynamical growth yield [25, 185]. Our work allows to rule out certain models, but more investigation is needed in order to better understand the (quite complex) bases of resource allocation.

Another striking result from our study is the level of heterogeneity in the resource allocation response of individual bacteria. It is quite remarkable that the stochasticity of cellular processes can result in such heterogeneous phenotypical response. Simple stochastic gene expression would predict a much lower variability of, for example, ribosome concentration. In the simplest case, the standard deviation would scale as the square root of the number of molecules. With around 10 000 ribosomes per cell, such a simple model would predict a variability on the order of 1%. Clearly, we observe a much higher heterogeneity. It would be interesting to investigate whether this response is a strategy that populations adopt in order to increase their chances of surviving in a

new or fluctuating environment. In other words: is this large heterogeneity a manifestation of some kind of bet-hedging strategy [186]? In order to decide whether the observed resource allocation strategies within a population of cells corresponds to bet-hedging, we would need to measure the fitness of different populations. Evolutionary experiments that impose different constraints on bacteria, such as a constant or rapidly changing environment, are needed to obtain an answer to this type of question.

All the studies that I carried out during this thesis project measure "general" properties (growth rates, allocation of resources to different classes of genes) of bacteria and bacterial populations, obviating any need to make reference to the underlying molecular mechanisms. Therefore, our experiments alone can not give us any indication about the molecular mechanisms that govern bacterial adaptation. For the ribosomes and Mdh, the change in resource allocation during an upshift is extremely fast. This kind of response could indicate a ppGpp mediation [23, 24]. This global signaling molecule accumulates and disappears rapidly in response to environmental clues, such as the availability of precursors (amino acids) and is well known to control the transcription of rRNA genes, which ultimately determine the rate of production of ribosomes. In order to clarify the role of ppGpp in the adaptation dynamics that we observe, we should repeat this type of experiments in a $ppGpp^{-}$ strain [187] and compare the corresponding resource allocation profiles with the ones from this study. Other global regulators, such as cAMP, may well be part of this global regulatory scheme [34, 35]. Deciphering the molecular mechanisms will require extensive experiments in several different genetic backgrounds.

During my thesis work, I acquired a very large amount of experimental data. Only part of these data have been analyzed, interpreted and presented in this thesis manuscript. We thus envision to further analyze the experimental data. More generally, recycling of experimental data has become crucial considering the amount of measurements one can obtain using new-era quantitative approaches. Many resources (in this case not precursors, but money and researcher time) are allocated to very expensive and sophisticated experiments, the resulting data of which are often under-exploited. Undoubtedly, the data obtained from the microfluidics experiments I conducted are quite

rich and can be exploited to study other phenomena in bacteria.

First, as a direct extension of our work, a stochastic model can be used instead of a deterministic one for the reconstruction of resource allocation. This would provide better estimates for $\alpha(t)/\beta$ since the stochasticity in the signal would be accounted for. With a stochastic model, we can better explore the different regimes of resource allocation at the single-cell level, by observing how the obtained probability histograms evolve over time after the switch.

More generally, the experiments can be used to study general stochastic phenomena in bacteria. The experiments contain long steady states followed by the repeated transitions between growth regimes, while measuring the expression of two different classes of proteins. The steady-state phase gives access to "equilibrium" parameters, whereas the transitions probe the dynamics of the system. The proteins chosen for this study have different copy numbers in the cell (there are for example 100 times more Mdh in the cell than ArgG [157]). The size of simple random fluctuations of the number of protein molecules are very dependent on the absolute number of these proteins per cell. We can therefore use our data about the expression dynamics of ArgG and Mdh to assess the stochasticity at the level of gene expression for different genes and individual bacteria.

Furthermore, we conducted two experiments not presented in this manuscript where some transitions consisted in adding arginine into the medium and observing the behavior of ArgG, which is no longer needed by the cell when the amino acid is present. These experiments could be used to study what happens when the dependency between a protein and the growth rate is "dissolved", for example.

Additionally, the fact that we used a mother machine means that the experiments contain information on the "progeny" of the mother cell, information we, unfortunately, did not have the time to exploit. These data can be used to study the heritability of traits like the growth rate from mother to daughter. These observations can be compared with theoretical studies conducted on the subject [188]. Going even further, the stochastic inheritance can be compared between steady-state and dynamical growth conditions.

References

[1] Swanson, M. S., G. Reguera, M. Schaechter, and F. C. Neidhardt, 2016. Microbe, 2nd edition, ASM Press.

[2] Smith, D. J., 2013. Microbes in the upper atmosphere and unique opportunities for astrobiology research. *Astrobiology* 13:981–990.

[3] Gilichinsky, D., G. Wilson, E. Friedmann, C. Mckay, R. Sletten, E. Rivkina, T. Vishnivetskaya, L. Erokhina, N. Ivanushkina, and G. Kochkina, 2007. Microbial populations in antarctic permafrost: biodiversity, state, age, and implication for astrobiology. *Astrobiology* 7:275–311.

[4] DasSarma, S., 2006. Extreme halophiles are models for astrobiology. *Microbe* 1:120–126.

[5] Westall, F., D. Loizeau, F. Foucher, N. Bost, M. Betrand, J. Vago, and G. Kminek, 2013. Habitability on Mars from a microbial point of view. *Astrobiology* 13:887–897.

[6] West, S. A., S. P. Diggle, A. Buckling, A. Gardner, and A. S. Griffin, 2007. The social lives of microbes. *Annu Rev Ecol Evol Syst* 38:53–77.

[7] Fuhrman, J. A., 2009. Microbial community structure and its functional implications. *Nature* 459:193–199.

[8] Asfahl, K. L., and M. Schuster, 2017. Social interactions in bacterial cell–cell signaling. *FEMS Microbiol Rev* 41:92–107.

[9] Grenham, S., G. Clarke, J. Cryan, and T. Dinan, 2011. Brain–gut–microbe communication in health and disease. *Front Physiol* 2.

[10] Lareen, A., F. Burton, and P. Schäfer, 2016. Plant root-microbe communication in shaping root microbiomes. *Plant Mol Biol* 90:575–587.

[11] Engel, P., W. K. Kwong, Q. McFrederick, K. E. Anderson, S. M. Barribeau, J. A. Chandler, R. S. Cornman, J. Dainat, J. R. de Miranda, V. Doublet, O. Emery, J. D. Evans, L. Farinelli, M. L. Flenniken, F. Granberg, J. A. Grasis, L. Gauthier, J. Hayer, and H. Koch, 2016. The bee microbiome: impact on bee health and model for evolution and ecology of host-microbe interactions. *mBio* 7:e02164–15.

[12] Wargo, M. J., and D. A. Hogan, 2006. Fungal—bacterial interactions: a mixed bag of mingling microbes. *Curr Opin Microbiol* 9:359–364.

[13] Cani, P. D., 2018. Human gut microbiome: hopes, threats and promises. *Gut* 67:1716–1725.

[14] Louis, P., K. Scott, S. Duncan, and H. Flint, 2007. Understanding the effects of diet on bacterial metabolism in the large intestine. *J Gen Microbiol* 102:1197–1208.

[15] Rinninella, E., P. Raoul, M. Cintoni, F. Franceschi, G. A. D. Miggiano, A. Gasbarrini, and M. C. Mele, 2019. What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms* 7:14.

[16] Yano, J. M., K. Yu, G. P. Donaldson, G. G. Shastri, P. Ann, L. Ma, C. R. Nagler, R. F. Ismagilov, S. K. Mazmanian, and E. Y. Hsiao, 2015. Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell* 161:264–276.

[17] Valdes, A. M., J. Walter, E. Segal, and T. D. Spector, 2018. Role of the gut microbiota in nutrition and health. *BMJ* 361:k2179.

[18] Miller, M. B., and B. L. Bassler, 2001. Quorum sensing in bacteria. *Annu Rev Microbiol* 55:165–199.

[19] Prindle, A., J. Liu, M. Asally, S. Ly, J. Garcia-Ojalvo, and G. M. Süel, 2015-11. Ion channels enable electrical communication in bacterial communities. *Nature* 527:59–63.

[20] Ishihama, A., 1997. Adaptation of gene expression in stationary phase bacteria. *Curr Opin Genet Dev* 7:582–588.

[21] Marles-Wright, J., and R. J. Lewis, 2007. Stress responses of bacteria. *Curr Opin Struct Biol* 17:755–760.

[22] Chiang, S. M., and H. E. Schellhorn, 2012. Regulators of oxidative stress response genes in *Escherichia coli* and their functional conservation in bacteria. *Arch Biochem Biophys* 525:161–169.

[23] Traxler, M. F., S. M. Summers, H.-T. Nguyen, V. M. Zacharia, G. A. Hightower, J. T. Smith, and T. Conway, 2008. The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Mol Microbiol* 68:1128–1148.

[24] Zhu, M., Y. Pan, and X. Dai, 2019. (p)ppGpp: the magic governor of bacterial growth economy. *Curr Genet* 65:1121–1125.

[25] Jun, S., F. Si, R. Pugatch, and M. Scott, 2018. Fundamental principles in bacterial physiology - history, recent progress, and the future with focus on cell size control: A review. *Rep Prog Phys* 81:056601.

[26] Schulz, H. N., T. Brinkhoff, T. G. Ferdelman, M. H. Mariné, A. Teske, and B. B. Jørgensen, 1999. Dense populations of a giant sulfur bacterium in namibian shelf sediments. *Science* 284:493–495.

[27] Monod, J., 1949. The Growth of bacterial cultures. *Annu Rev Microbiol* 3:371–394.

[28] Schaechter, M., O. Maaløe, and N. O. Kjeldgaard, 1958. Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. *J Gen Microbiol* 19:592–606.

[29] Kjeldgaard, N. O., O. Maaløe, and M. Schaechter, 1958. The transition between different physiological states during balanced growth of *Salmonella typhimurium*. *J Gen Microbiol* 19:607–616.

[30] Koch, A., 2001. Bacterial growth and form, Springer Science & Business Media.

[31] Neidhardt, F. C., and B. Magasanik, 1960. Studies on the role of ribonucleic acid in the growth of bacteria. *Biochim Biophys Acta* 42:99–116.

[32] 2012. Origins of molecular biology: A tribute to Jacques Monod, Elsevier.

[33] Scott, M., C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa, 2010. Interdependence of cell growth and gene expression: origins and consequences. *Science* 330:1099–1102.

[34] You, C., H. Okano, S. Hui, Z. Zhang, M. Kim, C. W. Gunderson, Y.-P. Wang, P. Lenz, D. Yan, and T. Hwa, 2013. Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* 500:301–306.

[35] Kochanowski, K., H. Okano, V. Patsalo, J. R. Williamson, U. Sauer, and T. Hwa, 2021. Global coordination of metabolic pathways in *Escherichia coli* by active and passive regulation. *Mol Syst Biol* 17:e10064.

[36] Hui, S., J. M. Silverman, S. S. Chen, D. W. Erickson, M. Basan, J. Wang, T. Hwa, and J. R. Williamson, 2015. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol Syst Biol* 11:784.

[37] Yang, L., J. T. Yurkovich, C. J. Lloyd, A. Ebrahim, M. A. Saunders, and B. O. Palsson, 2016. Principles of proteome allocation are revealed using proteomic data and genome-scale models. *Sci Rep* 6:36734.

[38] Mori, M., Z. Zhang, A. Banaei-Esfahani, J. Lalanne, H. Okano, B. C. Collins, A. Schmidt, O. T. Schubert, D. Lee, G. Li, R. Aebersold, T. Hwa, and C. Ludwig, 2021. From coarse to fine: the absolute *Escherichia coli* proteome under diverse growth conditions. *Mol Syst Biol* 17.

[39] Schaechter, M., 2006. From growth physiology to systems biology. *Int Microbiol* 9:157–161.

[40] Campbell, A., 1957. Synchronization of cell division. *Bacteriol Rev* 21:263–272.

[41] Painter, P. R., and A. G. Marr, 1968. Mathematics of microbial populations. *Annu Rev Microbiol* 22:519–548.

[42] Fisher, R. A., B. Gollan, and S. Helaine, 2017. Persistent bacterial infections and persister cells. *Nat Rev Microbiol* 15:453–464.

[43] Hobbie, J., and E. Hobbie, 2013. Microbes in nature are limited by carbon and energy: the starving-survival lifestyle in soil and consequences for estimating microbial rates. *Front Microbiol* 4.

[44] Blount, Z. D., 2015. The unexhausted potential of *E. coli*. *eLife* 4:e05826.

[45] Grossman, G. D., P. B. Moyle, and J. O. Whitaker, 1982. Stochasticity in structural and functional characteristics of an Indiana stream fish assemblage: A test of community theory. *Am Nat* 120:423–454.

[46] Fox, G. A., and B. E. Kendall, 2002. Demographic stochasticity and the variance reduction effect. *Ecology* 83:1928–1934.

[47] Calder, C., M. Lavine, P. Müller, and J. S. Clark, 2003. Incorporating multiple sources of stochasticity into dynamic population models. *Ecology* 84:1395–1402.

[48] Lenormand, T., D. Roze, and F. Rousset, 2009. Stochasticity in evolution. *Trends Ecol Evol* 24:157–165.

[49] Hubbell, S. P., 2001. The unified neutral theory of biodiversity and biogeography, Princeton University Press.

[50] De Vrieze, J., T. De Mulder, S. Matassa, J. Zhou, L. T. Angenent, N. Boon, and W. Verstraete, 2020. Stochasticity in microbiology: managing unpredictability to reach the Sustainable Development Goals. *Microb Biotechnol* 13:829–843.

[51] Kiviet, D. J., P. Nghe, N. Walker, S. Boulineau, V. Sunderlikova, and S. J. Tans, 2014. Stochasticity of metabolism and growth at the single-cell level. *Nature* 514:376–379.

[52] Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, and J. F. Kim, 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461:1243–1247.

[53] Vulić, M., R. E. Lenski, and M. Radman, 1999. Mutation, recombination, and incipient speciation of bacteria in the laboratory. *Proc Natl Acad Sci U S A* 96:7348–7351.

[54] Bamford, R. A., A. Smith, J. Metz, G. Glover, R. W. Titball, and S. Pagliara, 2017. Investigating the physiology of viable but non-culturable bacteria by microfluidics and time-lapse microscopy. *BMC Biol* 15:121.

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

[56] Wakamoto, Y., N. Dhar, R. Chait, K. Schneider, F. Signorino-Gelo, S. Leibler, and J. D. McKinney, 2013. Dynamic persistence of antibiotic-stressed mycobacteria. *Science* 339:91–95.

[57] Amato, S. M., C. H. Fazen, T. C. Henry, W. W. K. Mok, M. A. Orman, E. L. Sandvik, K. G. Volzing, and M. P. Brynildsen, 2014. The role of metabolism in bacterial persistence. *Front Microbiol* 5:70.

[58] Greulich, P., M. Scott, M. R. Evans, and R. J. Allen, 2015. Growth-dependent bacterial susceptibility to ribosome-targeting antibiotics. *Mol Syst Biol* 11:796.

[59] Windels, E. M., Z. Ben Meriem, T. Zahir, K. J. Verstrepen, P. Hersen, B. Van den Bergh, and J. Michiels, 2019. Enrichment of persisters enabled by a ß-lactam-induced filamentation method reveals their stochastic single-cell awakening. *Commun biol* 2:1–7.

[60] Carballo-Pacheco, M., M. D. Nicholson, E. E. Lilja, R. J. Allen, and B. Waclaw, 2020. Phenotypic delay in the evolution of bacterial antibiotic resistance: Mechanistic models and their implications. *PLOS Comput Biol* 16:e1007930.

[61] Visco, P., R. J. Allen, S. N. Majumdar, and M. R. Evans, 2010. Switching and growth for microbial populations in catastrophic responsive environments. *Biophys J* 98:1099–1108.

[62] Taheri-Araghi, S., S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun, 2015. Cell-size control and homeostasis in bacteria. *Curr Biol* 25:385–391.

[63] Wallden, M., D. Fange, E. G. Lundius, O. Baltekin, and J. Elf, 2016. The synchronization of replication and division cycles in individual *E. coli* cells. *Cell* 166:729–739.

[64] Ferrezuelo, F., N. Colomina, A. Palmisano, E. Garí, C. Gallego, A. Csikász-Nagy, and M. Aldea, 2012. The critical size is set at a single-cell level by growth rate to attain homeostasis and adaptation. *Nat Commun* 3:1012.

[65] Pedraza, J. M., and A. van Oudenaarden, 2005. Noise propagation in gene networks. *Science* 307:1965–1969.

[66] Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain, 2002. Stochastic gene expression in a single cell. *Science* 297:1183–1186.

[67] Munsky, B., B. Trinh, and M. Khammash, 2009. Listening to the noise: random fluctuations reveal gene network parameters. *Mol Syst Biol* 5:318.

[68] Thomas, P., G. Terradot, V. Danos, and A. Y. Weiße, 2018. Sources, propagation and consequences of stochasticity in cellular growth. *Nat Commun* 9:4528.

[69] Neidhardt, F. C., P. L. Bloch, and D. F. Smith, 1974. Culture medium for enterobacteria. *J Bacteriol* 119:736–747.

[70] Takahashi, C. N., A. W. Miller, F. Ekness, M. J. Dunham, and E. Klavins, 2015. A low-cost, customizable turbidostat for use in synthetic circuit characterization. *ACS Synth Biol* 4:32–38.

[71] Wong, B. G., C. P. Mancuso, S. Kiriakov, C. J. Bashor, and A. S. Khalil, 2018. Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. *Nat Biotechnol* 36:614–623.

[72] Steel, H., R. Habgood, C. L. Kelly, and A. Papachristodoulou, 2020. In situ characterisation and manipulation of biological systems with Chi.Bio. *PLOS Biol* 18:e3000794.

[73] Aditya, C., F. Bertaux, G. Batt, and J. Ruess, 2021. A light tunable differentiation system for the creation and control of consortia in yeast. *Nature commun* 12:1–10.

[74] Bremer, H., and P. P. Dennis, 2008. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus* 3.

[75] Volkmer, B., and M. Heinemann, 2011. Condition-dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling. *PLOS ONE* 6:e23126.

[76] Day, R. N., and M. W. Davidson, 2009. The fluorescent protein palette: tools for cellular imaging. *Chem Soc Rev* 38:2887–2921.

[77] de Jong, H., C. Ranquet, D. Ropers, C. Pinel, and J. Geiselmann, 2010. Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria. *BMC Syst Biol* 4:55.

[78] Bennett, B. D., E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, and J. D. Rabinowitz, 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol* 5:593–599.

[79] Young, J. W., J. C. W. Locke, A. Altinok, N. Rosenfeld, T. Bacarian, P. S. Swain, E. Mjolsness, and M. B. Elowitz, 2012. Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nat Protoc* 7:80–88.

[80] Moffitt, J. R., J. B. Lee, and P. Cluzel, 2012. The single-cell chemostat: an agarose-based, microfluidic device for high-throughput, single-cell studies of bacteria and bacterial communities. *Lab Chip* 12:1487–1494.

[81] Bakshi, S., A. Siryaporn, M. Goulian, and J. C. Weisshaar, 2012. Super-resolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Mol Microbiol* 85:21–38.

[82] Wang, P., L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, 2010. Robust growth of *Escherichia coli*. *Curr Biol* 20:1099–1103.

[83] Long, Z., E. Nugent, A. Javer, P. Cicuta, B. Sclavi, M. C. Lagomarsino, and K. D. Dorfman, 2013. Microfluidic chemostat for measuring single cell dynamics in bacteria. *Lab Chip* 13:947–954.

[84] Kaiser, M., F. Jug, T. Julou, S. Deshpande, T. Pfohl, O. K. Silander, G. Myers, and E. v. Nimwegen, 2018. Monitoring single-cell gene regulation under dynamically controllable conditions with integrated microfluidics and software. *Nat Commun* 9:1–16.

[85] Wu, F., and C. Dekker, 2016. Nanofabricated structures and microfluidic devices for bacteria: from techniques to biology. *Chem Soc Rev* 45:268–280.

[86] Marinkovic, Z. S., C. Vulin, M. Acman, X. Song, J. M. Di Meglio, A. B. Lindner, and P. Hersen, 2020. Observing Nutrient Gradients, Gene Expression and Growth Variation Using the "Yeast Machine" Microfluidic Device. *Bio Protoc* 10:e3668.

[87] Tsien, R. Y., 1998. The Green Fluorescent Protein. *Annu Rev Biochem* 67:509–544.

[88] Shu, X., N. C. Shaner, C. A. Yarbrough, R. Y. Tsien, and S. J. Remington, 2006. Novel chromophores and buried charges control color in mFruits. *Biochemistry* 45:9639–9647.

[89] Specht, E. A., E. Braselmann, and A. E. Palmer, 2017. A critical and comparative review of fluorescent tools for live-cell imaging. *Annu Rev Physiol* 79:93–117.

[90] Zaslaver, A., A. Bren, M. Ronen, S. Itzkovitz, I. Kikoin, S. Shavit, W. Liebermeister, M. G. Surette, and U. Alon, 2006. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat Methods* 3:623–628.

[91] English, B. P., V. Hauryliuk, A. Sanamrad, S. Tankov, N. H. Dekker, and J. Elf, 2011. Single-molecule investigations of the stringent response machinery in living bacterial cells. *Proc Natl Acad Sci U S A* 108:E365–E373.

[92] Forsberg, A. J., G. D. Pavitt, and C. F. Higgins, 1994. Use of transcriptional fusions to monitor gene expression: a cautionary tale. *J Bacteriol* 176:2128–2132.

[93] Balleza, E., J. M. Kim, and P. Cluzel, 2018. Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat Methods* 15:47–51.

[94] Remington, S. J., 2006. Fluorescent proteins: maturation, photochemistry and photophysics. *Curr Opin Struct Biol* 16:714–721.

[95] Heux, S., B. Philippe, and J.-C. Portais, 2011. High-Throughput Workflow for monitoring and mining bioprocess data and its application to inferring the physiological response of *Escherichia coli* to perturbations. *Appl Environ Microbiol* 77:7040–7049.

[96] Lazebnik, Y., 2002. Can a biologist fix a radio?—Or, what I learned while studying apoptosis. *Cancer Cell* 2:179–182.

[97] May, R. M., 2004. Uses and abuses of mathematics in biology. *Science* 303:790–793.

[98] Noid, W. G., 2013. Perspective: Coarse-grained models for biomolecular systems. *J Chem Phys* 139:090901.

[99] de Jong, H., S. Casagranda, N. Giordano, E. Cinquemani, D. Ropers, J. Geiselmann, and J.-L. Gouzé, 2017. Mathematical modelling of microbes: metabolism, gene expression and growth. *J R Soc Interface* 14:2017050.

[100] Klipp, E., W. Liebermeister, C. Wierling, and A. Kowald, 2016. Systems biology: A textbook, 2nd edition, Wiley.

[101] Weiße, A. Y., D. A. Oyarzún, V. Danos, and P. S. Swain, 2015. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc Natl Acad Sci U S A* 112:E1038–E1047.

[102] Mauri, M., J.-L. Gouzé, H. d. Jong, and E. Cinquemani, 2020. Enhanced production of heterologous proteins by a synthetic microbial community: Conditions and trade-offs. *PLOS Comput Biol* 16:e1007795.

[103] Covert, M. W., N. Xiao, T. J. Chen, and J. R. Karr, 2008. Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. *Bioinformatics* 24:2044–2050.

[104] Lerman, J. A., D. R. Hyduke, H. Latif, V. A. Portnoy, N. E. Lewis, J. D. Orth, A. C. Schrimpe-Rutledge, R. D. Smith, J. N. Adkins, K. Zengler, and B. O. Palsson, 2012. In silico method for modelling metabolism and gene product expression at genome scale. *Nat Commun* 3:929.

[105] Karr, J. R., J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival, N. Assad-Garcia, J. I. Glass, and M. W. Covert, 2012. A whole-cell computational model predicts phenotype from genotype. *Cell* 150:389–401.

[106] Carrera, J., and M. Covert, 2015. Why build whole-cell models? *Trends cell biol* 25:719–722.

[107] Macklin, D. N., T. A. Ahn-Horst, H. Choi, N. A. Ruggero, J. Carrera, J. C. Mason, G. Sun, E. Agmon, M. M. DeFelice, I. Maayan, K. Lane, R. K. Spangler, T. E. Gillies, M. L. Paull, S. Akhter, S. R. Bray, D. S. Weaver, I. M. Keseler, P. D. Karp, J. H. Morrison, and M. W. Covert, 2020. Simultaneous cross-evaluation of heterogeneous *E. coli* datasets via mechanistic simulation. *Science* 369:eaav3751.

[108] Orth, J. D., I. Thiele, and B. Ø. Palsson, 2010. What is flux balance analysis? *Nat Biotechnol* 28:245–248.

[109] Price, N. D., J. L. Reed, and B. Ø. Palsson, 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol* 2:886–897.

[110] Senne de Oliveira Lino, F., D. Bajic, J. C. C. Vila, A. Sánchez, and M. O. A. Sommer, 2021. Complex yeast–bacteria interactions affect the yield of industrial ethanol fermentation. *Nat Commun* 12:1498.

[111] Bulović, A., S. Fischer, M. Dinh, F. Golib, W. Liebermeister, C. Poirier, L. Tournier, E. Klipp, V. Fromion, and A. Goelzer, 2019. Automated generation of bacterial resource allocation models. *Metab Eng* 55:12–22.

[112] Allen, R. J., and B. Waclaw, 2019. Bacterial growth: a statistical physicist's guide. *Rep Prog Phys* 82:016601.

[113] Wilkinson, D. J., 2011. Stochastic modelling for systems biology, CRC Press.

[114] Dutta, D., and S. Saini, 2021. Cell growth model with stochastic gene expression helps ynderstand the growth advantage of metabolic exchange and auxotrophy. *mSystems* 6:e0044821.

[115] Molenaar, D., R. van Berlo, D. de Ridder, and B. Teusink, 2009. Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Syst Biol* 5:323.

[116] Zaslaver, A., S. Kaplan, A. Bren, A. Jinich, A. Mayo, E. Dekel, U. Alon, and S. Itzkovitz, 2009. Invariant distribution of promoter activities in *Escherichia coli*. *PLoS Comput Biol* 5:e1000545.

[117] Bosdriesz, E., D. Molenaar, B. Teusink, and F. J. Bruggeman, 2015. How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. *FEBS J* 282:2029–2044.

[118] Erickson, D. W., S. J. Schink, V. Patsalo, J. R. Williamson, U. Gerland, and T. Hwa, 2017. A global resource allocation strategy governs growth transition kinetics of Escherichia coli. *Nature* 551:119–123.

[119] Giordano, N., F. Mairet, J.-L. Gouzé, J. Geiselmann, and H. de Jong, 2016. Dynamical allocation of cellular resources as an optimal control problem: Novel insights into microbial growth strategies. *PLoS Comput Biol* 12:e1004802.

[120] Engl, H. W., C. Flamm, P. Kügler, J. Lu, S. Müller, and P. Schuster, 2009. Inverse problems in systems biology. *Inverse Probl* 25:123014.

[121] Guzzi, R., T. Colombo, and P. Paci, 2018. Inverse problems in systems biology: A critical review, Springer.

[122] Cinquemani, E., and L. Paulevé, 2021. Computational methods in systems biology: 19th International Conference, CMSB 2021, Proceedings, Springer Nature.

[123] Bertero, M., 1989. Linear inverse and ill-Posed problems. *Adv. Electron. El. Phys.* 75:1 – 120.

[124] De Nicolao, G., G. Sparacino, and C. Cobelli, 1997. Nonparametric input estimation in physiological systems: Problems, methods, and case studies. *Automatica* 33:851–870.

[125] Wahba, G., 1990. Spline models for observational data, SIAM.

[126] Hastie, T., R. Tibshirani, and J. H. Friedman, 2001. The elements of statistical learning: Data mining, inference, and prediction, Springer Science & Business Media.

[127] Bansal, M., V. Belcastro, A. Ambesi-Impiombato, and D. d. Bernardo, 2007. How to infer gene networks from expression profiles. *Mol Syst Biol* 3:122.

[128] Porreca, R., E. Cinquemani, J. Lygeros, and G. Ferrari-Trecate, 2010. Structural identification of unate-like genetic network models from time-lapse protein concentration measurements, Proc IEEE Conf Decis Control.

[129] Lichten, C. A., R. White, I. B. Clark, and P. S. Swain, 2014. Unmixing of fluorescence spectra to resolve quantitative time-series measurements of gene expression in plate readers. *BMC Biotechnol* 14:11.

[130] Zulkower, V., M. Page, D. Ropers, J. Geiselmann, and H. de Jong, 2015. Robust reconstruction of gene expression profiles from reporter gene data using linear inversion. *Bioinform* 31:i71–i79.

[131] Kannan, S., T. Sams, J. Maury, and C. T. Workman, 2018. Reconstructing dynamic promoter activity profiles from reporter gene data. *ACS Synth Biol* 7:832–841.

[132] De Nicolao, G., and G. Ferrari-Trecate, 2003. Regularization networks for inverse problems: A state-space approach. *Automatica* 39:669–676.

[133] Kailauth, T., 2005. Linear Estimation. *IEEE Transactions on Information Theory* 51:2236–2240.

[134] Jazwinski, A. H., 1970. Stochastic processes and filtering theory, Elsevier.

[135] Tikhonov, A. Nonlinear ill-posed problems Springer Dordrecht.

[136] Leveau, J. H. J., and S. E. Lindow, 2001. Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. *J Bacteriol* 183:6752–6762.

[137] Wang, X., B. Errede, and T. C. Elston, 2008. Mathematical analysis and quantification of fluorescent proteins as transcriptional reporters. *Biophys J* 94:2017–2026.

[138] Finkenstädt, B., E. A. Heron, M. Komorowski, K. Edwards, S. Tang, C. V. Harper, J. R. E. Davis, M. R. H. White, A. J. Millar, and D. A. Rand, 2008. Reconstruction of transcriptional dynamics from gene reporter data using differential equations. *Bioinform* 24:2901–2907.

[139] Verkhusha, V. V., D. M. Chudakov, N. G. Gurskaya, S. Lukyanov, and K. A. Lukyanov, 2004. Common pathway for the red chromophore formation in fluorescent proteins and chromoproteins. *Chem Biol* 11:845–854.

[140] Strack, R. L., D. E. Strongin, L. Mets, B. S. Glick, and R. J. Keenan, 2010. Chromophore formation in DsRed occurs by a branched pathway. *J Am Chem Soc* 132:8496–8505.

[141] Rasmussen, C. E., and C. K. I. Williams, 2006. Gaussian processes for machine learning, MIT Press.

[142] Kailath, T., A. H. Sayed, and B. Hassibi, 2000. Linear estimation, Prentice Hall.

[143] Davis, J. H., A. J. Rubin, and R. T. Sauer, 2011. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res* 39:1131–1141.

[144] Miyawaki, A., D. M. Shcherbakova, and V. V. Verkhusha, 2012. Red fluorescent proteins: chromophore formation and cellular applications. *Curr Opin Struct Biol* 22:679–688.

[145] Gibson, D. G., L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith, 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6:343–345.

[146] Berthoumieux, S., H. de Jong, G. Baptist, C. Pinel, C. Ranquet, D. Ropers, and J. Geiselmann, 2013. Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol Syst Biol* 9:634.

[147] Martin, Y., M. Page, C. Blanchet, and H. de Jong, 2019. WellInverter: a web application for the analysis of fluorescent reporter gene data. *BMC Bioinform* 20:309.

[148] Cinquemani, E., V. Laroute, M. Cocaign-Bousquet, H. de Jong, and D. Ropers, 2017. Estimation of time-varying growth, uptake and excretion rates from dynamic metabolomics data. *Bioinform* 33:i301–i310.

[149] Sakamoto, Y., M. Ishiguro, and G. Kitagawa, 1986. Akaike information criterion statistics, Springer Netherlands.

[150] Stefan, D., C. Pinel, S. Pinhal, E. Cinquemani, J. Geiselmann, and H. de Jong, 2015. Inference of quantitative models of bacterial promoters from time-series reporter gene data 11:e1004028.

[151] Mairet, F., J.-L. Gouzé, and H. de Jong, 2021. Optimal proteome allocation and the temperature dependence of microbial growth laws. *npj Syst Biol Appl* 7:1–11.

[152] Scott, M., S. Klumpp, E. M. Mateescu, and T. Hwa, 2014. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Mol Syst Biol* 10:747.

[153] Bremer, H., and P. P. Dennis, 1975. Transition period following a nutritional shift-up in the bacterium *Escherichia coli* B/r: stable RNA and protein synthesis. *J Theor Biol* 52:365–382.

[154] Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori, 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2:2006.0008.

[155] Ollion, J., M. Elez, and L. Robert, 2019. High-throughput detection and tracking of cells and intracellular spots in mother machine experiments. *Nat Protoc* 14:3144–3161.

[156] Basan, M., T. Honda, D. Christodoulou, M. Hörl, Y.-F. Chang, E. Leoncini, A. Mukherjee, H. Okano, B. R. Taylor, J. M. Silverman, C. Sanchez, J. R. Williamson, J. Paulsson, T. Hwa, and U. Sauer, 2020. A universal tradeoff between growth and lag in fluctuating environments. *Nature* 584:470–474.

[157] Schmidt, A., K. Kochanowski, S. Vedelaar, E. Ahrné, B. Volkmer, L. Callipo, K. Knoops, M. Bauer, R. Aebersold, and M. Heinemann, 2016. The quantitative and condition-dependent *Escherichia coli* proteome. *Nat Biotechnol* 34:104–110.

[158] Basan, M., M. Zhu, X. Dai, M. Warren, D. Sévin, Y.-P. Wang, and T. Hwa, 2015. Inflating bacterial cells by increased protein synthesis. *Mol Syst Biol* 11:836.

[159] Pavlou, A., E. Cinquemani, J. Geiselmann, and H. de Jong, 2021. Maturation models are necessary to obtain unbiased estimates of promoter activity, submitted.

[160] Gerosa, L., K. Kochanowski, M. Heinemann, and U. Sauer, 2013. Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol Syst Biol* 9:658.

[161] Mori, M., S. Schink, D. W. Erickson, U. Gerland, and T. Hwa, 2017. Quantifying the benefit of a proteome reserve in fluctuating environments. *Nat Commun* 8:1225.

[162] Caldara, M., D. Charlier, and R. Cunin, 2006. The arginine regulon of *Escherichia coli*: whole-system transcriptome analysis discovers new genes and provides an integrated view of arginine regulation. *Microbiology* 152:3343–3354.

[163] Sutherland, P., and L. McAlister-Henn, 1985. Isolation and expression of the *Escherichia coli* gene encoding malate dehydrogenase. *J Bacteriol* 163:1074–1079.

[164] van den Berg, H. A., Y. N. Kiselev, S. A. L. M. Kooijman, and M. V. Orlov, 1998. Optimal allocation between nutrient uptake and growth in a microbial trichome. *J Math Biol* 37:28–48.

[165] Hu, X.-P., H. Dourado, P. Schubert, and M. J. Lercher, 2020. The protein translation machinery is expressed for maximal efficiency in *Escherichia coli*. *Nat Commun* 11:5260.

[166] Balakrishnan, R., R. T. de Silva, T. Hwa, and J. Cremer, 2021. Suboptimal resource allocation in changing environments constrains response and growth in bacteria. *Mol Syst Biol* 17:e10597.

[167] Koch, A., 1988. Why can't a cell grow infinitely fast? *Can J Microbiol* 34:421–426.

[168] Cashel, M., and J. Gallant, 1969. Two Compounds implicated in the Function of the RC Gene of *Escherichia coli*. *Nature* 221:838–841.

[169] Wada, A., K. Igarashi, S. Yoshimura, S. Aimoto, and A. Ishihama, 1995. Ribosome Modulation Factor: stationary growth phase-specific inhibitor of ribosome functions from *Escherichia coli*. *Biochem Biophys Res Commun* 214:410–417.

[170] Lemke, J. J., P. Sanchez-Vazquez, H. L. Burgos, G. Hedberg, W. Ross, and R. L. Gourse, 2011. Direct regulation of *Escherichia coli* ribosomal protein promoters by the transcription factors ppGpp and DksA. *Proc Natl Acad Sci USA* 108:5712–5717.

[171] Braun, V., 2009. FhuA (TonA), the career of a protein. *J Bacteriol* 191:3431–3436.

[172] Sharan, S. K., L. C. Thomason, S. G. Kuznetsov, and D. L. Court, 2009. Recombineering: a homologous recombination-based method of genetic engineering. *Nat Protoc* 4:206–223.

[173] Guzman, L. M., D. Belin, M. J. Carson, and J. Beckwith, 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* 177:4121–4130.

[174] Miller, J. H., 1972. Experiments in molecular genetics. *Cold Spring Harbor Laboratory* Cold Spring Harbor, New York.

[175] Edelstein, A., N. Amodaj, K. Hoover, R. Vale, and N. Stuurman, 2010. Computer control of microscopes using μManager. *Curr Protoc Mol Biol* 92:14.20.1–14.20.17.

[176] Likas, A., N. Vlassis, and J. J. Verbeek, 2003. The global k-means clustering algorithm. *Pattern Recognit* 36:451–461.

[177] Buitinck, L., G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, 2013. API design for machine learning software: experiences from the scikit-learn project.

In ECML PKDD Workshop: Languages for Data Mining and Machine Learning. 108–122.

[178] Tibshirani, R., G. Walther, and T. Hastie. Estimating the number of clusters in a data set via the gap statistic. *J R Stat Soc, B: Stat Methodol* 63:411–423.

[179] James, G., D. Witten, T. Hastie, and R. Tibshirani, 2013. An introduction to statistical learning: with applications in R, Springer Science & Business Media.

[180] Pham, D. T., S. S. Dimov, and C. D. Nguyen, 2005. Selection of K in K-means clustering. *Proc Inst Mech Eng C J Mech Eng Sci P I MECH ENG C-J MEC* 219:103–119.

[181] Milo, R., 2013. What is the total number of protein molecules per cell volume? A call to rethink some published values. *Bioessays* 35:1050–5.

[182] Basan, M., M. Zhu, X. Dai, M. Warren, D. Sévin, Y.-P. Wang, and T. Hwa, 2015. Inflating bacterial cells by increased protein synthesis. *Mol Syst Biol* 11:836.

[183] Larrabee, K., J. Phillips, G. Williams, and A. Larrabee, 1980. The relative rates of protein synthesis and degradation in a growing culture of *Escherichia coli*. *J Biol Chem* 255:4125–4130.

[184] Mosteller, R., R. Goldstein, and K. Nishimoto, 1980. Metabolism of individual proteins in exponentially growing Escherichia coli. *J Biol Chem* 255:2524–2532.

[185] Jackson, B. E., and M. J. McInerney, 2002. Anaerobic microbial metabolism can proceed close to thermodynamic limits. *Nature* 415:454–456.

[186] de Jong, I. G., P. Haccou, and O. P. Kuipers, 2011. Bet hedging or not? A guide to proper classification of microbial survival strategies. *BioEssays* 33:215–223.

[187] ppGpp is a bacterial cell size regulator 32:870–877.

[188] Marguet, A., M. Lavielle, and E. Cinquemani, 2019. Inheritance and variability of kinetic gene expression parameters in microbial cells: modeling and inference from lineage tree data. *Bioinformatics* 35:i586–i595.

List of Figures

List of Tables