TU Delft

Delft University of Technology

A single-cell study on stochasticity growth and gene expression

Walker, Noreen

DOI
10.4233/uuid:7f3b3281-9136-4876-b2a5-a151d69b659b

Publication date
2016

Document Version
Final published version

Citation (APA)
Walker, N. (2016). *A single-cell study on stochasticity growth and gene expression*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:7f3b3281-9136-4876-b2a5-a151d69b659b

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

A SINGLE-CELL STUDY ON STOCHASTICITY IN GROWTH AND GENE EXPRESSION

Noreen E. Walker

A SINGLE-CELL STUDY ON STOCHASTICITY IN GROWTH AND GENE EXPRESSION

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op vrijdag 22 april 2016 om 15.00 uur

door

Noreen Evelyn WALKER

Diplom-Physikerin,
Ludwig-Maximilians-Universität München, Duitsland,
geboren te München, Duitsland.

Dit proefschrift is goedgekeurd door de

promotor: prof. dr. ir. S. J. Tans

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. S. J. Tans	Technische Universiteit Delft
Onafhankelijke leden:	
Prof. dr. M. Dogterom	Technische Universiteit Delft
Prof. dr. P. R. ten Wolde	Vrije Universiteit Amsterdam
Prof. dr. M. Cosentino Lagomarsino	Université Pierre et Marie Curie, Paris, France
Dr. A. S. Meyer	Technische Universiteit Delft
Dr. R. Hermsen	Universiteit Utrecht
Prof. dr. A. H. Engel	Technische Universiteit Delft, reservelid

The work described in this thesis was performed at the FOM Institute AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands. This work is part of the research program of the Foundation for Fundamental Research on Matter (FOM), which is financially supported by the Netherlands Organisation for Scientific Research (NWO).

Printed by: GVO, Ede, The Netherlands

Cover: Overlay of fluorescence and phase contrast microscopy images, which were the imaging methods used throughout this thesis. The sample is a fluorescent dye, interrupted by spots of non-fluorescent, dry surface.

Copyright © 2016 by N. Walker

ISBN 978-94-92323-02-6

A digital version of this thesis can be obtained from `http://www.amolf.nl` and from `http://repository.tudelft.nl`. Printed copies can be obtained by request via email to library@amolf.nl.

CONTENTS

1

INTRODUCTION

1.1. A STOCHASTIC MICROSCOPIC WORLD

The physical rules dominating the world of bacteria are quite different than those of the human world. Bacteria are around 1,000,000 times smaller in length than humans - only a few µm in size - and weigh an extreme factor of approximately 10^{13} less [1]. On that scale, physical properties like mass and inertia are irrelevant and, for example, bacterial movement is instead governed by viscous drag [2]. But maybe the biggest difference between the macroscopic and the microscopic world is the predictability of processes and reactions: For a finite temperature, the elementary particles of fluids and gases are in constant random motion. In macroscopic objects such stochastic movements average out because of the sheer number of involved molecules, and macroscopic processes are thus deterministic. Within cells, however, the copy number of particles such as proteins or molecular machines can be very small and hence stochastic fluctuations caused by thermal motion do not simply average out. For many types of molecules bacterial cells even rely on that random motion to distribute particles, instead of using active transport mechanisms [3, 4]. As consequence of thermal fluctuations and low copy numbers, virtually all processes in a cell are to some extent stochastic, or, in other terms, noisy.

Cells can for example not perfectly control the rate at which proteins are produced, and the concentration of proteins fluctuates over time (Fig. 1.1A) [5, 6]. As consequence of cellular noise, genetically identical cells can behave differently - even in the same environment. This variability can range from variation in protein concentration [6–8] (Fig. 1.1A) to different growth states and sensitivity to antibiotic treatment [9–11], as well as to differentiation into specialized cell types (Fig. 1.1B) [12]. Even in higher organisms numerous cell fate decisions are thought to be based on stochastic processes [13, 14].

Many questions arise when considering molecular noise on the one hand and the robust growth of bacteria on the other hand. For example: How large is the cell-to-cell variability, for instance in protein content? How do cells cope with this randomness? Are the consequences of noise detrimental or also advantageous? In recent years, quantitative tools for measuring cellular variability became increasingly available and the-

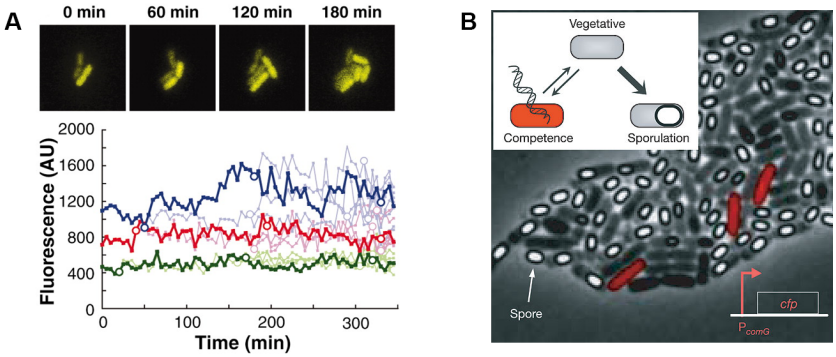


Figure 1.1: Examples of cell-to-cell variability. (**A**) Genetically identical *E. coli* bacteria produce different amounts of proteins (upper panel) and the protein concentration in single cells fluctuates over time (lower graph). (**B**) Clonal *B. subtilis* bacteria can differentiate into different cell types. The differentiation is driven by stochastic protein production. Figure (**A**) was taken from [6], (**B**) was taken from [12].

oretical work created models to better understand origins of noise and robustness to fluctuations [5, 7, 15–20]. Much progress in understanding cellular noise has therefore been made and some of the results are: The cell-to-cell variability of protein concentration depends inversely on the average concentration, but does not decrease below a certain noise floor [6, 18, 19, 21]. Cells can actively suppress and enhance protein fluctuations with different feedback networks [22–25]. Fluctuations in proteins can propagate to other proteins downstream in a regulatory network [26–28]. While noise is considered to be generally undesired, cellular heterogeneity can be beneficial and increase survival in unpredictable environments ("bet hedging") [15, 29, 30].

1.1.1. MORE TO BE DISCOVERED

Despite these advances, many aspects of cellular variability still remain poorly understood. The research done in this thesis aims to contribute to a better understanding of some of these aspects. One main point that we study is cellular growth rate: Single-cell growth has not been under much investigation (for recent exceptions, see [31–33]), although it is an important cellular property and directly relates to the fitness of cells. A central goal of bacterial cells is to grow as fast as possible, but for growth cells must rely on their protein-machinery - which is noisy. We investigate whether single-cell growth rates fluctuate as well, how growth and protein fluctuations are related and whether fluctuations propagate. A second focus point of this thesis are specific sources of cellular variation because little is still known on which sources are the most relevant ones. Therefore, we investigate two key parameters of bacteria: cell cycle stage and ribosome content (ribosomes are the protein producing machines). We test whether and to what extent they cause noise in protein production and growth and, for the latter, whether they are dynamically limiting cellular growth.

To study these questions, we follow a systems biology approach. Contrary to reduc-

tionist methods, where a cell is split into increasingly smaller parts, we investigate the intact system as a whole. For example, we aim to understand fluctuations within the context of the whole cell and its interaction with the environment. The next section introduces available and chosen methods to study questions on cellular variability: A model organism, experimental techniques and theoretical approaches.

1.2. TOWARDS UNDERSTANDING NOISE IN SINGLE CELLS

1.2.1. ESCHERICHIA COLI: A "SIMPLE" MODEL ORGANISM

To investigate variability in cells, we work with a well-studied model organism, the bacterium *Escherichia coli* (*E. coli*) (Fig. 1.2). Many years of research on *E. coli* provide a good knowledge base to study fundamental biological processes. *E. coli* can be readily cultured in the lab and, importantly, can be genetically manipulated. It is a rod-shaped, gram-negative bacterium and like all prokaryotes unicellular [34]. Its length is around 3 µm, the diameter roughly 1 µm and, depending on the environment, cells can grow very fast with doubling times as short as 20 min [1, 35]. The natural habitat of *E. coli* is the gut of humans and animals and most *E. coli* strains live in symbiosis with their host, while a few are pathogenic. Cells can also survive in many different non-native environments and can grow on a wide variety of nutrients. Like many bacteria, *E. coli* cells reproduce (normally) by binary fission and therefore daughter cells are typically genetically identical [34].

The inside of *E. coli* cells (cytoplasm) is enclosed by a cell wall and an inner and outer membrane (inside, respectively outside of the cell wall) (Fig. 1.2). Characteristic for bacteria, the cytoplasm is not sub-structured, and cells do, for example, not have a nucleus. The DNA encodes around 4,000 genes [36] and has a contour length of 1.5 mm [37]. Therefore, it is tightly packed in the cell. The cytoplasm of *E. coli* is very crowded and molecules continuously bump into each other [38]. Of the cellular dry mass, proteins constitute the largest fraction (50%), followed by RNA and DNA (20% together) [1].

In this thesis we used the non-pathogenic strain MG1655. It has been grown in the lab for decades and is based on the natural isolate K-12. The complete genome sequence of MG1655 was published in 1997 [36], which makes *E. coli* one of the first organisms to be sequenced.

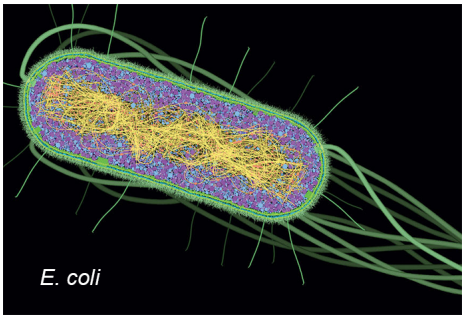


Figure 1.2: Illustration of the bacterium *E. coli*. Cell wall and membranes (as well as the filament-like flagella which are required for swimming) are depicted in green. The cytoplasmic area, containing for example proteins and ribosomes, is colored in blue-purple. The DNA is shown in yellow and nucleoid-associated proteins are depicted in orange. The illustration was created by David Goodsell (http://www.scripps.edu/goodsell/).

1.2.2. SINGLE-CELL EXPERIMENTS

Investigating cell-to-cell variability requires first of all an experimental technique that provides single-cell resolution. Second, throughput of cell measurements needs to be sufficiently high so that the distribution of cellular states, not just its average, can be obtained. The latter typically requires the use of automated (computer-controlled) setups. Multiple techniques fulfill these criteria and new approaches are being developed [39–42].

Microscopy is one of the most widespread single-cell methods. It can, for example, be used to observe variability in mRNA or protein concentration, in growth rate and phenotype [7, 9, 12]. To visualize expression of native proteins, they can be fused to (or replaced by) fluorescent proteins [43]. Importantly, fluorescent proteins can be measured in live cells and thus allow monitoring of expression over time. With the help of microfluidics, the environment can be precisely controlled and changed [31, 44]. Detection of mRNA levels is more difficult and is usually done in fixed (killed) cells. A common method is fluorescence *in situ* hybridization (FISH), in which fluorescent tags are designed to bind to a specific mRNA sequence [40]. Then, single mRNAs become visible as bright fluorescent spots.

Fluorescence-activated cell sorting (FACS) is a different and high-throughput method to measure cellular heterogeneity [45]. FACS can be used to measure live cells, but individual cells cannot be tracked over time. Similar to microscopy, it first requires fluorescent labeling of the quantity of interest, for example, of proteins. For sorting, the cell suspension is transformed into a very thin and fast stream of liquid droplets, with typically one cell per droplet. The droplets can then be sorted based on scattering and fluorescence properties. With this method data acquisition of thousands of cells is readily achieved, but interpretation of FACS data is sometimes not trivial [46]. Another single-cell method is real-time qPCR (polymerase chain reaction), which can measure mRNA copy number in individual cells at rather high throughput, but requires lysis of cells [47]. The advantage is that this method does not require labeling, and the mRNA of interest is directly selected by adjusting the primer sequence. New single-cell methods are being developed, and especially advances in the "omics"-field, that is the complete profiling of the cellular proteome, transcriptome, etc., as well as new methods for single-cell (genome) sequencing, will likely give new insights [31, 41, 42, 48, 49].

In this thesis we study gene expression and growth rate fluctuations and their propagation. To address these questions, time-resolved information of growing *E. coli* cells is required. Therefore, the only suitable method is time-lapse microscopy, in which a movie of multiple generations of growing cells can be acquired. Cells can therefore be tracked and growth rate and, by using fluorescent proteins, gene expression of single cells can be determined.

1.2.3. MODELING

Modeling in biology is of ever growing importance, especially since in recent years quantitative experimental data with larger statistical power became available [50]. In the context of cellular variability, models can, for example, give valuable insight into the distribution of protein concentrations [17, 18], the sensitivity of networks to noise [51] or transmission of fluctuations to downstream components [26, 27]. Natural challenges are

the complexity of the studied system and modeling of emergent properties, which are "more" than the physics of its components [50]. While the focus of this thesis is clearly experimental, we develop and make use of models on several occasions (chapter 4-6).

A detailed overview on methods is beyond the scope of this introduction, therefore we here focus on one aspect relevant for models in systems biology. To explain an observation, a model can be either rather mechanistic or rather phenomenological (or lie in the continuum in between) [52]. "Mechanistic" here means that the equations directly describe the involved biochemical processes and that parameters have an immediate biological meaning. The purpose of phenomenological models is to reproduce the experimental data (in a simple way) and to make predictions, without requiring information on all underlying processes [52]. Such an approach is usually chosen when mechanistic details are unknown, would be too complex too include or would hide the relevant (global) parameters. Naturally, interpretation of model parameters can then be difficult. Given the cellular complexity, phenomenological models are regularly chosen in systems biology and work surprisingly well [32, 53, 54]. Apparently, biological quantities frequently obey simple relations on the systems level [55]. That aside, systems biology might benefit from a larger effort to unify these models, by integrating published results and theories more strongly into new "black box" models.

We develop and use phenomenological models in this thesis. Arguably, they are to date the better choice for understanding cellular growth, since on the molecular level thousands of biomolecular processes might need to be considered. Modeling allows us to connect experimental observations, to make predictions and to, for example, determine which elements (such as noise sources, regulatory links) are relevant in which environment.

1.3. THESIS OUTLINE

In this thesis, we address questions around cellular variability which can be categorized into two main topics: In chapter 4, 6 and 7 we study the connection between fluctuations in gene expression and fluctuations in cellular growth rate. In chapter 5, we address specific sources of noise and their relevance for the cell. These topics are not strictly separated and some projects can be found in between these categories (see chapter 6 and 7). In more detail, the thesis is structured as follows:

In Chapter 2 we explain the general experimental and analysis methods used in this thesis. We describe the construction of fluorescent reporters into *E. coli* cells and the experimental protocol used for time-lapse microscopy. Additionally, the methods and software used for image analysis and cell tracking are introduced. We also describe the calculation of temporal cross-correlations, which we will use in later chapters to analyze fluctuations.

In Chapter 3 we characterize different fluorescent proteins and assess their quality as reporter of dynamic gene expression fluctuations. To this end, we compare cross-correlations, maturation times (time needed for a produced protein to become fluorescent) and photostability of six different fluorescent proteins. We find that for our experimental design a green fluorescent protein variant, as well as a combination of yellow and cyan emitting proteins (if two reporters are needed), are good choices.

In Chapter 4 we investigate the general relation between noise in gene expression and growth rate. To this end, expression noise is decomposed into global cell-wide fluctuations (extrinsic noise) and local promoter-specific fluctuations (intrinsic noise). We find that growth rate fluctuations are modestly coupled to extrinsic expression fluctuations, but essentially uncoupled from intrinsic expression fluctuations. In contrast, the magnitudes of both expression and growth noise are very strongly correlated and scale linearly. A model is used to explain the different scaling behaviors observed for different noise types and expression levels.

In Chapter 5 we turn towards a specific source of fluctuations in gene expression: the cell cycle. We show that about half of the noise in protein production rate is caused by the gene duplication occurring during cell cycle progression. In contrast to that, protein concentration is hardly affected by the cell cycle. We further show that cell cycle related fluctuations can be decomposed into deterministic oscillations and a stochastic contribution. The influence of the location of a gene on the chromosome is predicted by a simple model.

In Chapter 6 we investigate whether random fluctuations in protein concentration can affect cellular growth. We show that fluctuations in the expression of needed genes can propagate and cause growth fluctuations. The amount of noise transmission depends on how strongly the enzyme is limiting growth. We further show that growth fluctuations also propagate back to disturb protein concentrations. An analytical model is developed to accurately predict noise transmission. Our results indicate that noise can be propagated by metabolic reactions and that cellular metabolism is inherently stochastic.

In Chapter 7 we study the influence of fluctuations in the number of ribosomes on the cellular growth rate. Since ribosomes produce all proteins, they play a central role in cellular growth, and their concentration is tightly adjusted to the supported growth rate. We test whether fluctuations of ribosomes are dynamically limiting cellular growth and whether ribosomes fluctuate around an optimum concentration. Despite the preliminary nature of this chapter, the data suggests that random fluctuations in ribosome content are not propagating to and thus not dynamically limiting cellular growth.

2

EXPERIMENTAL AND ANALYSIS METHODS

In this chapter we introduce the general experimental and analysis methods used in this thesis. We describe the preparation and carrying out of time-lapse microscopy experiments with E. coli cells. Image analysis methods are introduced, and we describe how time-resolved data and finally cross-correlations are obtained. Additionally, we discuss how cross-correlations should be interpreted.

2.1. SHORT SUMMARY

To answer biological questions about dynamic fluctuations in single *E. coli* cells, we acquired time-resolved data on gene expression and growth. In a typical experiment we obtained movies of growing *E. coli* cells with an automatized microscope. The cells contained fluorescent reporter genes to monitor gene expression. Subsequently, the movie was segmented automatically to detect cells and then cells were tracked from image to image. We then determined time-resolved growth rate and protein expression for each cell at sub-cell-cycle resolution. Finally, we calculated cross-correlations between the fluctuating signals, taking the tree-like data structure of the growing and dividing colony into account. Cross-correlations are a useful tool to infer transmission of noise between fluctuating signals. In the following sections we describe these experimental and analysis methods in detail. The sections 2.2 - 2.3.6 are based on our publication [56] (chapter 6), but provide more in-depth explanations and were extended with additional methods developed. As the interpretation of cross-correlations is not trivial and can be misleading, we conclude this chapter by giving guidelines on interpretation. Most of the experiments throughout this thesis are based on the methods introduced here. Details on bacterial strains and deviating analysis methods will be explained specifically in each chapter.

2.2. EXPERIMENTAL METHODS

2.2.1. BACTERIAL STRAINS

We used strains based on wild-type *E. coli* MG1655 (*rph-1 ilvG- rfb-50*) for our experiments. MG1655 is a common laboratory strain and has been fully sequenced [36]. Of its three mutations, only *rph-1* is relevant in laboratory conditions, resulting in lower *pyrE* expression and a growth defect of ca. 10% due to pyrimidine starvation [57]. Full growth rate can be recovered by adding the pyrimidine derivative uracil to the medium [57].

Gene expression was measured by inserting fluorescent reporter genes into the bacterial DNA. There are many different reporter genes available and in the following chapter 3 we will analyze the quality of different selected fluorescent proteins. The reporter was placed either under direct control of a promoter or translationally fused to a native gene (more details in chapter 3.2.3). Cloning of fluorescent reporter genes into the bacterial chromosome was achieved by homologous recombination, following the Datsenko and Wanner protocol [58]. Briefly, the external linear DNA was inserted into bacteria by electroporation [59]. This technique exposes cells to high voltage (kV) for short times and thus creates temporary openings in the cell membrane, through which DNA can be taken up. The external DNA sequence was constructed such that its end sequences overlap with the sequence at the desired chromosomal location. Then, recombinase enzymes (λRed), which were provided on a plasmid, can exchange native DNA with the external DNA on the chromosome. The inserted DNA typically also contained a gene encoding an antibiotic resistance protein. Successful insertions can then be detected as growing colonies on a selective agar plate (i.e. with added antibiotics). Some constructs (e.g. rRNA reporters, chapter 7) were difficult to create and accompanied by a lot of false-positive colonies. Then, fluorescent colonies were directly searched for by imaging the

complete agar plate on a microscope at low magnification.

In a few constructs a two-step cloning and selection protocol was used to remove the antibiotic resistance cassette in the second selection step [60]. Some of the used strains in this thesis were gifts of other research groups, which will be mentioned in the specific chapters.

2.2.2. GROWTH MEDIA AND PREPARATION OF CELLS

In most of the time-lapse experiments cells were grown in M9 minimal medium (47.7mM Na_2HPO_4, 25mM KH_2PO_4, 9.3mM NaCl, 17.1mM NH_4, 2.0mM $MgSO_4$, 0.1mM $CaCl_2$) supplemented with 0.2mM uracil. We added one of various organic components as carbon and energy source: 0.1% lactulose (=2.9mM), 0.1% lactose (=2.9mM), 0.1% glucose (=5.6mM), 0.1% succinate (=8.5mM), 0.1% maltose (=2.9mM) or 0.18% acetate (=30mM). These carbon sources supported growth rates ranging from 0.2-1dbl/hr. For faster growth conditions we used rich medium ("MOPS EZ Rich Defined") from Teknova supplemented with 0.2% glycerol or glucose. In many experiments the *lac* promoter P_{lac} controlled expression of fluorescent reporter genes. For induction of P_{lac} , various levels (0-200μM) of the gratuitous inducer IPTG (Isopropyl β-D-1-thiogalactopyranoside) was added. If not mentioned otherwise, chemicals were obtained from Sigma Aldrich (Sigma Aldrich Chemie BV., the Netherlands).

Bacteria are stored as glycerol stock at -80°C which keeps them viable for many years. In the morning before the experiment, cells were inoculated into TY (tryptone yeast) medium and grown at 37°C. At an OD of 0.02-0.50 (optical density at 600nm), which required around seven hours of growth, the cells were strongly diluted into the defined experimental medium (see above). Cells were diluted into three different final concentrations to ensure that one of the flasks would contain exponentially growing cells on the next day. TY concentrations were thereby diluted to <0.05 vol% and remaining TY was consumed by the cells. Cells were grown overnight and the following day a flask that contained still exponentially growing cells (OD<0.2) was chosen and diluted again to an OD≈0.005, which was then used for microscopy.

To observe cell growth for multiple hours, cells need to stay in place and have to grow into a single layered microcolony. Therefore, we confined growth of bacteria to a narrow space between a glass slide (coverslip) and a polyacrylamide gel (Fig. 2.1) [44, 61]. Polyacrylamide gels fulfill the same role as the more commonly used gels based on agarose. That is, they confine cells, are soft enough to not disturb growth and act as medium reservoir. However, gels based on polyacrylamide are easier to handle and cannot be metabolized by bacteria - contrary to the sugar-based agar gels [62]. We produced the gels by mixing 1.25 ml 40% acrylamide, 3.7 mL water, 50 μL fresh 10% ammonium persulfate, and 5 μL TEMED, then poured 900 μL into a cavity glass slide and closed it with a coverslip (same geometry as glass chamber for microscopy, see Fig. 2.1). The glass was previously silanized to avoid sticking of the solution. Polymerized acrylamide is safe but monomers are toxic and we therefore prepared the gels very carefully (gloves, fume hood). Polymerization started within minutes after addition of TEMED, and after one and a half hours the gel was cut into pieces of ca. 5mm x 10mm. We stored the gel pads in water and washed them to remove monomers by exchanging the water. In sterile water, gels can be stored for several months. Before the actual experiment, a gel was

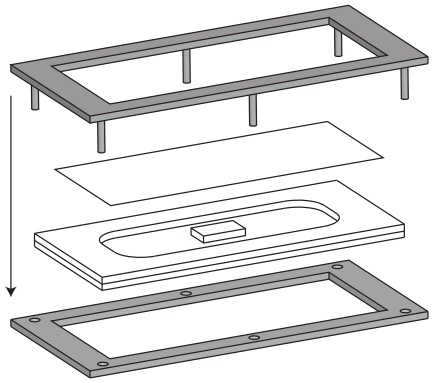


Figure 2.1: Sample chamber for microscopy. A hole glass slide (manufactured by the mechanical workshop of Amolf) is glued onto a microscope slide with silicon grease. The gel pad is transfered to the center of the glass chamber and 1µL of cells is pipetted onto it. A coverslip closes the chamber. The metal holders (gray), which are subsequently fixated by screws, clamp it. For microscopy, the sample chamber needs to be turned upside down.

transfered to a falcon tube with the desired medium and washed three times with fresh medium.

2.2.3. TIME LAPSE MICROSCOPY

To prepare the sample for microscopy, we transfered all required components to the warm room. The presoaked polyacrylamide gel was placed into a glass chamber and approximately 1µL of the diluted cells (OD=0.005) were applied onto it (Fig. 2.1). The gel was then covered with a coverslip and sealed tightly with a metal clamp to avoid drying out. The large oxygen reservoir relative to applied amount of bacteria allowed for exponential growth for many hours. Cells were distributed sparsely on the gel pad and single cells typically grew into microcolonies of several hundred cells before forming multilayers or interfering with other colonies. Growth rate under the microscope was comparable to bulk growth rates measured with a plate reader. In some conditions, however, multilayers formed rather quickly, which then made analysis impossible. The formation could be delayed by adding 0.001% of the detergent Tween20 to the medium. Addition of Tween20 did not lead to growth reduction. On the contrary, we found that it could sustain bacterial growth (0.5 dbl/hr) due to its organic nature, but we did not observe co-utilization in the presence of a primary carbon source. Still, caution should be taken for slow growth conditions or when the type of carbon source is important.

In a few experiments the described static chamber could not be used because a medium switch was required. Then, the chamber was replaced with a microfluidic setup [44] which will be described together with the experiment in chapter 3.

In most of the experiments cells were imaged with an inverted TE2000 microscope from Nikon, using a 100X oil objective (Nikon, Plan Fluor NA 1.3) and 1.5x intermediate magnification. Images were acquired with a cooled CCD camera (Photometrix, CoolSnap HQ) or a CMOS camera (Hamamatsu, Orca Flash 4.0). For both cameras the pixel size corresponded to around 41nm. A xenon arc lamp (Sutter, Lambda LS) was used for fluorescence illumination and was connected to the microscope with a liquid light guide (Sutter). Fluorescence filters were obtained from Chroma. We used 49001 for CFP like proteins, 41001 (older series) for GFP, 49003 for YFP and 49008 for red emitting proteins like mCherry. Shutters were computer controlled (Sutter, Lambda 10-3 with

SmartShutter) and the microscope stage was automated (Märzhäuser). The microscope was located inside an incubation chamber (Solent) to allow for a constant temperature of 37°C. Microscope and automated image acquisition were controlled with MetaMorph software (Molecular Devices).

Some of the experiments were performed on a newer version of the microscope (TI-E) with very similar functionality but some differences in equipment. It was equipped with a CMOS camera (Hamamatsu, Orca Flash 4.0), LED lamp for fluorescence illumination (Lumencor, SOLA II) and liquid light guide. Fluorescence filters were identical except for GFP (Chroma 49002). LED lamps were software controlled and phase contrast light path was controlled by a shutter (Sutter, Lambda SC Smart Shutter). The incubation chamber was fabricated in house. Microscope and image acquisition were controlled with Micromanager software [63].

At the beginning of each experiment, we searched the gel for isolated cells, selected 4-12 of them (depending on desired imaging loop frequency) and saved their stage positions. Then, an automated script acquired phase contrast and fluorescence movies of the growing cells at each position and for up to 40 hours. Each loop, the routine automatically refocused based on image contrast, using the Brenner algorithm. We acquired phase contrast images at three different heights and at a frequency of 30-50 images per cell-cycle and fluorescence images at roughly 6 images/cell-cycle. Illumination time for fluorescence images was kept as short as possible to reduce photodamage and bleaching. To increase signal-to-noise ratio for fluorescence images, 2x2 binning was used (effective for CCD cameras) and all disturbing light sources (room light, computer screen) were turned off.

2.3. DATA ANALYSIS

2.3.1. CELL SEGMENTATION AND TRACKING

After termination of the experiment, we analyzed the acquired movies offline with custom software written in Matlab (MathWorks). The software is based on Schnitzcells [39] and was largely changed and extended in our group. First, the movie was restricted to the frame range with single-layer growth and before cells outgrew the field of view. That meant usually around 500 cells in the last image, respectively 9 generations of growth. Then, a segmentation algorithm automatically determined cell areas (see Fig 2.2): The three phase contrast images were first averaged (a slightly blurred image can decrease oversegmentation artifacts). Then, the colony region was located by finding the region of most contrast variation. Primary cell outlines were determined automatically by applying a Laplacian of Gaussian filter on the phase contrast images. Segmented areas that were much smaller than a typical cell were automatically discarded. Accidentally connected cells were then separated in multiple ways: Morphological opening separated barely connected cells. Suspiciously long cells were additionally cut based on concavity of the cell outline. Subsequently, all cell outlines were filled by seeding them and applying the watershed algorithm. This segmentation procedure had a low error rate in cell detection, but all images were visually checked and, if necessary, corrected manually.

After segmentation cells were tracked between consecutive images. Briefly, for each cell in the next frame, an algorithm determined the nearest cell of the previous frame.

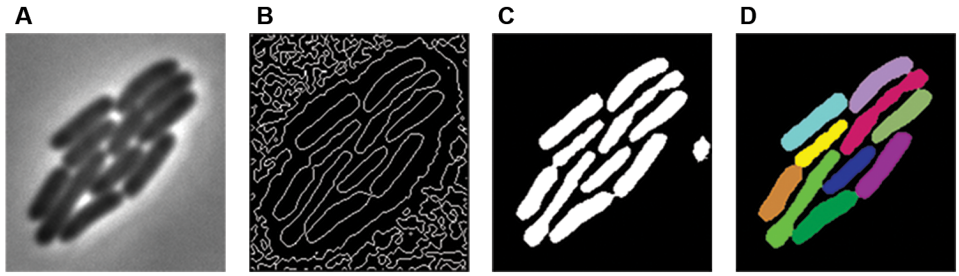


Figure 2.2: Steps during image segmentation. (**A**) Phase contrast image is the start image for segmentation. (**B**) Applying a Laplacian of Gaussian filter to the phase contrast image detects outlines of cells and background variations. (**C**) After removing lines outside of the colony region mask and small areas, cell areas are properly detected. However, cells that are in contact are not separated into distinguishable cells yet. (**D**) After morphological opening and cutting of long cells, the final segmentation result is obtained by seeding and applying a watershed algorithm.

To determine what is "nearest", we used two alternative criteria: The standard routine determined the skeleton of each cell (morphological thinning) and three characteristic points located at 1/4, 1/2 and 3/4 along that line. Then, the mean square distances between this triplet and each of the triplets from previous-frame cells were determined (Fig. 2.3A). The next-frame cell was linked to the previous-frame cell with the smallest distance. In some cases, for example for filamentous (dividing) cells or when cells moved a lot, the routine would produce insufficient results. Then, we applied a more robust but slower algorithm that searched for minimal distances between new-frame cell centers and previous-frame cell areas (Fig. 2.3B). Both routines took the increase in total colony size into account and automatically checked for tracking errors such as disappearing cells or cells dividing into three daughters. Each tracking result was checked manually and we found that for efficient tracking around 30 frames per cell cycle were required. Once tracking was finished, we obtained a complete lineage tree of the colony with full information on history of cells. In the next steps, growth and expression data were determined for each cell.

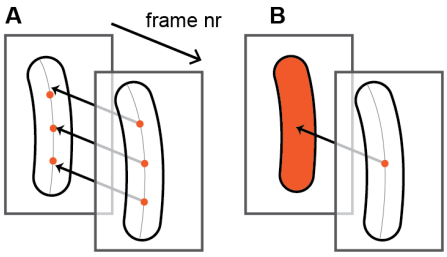


Figure 2.3: Illustration of cell tracking. Displayed are the two different algorithms used for an example of non-dividing cells. (**A**) Standard algorithm: A cell is characterized by three points (orange) along its skeleton (thin gray line). Each new cell is connected to the previous-frame cell whose characteristic points lie closest by (arrows). (**B**) Modified algorithm. Shift between cells is determined as distance of next-frame cell center to previous-frame cell areas and cells with the smallest distance were linked. In the example the distance is zero.

2.3.2. DETERMINATION OF CELL LENGTH AND GROWTH RATE

For calculation of growth rate we first determined cell length over time and then calculated growth rate as exponential fit to the cell length. Cell length is similar to cellular volume because the rod shaped *E. coli* bacteria only grow along their long axis [32]. Additionally, we tested that our analysis was robust to different methods for cell size determination which also take cell width into account (see chapter 6, Fig. 6.6). We determined cell length at high resolution by analyzing the phase contrast profile along the cell axis [61, 64]: To find the cell axis, we fitted a third degree polynomial $f(x)$ through the cell area (Fig. 2.4A). Choosing a non-linear function took into account that even short cells can be slightly curved. In experiments with filamentous cells which can be strongly bent a polynomial of up to seventh degree was used to fit the complex shapes. The cell poles (denoted x_0 and x_1) were determined with a method similar to a sliding-window approach (Fig. 2.4B): For each point on the cell axis the sum of distances between that point and the 25 closest segmentation pixels, here termed silhouette proximity, was computed. For points located in the center of the cell area this sum is constant. However, when approaching the cell poles the sum rapidly increases and diverges. We defined the cell pole as the location where the silhouette proximity increased 10% above the value in the cen-

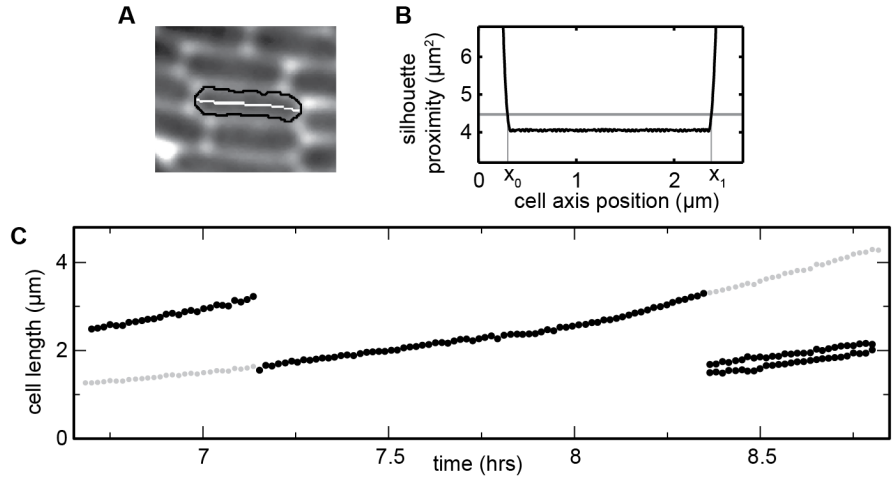


Figure 2.4: Determination of cell length and growth rate. (**A**) Phase contrast image with segmented cell outline (black). The cell axis (white) is determined by fitting a third degree polynomial through the cell area. (**B**) Cell length determination. The silhouette proximity (distance of a point on the cell axis from the 25 closest segmentation pixels) is constant (4.06 μm^2) in the center of the cell. At the poles it rapidly increases and the pole locations x_0 and x_1 where defined where the threshold 4.47 μm^2 was crossed. (**C**) Determination of growth rate of a single cell. The length of a single cell, its parent and its offspring is plotted over time (dark circles). Instantaneous growth rate is determined by fitting an exponential function to cell length for a fraction of the cell-cycle. At the beginning and end of each cell cycle, length data of the parent or the offspring are used for this fitting process (gray circles). Figure and description are from our publication [56] and slightly modified.

ter of the cell. The advantage of this involved method for determination of cell poles is its robustness against variations in segmented cell outline and erratic pixels along the cell axis. The cell length L was then determined by numerically integrating the cell axis curve:

$$L = \int_{x_0}^{x_1} \sqrt{1 + f'(x)^2} \quad . \tag{2.1}$$

Growth rates μ were determined by fitting an exponential function to the cell length vector:

$$L(t) = L_0 e^{\mu \cdot \ln(2) \cdot t} = L_0 2^{\mu \cdot t} \quad . \tag{2.2}$$

Here, L_0 and μ are the fitting parameters. The average growth rate of a cell was calculated by fitting all length data available for that cell. More importantly, instantaneous growth rate $\mu(t)$ at sub-cell-cycle resolution was determined by fitting the length data within a sliding time window. The window was centered around t and the size of the window was chosen to be one third of the mean interdivision time. Using this size is a compromise between smoothing sufficiently strong such that growth fluctuations are not dominated by segmentation uncertainties, but still capturing the fast growth fluctuations with typical time scale of around half a cell cycle (see chapter 6). We also note that the absolute number of frames within that window has little influence on determined growth characteristics (tested with 10-60 images per cell cycle), but the window size relative to the cell-cycle time is relevant. For time points where the sliding window overlapped with a division event we took parent resp. daughter information into account: If a cell divided, its growth rate was extrapolated by summing up the length of the daughter cells (see Fig. 2.4C). At the beginning of the cell-cycle, length data was extrapolated by using the measurements of the mother cell and roughly dividing the length by two. More precisely, the division factor was the fraction of the daughter size relative to the mother size, to account for asymmetric division events. Example traces of fluctuating growth rate can be found in Fig. 2.6B.

2.3.3. PROTEIN CONCENTRATION AND PRODUCTION RATE

Protein concentration was measured in (arbitrary) fluorescence units. Phase contrast images and fluorescent images were slightly shifted (<10 pixels) relative to each other. To determine that deviation, each fluorescence image was overlaid with the segmentation image at various shifts. Then, the sum of all fluorescence intensity values within the segmented outline was determined. The optimal shift was calculated as the value for which that sum was maximized for the whole movie and images were shifted accordingly.

Fluorescence images were then corrected for camera noise, which produced a rather homogeneous offset signal, and for uneven illumination of the field of view. The correction images were termed I_b for background image and I_s for shading image (uneven illumination). The background image was obtained by acquiring multiple images with closed camera shutter (that is, no incoming light on the camera) and averaging them. For the shading image, we acquired and averaged images of a thin and homogeneous layer of fluorescent dye deposited between a coverslip and a microscope slide. As the illumination pattern depended somewhat on the fluorescence filter used, we acquired different shading images for all filters, using the dye fluorescein for cyan, green and yellow fluorescence filters and rhodamine for the red filter. The corrected output image I_c

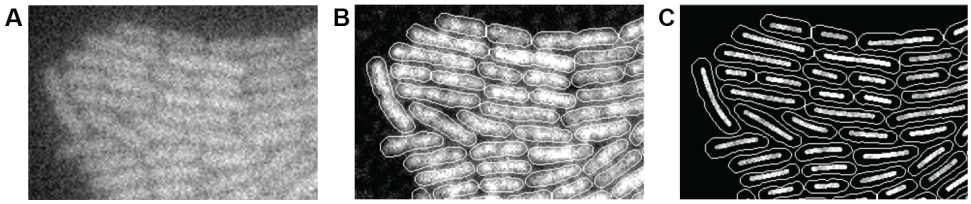


Figure 2.5: Determination of protein concentration and production rate. (**A**) Initial fluorescence image. (**B**) Image after background correction, shading correction and deconvolution with a point spread function. Total cell fluorescence is determined as the sum of fluorescence values within the cell outline. (**C**) Fluorescence concentration was usually determined by averaging total fluorescence within a central box. For exceptions, see main text. Figure and description are from [56] and slightly modified.

can then be calculated from the input image I via

$$I_c = \frac{I - I_b}{I_s - I_b} \quad . \tag{2.3}$$

The imaging system produces a slightly blurred image of the sample which is effectively a convolution of the ideal image with an unknown point spread function. To invert blurring, we first experimentally determined a point-spread function using 0.02 µm sized beads (FluoSpheres from Invitrogen). Then the image was deconvolved with a stochastic algorithm in Matlab (Lucy-Richardson algorithm), see Fig. 2.5A-B. One effect of image blurring is, for example, that cells located in the center of the microcolony appear brighter. The deconvolution algorithm could fully remove that bias.

The total fluorescence of each cell was determined by summing up all fluorescence pixels within the cell outline. Values were normalized for binning and illumination time to make comparison between different experiments (with the same fluorophore) possible. Usually, the protein concentration is determined by dividing total fluorescence by cell area. However, since we already used cell size to determine growth rate it would be possible that then spurious artificial correlations occur between concentration and growth. Therefore we determined protein concentration in a different way: Essentially, we placed a box in the center of the cell and calculated fluorescence concentration by dividing total fluorescence within that box by its size (Fig. 2.5C). The box was centered around the cell axis and 0.4 µm wide. It extended until 0.3 µm away from the cell poles. In some reporter constructs proteins were localized within the cell (e.g. ribosomes, see chapter 7) and determination of concentration via the box could then be inaccurate. In these cases, we determined concentration directly by dividing total fluorescence by the cell area. The same approach was used when absolute units of concentration relative to production rate mattered (see chapter 5). Independent of determination method, the concentration was further corrected by subtracting background fluorescence (determined from pixels outside the microcolony) and, if needed, autofluorescence (determined from fluorescence of wild-type MG1655). Protein production rate at time t was calculated from three consecutive total fluorescence data points centered around t. We determined production rate as slope of a linear fit to these points, taking parent and

daughter information into account for time points close to division events.

2.3.4. CORRECTION FOR CELL-CYCLE FLUCTUATIONS

Protein production rate depends strongly and systematically on cell-cycle phase and, to a lesser extent, protein concentration and growth rate depend on the phase as well. A detailed analysis of these quasi-periodic fluctuations can be found in chapter 5. In the remaining chapters, we focused on "random" noise and therefore automatically corrected these three signals for cell-cycle fluctuations. We first determined the phase for each data point, with 0 being cell birth and 1 being cell division. For cells with incomplete observed cell-cycle we estimated the phase based on the available cell length data points. We determined the cell-cycle dependence of each signal by fitting a third degree polynomial to the phase-vs.-signal dataset of all cells from an experiment. Choosing that function was heuristic but it was general enough to capture typical dependencies (increase, decrease or both) sufficiently well. Then, all data points were cell-cycle corrected by subtracting the fitted phase-dependent average.

An alternative method (described in detail in chapter 5) used spline interpolation instead of a 3rd degree polynomial to find the average dependency of a parameter on cell-cycle phase. The signal was first binned according to its phase and then a spline polynomial fitted between the binned averages. It could produce equal or better results because it allowed for more general functional shapes. It was, however, only semi-automatic so we resolved to applying spline interpolation only if automatic fitting was insufficient or we were specifically interested in the functional shape (chapter 5).

2.3.5. DATA STRUCTURE AND SELECTION

The analysis methods described so far are the standard analysis steps used for data extraction in every experiment. Summarized, as result we obtained a branched lineage tree with full history information of all cells (except the first one). For every cell at every acquired time point the data structure contained instantaneous growth rate, protein production rate and concentration. Additionally, it contained many more parameters such as cell length, number of the sister cell, location within the colony etc. For example, the data allowed to investigate the distribution of a quantity at one given time point, as well as the time evolution of its fluctuations over several generations (see Fig. 2.6). We typically obtained several hundred lineages (defined by the final number of cells), a time span of nine generations, and a few thousand unique data points. The datasets usually contained ca. 300-500 cells with complete observed cell-cycle.

For further analysis, the data was restricted to steady state conditions, that is, constant colony average values for growth rate etc. over time. Often, the full dataset was used but sometimes cells grew slower in the first one or two generations (presumably a stress response to deposition on the pad). Then, a time window of steady-state growth was introduced. Some microcolonies also contained a few elongated cells, cells that grew on top of each other or that ceased to grow. These cells were excluded from further analysis. We also always tested whether spatial effects such as slower growth in the center of a colony existed, but these effects were usually absent and did not require additional data restriction.

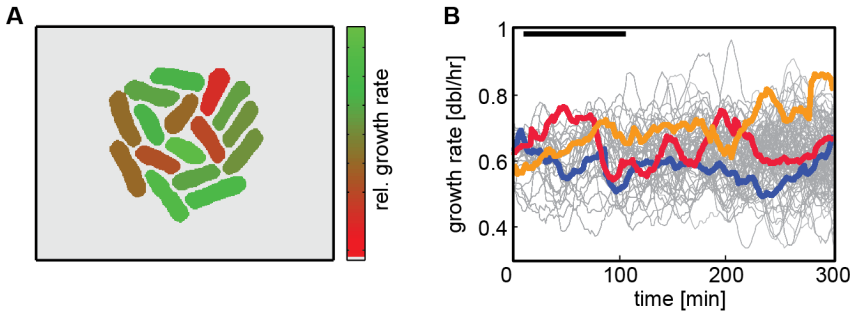


Figure 2.6: Typical data after standard analysis. (**A**) For every time point, the distribution of instantaneous growth rates within the colony is available. (**B**) Fluctuations of growth can be followed along cell lineages over time and for several generations. The black bar denotes average interdivision time. Examples are given for growth rate but also apply to any other quantity, such as protein concentration.

Further analysis methods specific to the projects will be described in each chapter. Only temporal cross-correlations, which we used often and for several projects, will be introduced in the next section.

2.3.6. CROSS-CORRELATIONS

Cross-correlation analysis is useful to reveal time delays and potential interaction between two fluctuating signals. Consider for example two signals $A(t)$ and $B(t)$ (either one could be growth rate, concentration etc.) as in Fig. 2.7A. Signal $B(t)$ fluctuates similarly to $A(t)$ but lags behind with a delay τ_D. This delay could for example occur when $B(t)$ is positively regulated by $A(t)$ but reacts slowly. Computing a standard (Pearson) correlation, which uses simultaneous time points for both signals, will yield a low correlation coefficient. Naturally, it cannot reveal the time delay. The cross-correlation, however, determines a correlation coefficient for every possible time delay τ between the signals and will peak at the characteristic delay τ_D (Fig. 2.7B). Thus, it can reveal interactions and noise transmission between the signals (see also section 2.4.2). Additionally, the steepness of the correlation curve (how fast the peak decays) indicates the time scale of fluctuations. This is especially useful for auto-correlation functions where a fluctuating signal $A(t)$ is correlated with itself (section 2.4.1): Fitting, for example, an exponential decay function to the auto-correlation reveals the characteristic time scale of fluctuations in $A(t)$.

CROSS-CORRELATIONS ALONG A LINEAGE

In our experiments we are interested in cross-correlations (and auto-correlations) between the fluctuating growth rate (μ), protein concentration (E) and protein production (p_E) (see also Fig. 2.6B). The definition of cross-correlations introduced here follows the description in our publication [56] closely (text on cross-correlations therein is by Daan Kiviet). For computation of cross-correlations we used the following notation (Fig. 2.7C): A microcolony consists of M lineages, each containing N data values which are sepa-

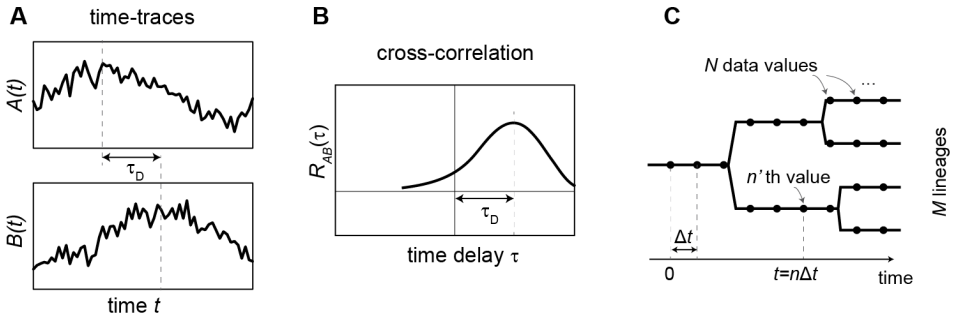


Figure 2.7: Concept of cross-correlations and definition of variables. (**A**) Two example traces $A(t)$ and $B(t)$ fluctuate over time. The signal $B(t)$ roughly follows $A(t)$ with a time lag τ_D. (**B**) Schematic cross-correlation function $R_{AB}(\tau)$ for the signals from (**A**) . The cross-correlation peaks at the characteristic delay time τ_D and can thus reveal delayed interaction. The standard (Pearson) correlation is the cross-correlation at delay $\tau = 0$ and could not have detected a coupling of the signals. (**C**) Variables needed for determination of cross-correlation for a branched data structure and observation at discrete time points. M lineages exist (thick lines) and each lineage contains N data points (black circles) which are not all unique (see also Fig. 2.8). Data is acquired at a time interval Δt and the n'th data point is at time $t = n \cdot \Delta t$.

rated by time intervals Δt. Then, the n'th data value originates from time point $t = n \cdot \Delta t$. For each signal we calculated the deviation of it from the population mean (at that time point). These deviation time traces were termed "noise" (not to be confused with noise as quantification of distribution width):

$$\begin{aligned} \tilde{\mu}_n &= \mu_n - \frac{1}{M}\sum_{m=1}^{M} \mu_{n,m} \\ \tilde{E}_n &= E_n - \frac{1}{M}\sum_{m=1}^{M} E_{n,m} \\ \tilde{p}_{E_n} &= p_{E_n} - \frac{1}{M}\sum_{m=1}^{M} p_{E_{n,m}} \quad . \end{aligned} \tag{2.4}$$

Here, the sum runs over all lineages. Usually the mean would be determined as time-average along each lineage [27] and in steady state these two methods are theoretically equivalent. Practically, mean determination along each lineage is unreliable due to small data size (ca. 50 points per lineage). It also turned out that using the population mean per time point in equation 2.4 was more robust for the few experiments in which steady state could not be perfectly achieved.

We can then determine the cross-covariance C within a single lineage between two signals at a time lag $\tau = r \cdot \Delta t$ [65]. The concept of cross-covariance is similar to a variance but it quantifies how much two different signals vary together and is generalized to non-zero time lags between the signals. We use as example E and μ but the definition applies

analogously to other quantities:

$$C_{E\mu}(r\Delta t) = \begin{cases} \frac{1}{N-r} \sum_{n=1}^{N-r} (\widetilde{E}_n \widetilde{\mu}_{n+r}) & \text{if } r \geq 0 \\ C_{\mu E}(-r\Delta t) & \text{if } r < 0. \end{cases} \tag{2.5}$$

The sum runs over all time points for which data also exists at a time τ later. To determine the cross-correlation within one lineage, the cross-covariance needs to be normalized with the standard deviations (σ) of both signals. σ is equivalent to the square root of the auto-covariance at time lag $\tau = 0$:

$$R_{E\mu}(r\Delta t) = \frac{C_{E\mu}(r\Delta t)}{\sqrt{C_{EE}(0)C_{\mu\mu}(0)}} \quad . \tag{2.6}$$

COMPOSITE CROSS-CORRELATIONS

Multiple lineages can be combined into a composite cross-covariance by averaging them:

$$C_{E\mu}(r\Delta t) = \begin{cases} \frac{1}{M}\frac{1}{N-r} \sum_{m=1}^{M} \sum_{n=1}^{N-r} (\widetilde{E}_{n,m} \widetilde{\mu}_{n+r,m}) & \text{if } r \geq 0 \\ C_{\mu E}(-r\Delta t) & \text{if } r < 0. \end{cases} \tag{2.7}$$

However, simply averaging all data points with equal weight ignores the fact that many of the data points are redundant and appear in several lineages (Fig. 2.8). Therefore, we introduced a correction factor that ensured that independent pairs of data points are all weighed equally strong in equation 2.7. For each pair we counted the number of lineages in which these data points are used (λ). Multiple contributions of the same pair of data points can then be corrected by weighing each summand with the inverse of λ, which we term weighing factor:

$$w_{n,m,r} = \frac{1}{\lambda_{(n+r),m}} \quad . \tag{2.8}$$

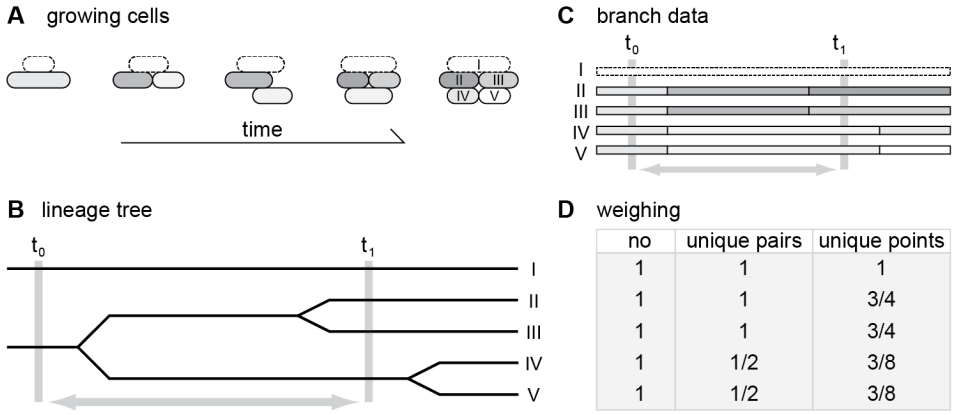
This weighing results effectively in the same calculation as in Dunlop et al. [27]. $w_{n,m,r}$ depends on the time point n, the lag r and the lineage m investigated. Many data pairs consist, however, of two points which are reused a different number of times in other lineages (see lineages II and III in Fig. 2.8). We used an additional correction to account for these "half-unique" data points:

$$w_{n,m,r} = \frac{1}{2\lambda_{n,m}} + \frac{1}{2\lambda_{(n+r),m}} \quad . \tag{2.9}$$

Using this weighing, we can calculate the corrected composite cross-covariance function:

$$C^M_{E\mu}(r\Delta t) = \begin{cases} \frac{1}{w_{tot}(r)} \sum_{m=1}^{M} \sum_{n=1}^{N-r} (E_{n,m} \mu_{n+r,m} w_{n,m,r}) & \text{if } r \geq 0 \\ C^M_{\mu E}(-r\Delta t) & \text{if } r < 0, \end{cases} \tag{2.10}$$

with $w_{tot}(r) = \sum_{m=1}^{M} \sum_{n=1}^{N-r} w_{n,m,r}$. The composite cross-correlation was then calculated by normalizing with σ as in equation 2.6.



no	unique pairs	unique points
1	1	1
1	1	3/4
1	1	3/4
1	1/2	3/8
1	1/2	3/8

Figure 2.8: Extracting and weighing lineages from a branched data set. (**A**) Depiction of a growing microcolony over time, starting with two cells on the left and growing into five cells on the right. (**B**) A lineage tree of the data shown in (**A**) . The tree starts with two lines (left), indicating the two starting cells, and at each division the line splits, resulting in five cells at the end (right). (**C**) Five lineages can be extracted from the data. Note that most lineages share part of their data. When correlating data points from t_0 with t_1, one pair consists of completely independent data points (lineage I). Two lineages provide exactly the same pairs of data points (lineages IV and V), and two lineages only share a data point at t_0 (lineages II and III). (**D**) Different types of weighing for the correlation of data points from t_0 with t_1 as used in equation 2.10. No: each lineage is weighed equally. Unique pairs: weighing such that only comparisons between unique data pairs are used. Unique points: lineages II and III are not completely independent, which can be corrected for by the weighing from equation 2.9. This weighing method was used throughout the thesis. Figure and description are taken from our publication [56].

To determine error bars for the cross-correlations, we split the microcolony into four independent branch trees which originated from the first four simultaneously present cells in the colony. We then determined the composite cross-correlation for each of these lineage trees independently. Subsequently, they were averaged and the standard deviation determined.

2.4. INTERPRETATION OF CROSS-CORRELATIONS

2.4.1. TIME SCALES

Cross-correlations are a powerful tool to learn about cellular fluctuations but have to be interpreted with care. Therefore, we discuss in this and the next section two important aspects: how to determine the time scale of fluctuations and how to interpret time delays and examine causality.

Stochastic fluctuations can be characterized by their amplitude and frequency. Slower fluctuations may be more relevant biologically because they persist and do not average out easily [5, 66, 67]. To determine the characteristic frequency of fluctuations, it is helpful to analyze autocorrelation functions (i.e. correlation of a signal with itself). The auto-

correlation is per construction one at zero time delay. For biological processes it typically decays in an exponential fashion to zero for large time delays (Fig. 2.9A). Then, the exponent of an exponential fit yields the frequency of fluctuations, and its inverse is the characteristic time scale τ_{AC} of the fluctuations. This is also the time delay after which the autocorrelation has dropped to $1/e$, or alternatively $1/2$ when using base 2 for the exponential function.

In reality, stochastic fluctuations may occur on multiple fast and slow time scales [5, 67, 68] and a single exponential fit may be insufficient. Additionally, measurement uncertainties cause noise in the observed quantities. Experimental fluctuations such as focusing accuracy are probably uncorrelated between consecutive images and therefore appear as a fast noise component ("shot noise"). Whether fast fluctuations are real or experimental noise is very hard to determine. In both cases the autocorrelation function is multi-exponential (simulated in Fig. 2.9B). Fitting a single exponential function then determines an incorrect intermediate characteristic time scale (Fig. 2.9B gray line) and a multi-exponential function should be used (black line). For experimental data we found both single or (at least) double exponential decays, depending on the signal investigated. τ_{AC} of the fast component is very hard to determine (see also the ranking approach used in [5]) due to limited time resolution and lack of knowledge what part of these fluctuations are real. To determine the biologically more relevant τ_{AC} of the slow component, exponential fitting should be restricted to time delays at which the fast decaying component is basically zero.

While experimental noise can lead to a fast drop in autocorrelations, also an opposite effect exists that artificially broadens the correlation function: Subsequent data points are often not independent but calculated from partly redundant raw data. For example, for growth rate determination the sliding window of cell lengths overlaps typically up to 50% for consecutive data points. This results in a widening of the cross-correlation at small delays (simulation: Fig. 2.9C, experimental growth data: Fig. 2.9D green circles). For determination of the correct decay time, such artificially correlated data points should be excluded from the exponential fit (empty circles in Fig. 2.9C,D). For the window size that we typically used to determine growth rate, the artificial broadening often precisely canceled the initial fast drop of the autocorrelation (Fig. 2.9D red circles and line). Then, the autocorrelation appeared to be a single exponential. This canceling is however coincidental and for each fit it should be examined whether data points at small time delays need to be excluded.

We also tested whether the amount of overlap of the fit windows (amount of artificial correlation) affected cross-correlations. We chose $R_{p\mu}(\tau)$ as example and varied the window size for growth rate determination (see also Fig. 2.9D). Calculation of production rate was not changed. The amplitude of the cross-correlation increased with smoothing (Fig. 2.9E). However, the shape stayed very similar, suggesting that observed features such as asymmetry are robust to the choice of fit window. Further, the robustness of the shape indicates that the correlation is probably caused by slow growth fluctuations and that fast fluctuations (for example present in the blue dataset in Fig. 2.9E) do not contribute to the cross-correlation. Similar analysis of changing the production rate determination could further corroborate these indications.

Summarizing, when determining decay times one should take into account that mea-

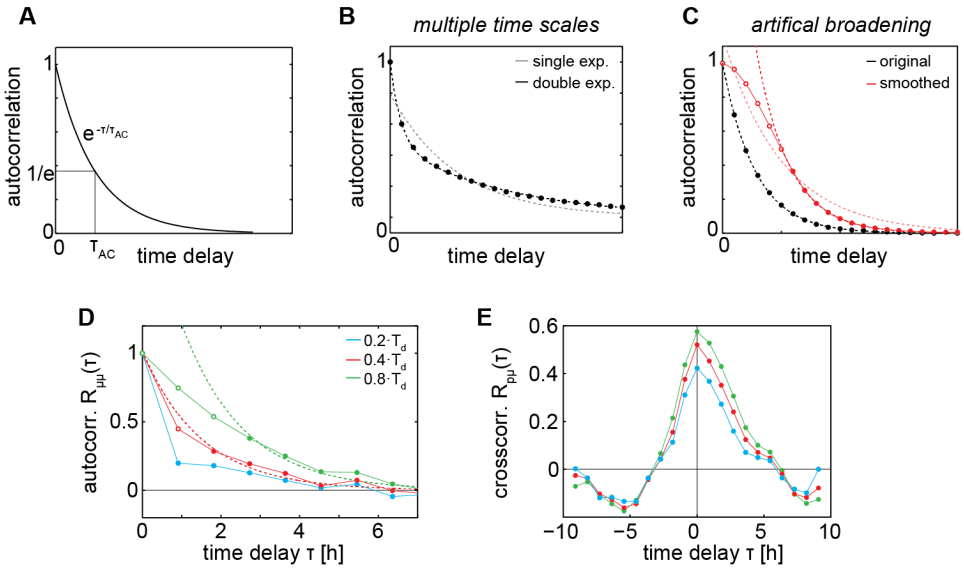


Figure 2.9: Extracting and interpreting time scales of fluctuations. (**A**) Illustration of a typical autocorrelation function that decays exponentially. After a characteristic time τ_{AC} the autocorrelation has dropped to $1/e$. (**B**) Autocorrelation of a simulated signal that fluctuates on two time scales that differ by a factor 10, with both noise sources contributing with equal intensity (circles). A single exponential fit (gray dashed line) fails, but a double exponential fit (black dashed line) can determine both decay times. Experimentally, the slow time scale can be determined by excluding all data points of the initial fast decrease and then fitting a single exponential function. (**C**) When data points are obtained from dependent measurements, the autocorrelation function is artificially broadened and a minimum time scale is imposed. An example is determination of growth rate where cell length data windows typically overlap for consecutive time points. This effect was illustrated by using simulated data (black circles) and smoothing it with a sliding window of size 7 (red circles). Using the complete smoothed dataset for decay time determination results in a too large τ_{AC} (light red dashed line). Excluding the artificially correlated data (empty red circles) recovers the correct τ_{AC}, which is identical for original and smoothed data (dashed black resp. red line). (**D**) Autocorrelation of growth rate for cells grown on acetate. Growth rate was determined with sliding windows of different sizes, denoted as fractions of interdivision time T_d. The red data corresponds to the window size used in this thesis. Data points are spaced $0.36 \cdot T_d$ (which is large and spacing in many experiments was around $\approx 0.2T_d$). For small sliding windows, fast noise, which is probably measurement noise, dominates (blue, see also (**B**)). For large windows the correlation is broadened (green, see also (**C**)). If artificially correlated points (empty circles) are excluded the decay time can be determined independently of the sliding window size used (red dashed line: $\tau_{AC} = 1.5h$, green dashed line: $\tau_{AC} = 1.6h$). In the red data set excluded data points lie on the fitted line only by chance. Such a collapse onto one line was regularly observed for growth rates but rarely for production rates. For very small fit window size (blue) fitting is unreliable. (**E**) Cross-correlation of growth and production rate for cells grown on acetate. Growth rate determination was varied and is colored as in (**D**) , production rate determination was kept constant. The magnitude of the cross-correlation changes but the shape is very robust.

surement noise can cause a fast decrease at small time delays and that data points can be artificially correlated due to overlapping fit windows.

2.4.2. CAUSALITY

Cross-correlations are used to determine whether fluctuations in a cellular network propagate, and to infer properties of the network structure. For example, the shape and asymmetry of the cross-correlation function between the concentration of a protein and its upstream transcription factor can determine whether the regulatory link is active or not [27]. Cross-correlations between enzyme concentrations and growth rate can reveal whether a certain enzyme is limiting and fluctuations propagate to growth (see chapter 6). However, when inferring causal relations between two observables one has to be careful: The mere existence of a non-zero (cross-)correlation between two signals does not mean that a causal link exists (a statement so famous that it even has its own wikipedia page). For example, two otherwise unconnected signals can be correlated because they are affected by the same global fluctuations (see "common noise" in chapter 6 or "extrinsic noise" in [27]). A more complex scenario arises, for example, when an unknown upstream component X affects the two measured signals Y and Z, but with a different time delay (Fig. 2.10). Then, the cross-correlation R_{YZ} is asymmetric, suggesting a transmission from Y to Z but in reality no causal link exists. For further examples, see also Dunlop et al. [27].

Interpretation of cross-correlations gets more difficult the less is known about the underlying network which connects the two signals. Direct regulatory interactions such as between repressor and target are relatively straightforward to interpret. However, in reality many fluctuating cellular parameters are inaccessible or regulatory links are unknown [27]. This is especially true for the cellular growth rate which is somehow linked to production of proteins and cell mass but with unknown links. Nonetheless, one can test whether a causal link between two signals is likely to exist. This can be achieved by manipulating the (suspected) network structure by e.g. introducing knockouts or changing the active metabolic pathways. The changes should be chosen such that they will affect the regulation between the two signals or shift limitations to other enzymes. Consider for example the case that an enzyme fluctuation preceded growth fluctuations, suggesting a limitation originating from that enzyme. Then, the carbon source can be altered such that a different pathway without that enzyme should be limiting (by inferring from literature). If asymmetry in the cross-correlation curves changed accordingly, it supported the hypothesis that the enzyme was indeed limiting in the first scenario, i.e. a causal

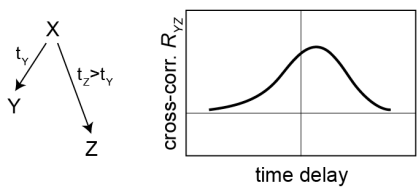


Figure 2.10: Asymmetric cross-correlations do not imply causality. Fluctuations in an unknown or unmeasured component X propagate to the observables Y and Z with different time delays $t_Z > t_Y$. Then, the cross-correlation R_{YZ} is asymmetric, suggesting a noise propagation from Y to Z. However, in reality this link is absent (see network structure on the left).

link existed (see for example chapter 6). Network structure and active pathways can be manipulated in multiple ways, such as by altering expression levels, changing nutrient sources or constructing knockouts.

An alternative approach (next to cross-correlations) for determining causal links is a method called Granger causality (GC) [69, 70]. This method has not been used in this thesis and would be interesting to investigate in future work. The GC measure determines how well a quantity Z can be predicted from its past values and tests whether the prediction is improved by taking the past of the suspected cause Y into account. If the prediction is significantly improved by Y, then Y is said to Granger-cause Z. This method is not equivalent to cross-correlations and can be more powerful. For example, consider the case that Y is a protein which is always expressed at the beginning of the cell-cycle while Z is expressed at the end, without any direct connection between Y and Z. Then their cross-correlation will show a time delay (falsely suggesting that Z is caused by Y) but the GC will be zero - because the past of Z already reveals the periodic cell-cycle fluctuations and Y does not add any extra information. It remains to be determined whether GC is also more powerful when Z only fluctuates randomly without an underlying rule (such as the cell-cycle timing), which is presumably the nature of fluctuations investigated mostly in this thesis. Note also that the confounding effects of an unknown upstream variable X (Fig. 2.10) cannot be revealed by the original definition of GC, and Y would be found to Granger-cause Z. However, more advanced versions of GC might be able to deal with such hidden variables [70].

I would like to thank Vanda Sunderlikova for constructing the *E. coli* strains used in this thesis. The growth protocol was developed by Daan Kiviet. The analysis software, obtained from [7], was extended in a first step by Daan Kiviet and in a second step by Philippe Nghe and the author.

3

CHARACTERIZATION OF FLUORESCENT PROTEINS

Throughout this thesis, we use fluorescent reporter proteins to measure gene expression in E. coli . These proteins exist in different variants that vary in spectral range and quality. In this chapter, we therefore characterize different fluorescent proteins. We first establish which quality criteria are important in our experiments and find several suitable candidates within the cyan to red emitting wavelength range. Then, we compare their performance experimentally in the categories cross-correlations, maturation time and photostability. We conclude that the tested GFP version is the best protein for single-label experiments. For double-label experiments we find that the yellow mVenus paired with a cyan emitting mCerulean variant are the best option based on the tested constructs.

3.1. INTRODUCTION

To measure expression of endogenous proteins over time we used fluorescent proteins that function as reporter [43]. Then, apparent expression can be measured in fluorescence units. To asses how well the reporter describes expression of the actual protein of interest it is crucial to understand the properties and quality of the construct and the used fluorescent protein. Ideally, an observed phenomenon should not depend on the specific choice of fluorescent reporter but in reality it may (see below, also [71]). Therefore, we tested and compared different fluorescent proteins. In this chapter, we first give a (non-exhaustive) overview of different existing variants of fluorescent proteins and reporter constructs. Then we determine which criteria define good fluorescent proteins for our type of measurements. In practice, choosing good proteins for our application was difficult because few studies exist that systematically compared different proteins (for good exceptions see [72] and [71]) and the existing studies focused on different quality criteria. Therefore, we constructed several strains with different fluorescent reporter genes to search for good reporters. We systematically tested and compared their performance based on the needs for our time-lapse microscopy studies.

3.2. OVERVIEW OF FLUORESCENT PROTEINS AND CONSTRUCTS

3.2.1. A BRIEF HISTORY OF DISCOVERY AND IMPROVEMENTS OF FLUORESCENT PROTEINS

We here start with a brief introduction to fluorescent proteins. For a very good extended overview we refer the reader to a scientific summary by Måns Ehrenberg [73], published as background information on the Nobel prize which was awarded for research on GFP [74]. Green fluorescent protein (GFP) originates from the jellyfish *Aequorea victoria* and was discovered by Shimomura et al. in 1962 [75]. Later, Chalfie et al. showed for the first time that GFP could be expressed in different organisms [43]. These discoveries opened unprecedented possibilities for imaging spatio-temporal processes in living organisms.

Today, GFP has been mutated to optimize properties and expression in laboratory organisms [76–78] and a whole range of spectral variants have been produced (blue: [79], cyan: [80, 81], yellow: [82, 83], see also Fig. 3.1A). Important improvements were for example increased brightness [76, 84] and folding efficiency at 37°C [77, 84]. Increased photostability allowed for longer imaging before proteins irreversibly transitioned to non-fluorescent states ("photobleached") [72, 85] (Fig. 3.1B). Especially for time-resolved imaging advancements in maturation time were crucial: When newly folded, GFP is still in a non-fluorescent state [79, 86] (Fig. 3.1B). The chromophore needs to transform through several stages before becoming functional, a process termed maturation [87]. One of these reactions requires molecular oxygen [79] and is considered rate-limiting, taking several hours in the native GFP [79] but only a few minutes in engineered variants [83, 88].

3.2.2. THE MODERN COLOR PALETTE

The discovery of spectrally distinct mutants of GFP made it possible to simultaneously image different structures or expression of different proteins in living cells [89] (Fig. 3.1A).

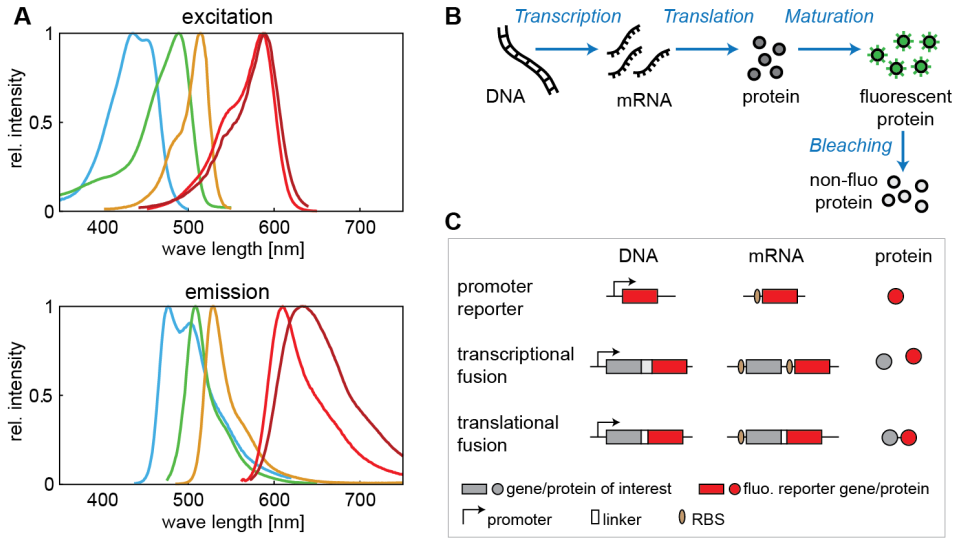


Figure 3.1: Fluorescent protein spectra, expression of fluorescent proteins and possible constructs. (**A**) Normalized excitation spectra (upper panel) and emission spectra (lower panel) of common fluorescent proteins. From short to long wave length: Cerulean (cyan), EGFP (green, we used the spectrally similar GFPmut2), Citrine (yellow, we used the spectrally similar Venus), mCherry (red), mKate2 (dark red). With standard filters cyan-yellow-red and green-red can be fully separated (taking either one of the red variants). Spectral data was obtained from Chroma Technology Corp (www.chroma.com). (**B**) Production and bleaching of fluorescent proteins. After transcription and translation proteins fold quickly but are still non-fluorescent (denoted in gray). Chromophore maturation results in fluorescent proteins (green). Illumination leads to photobleaching which irreversibly destroys the protein's fluorescence. (**C**) Different genetic constructs that report on gene expression. Upper row: Promoter (or transcriptional) reporter: A fluorescent reporter gene is directly under the control of the promoter of interest. Middle row: Gene of interest and fluorescent reporter are transcribed together but translated independently because each sequence contains its own ribosome binding sites (RBS). Bottom row: In a translational fusion gene of interest and reporter are translated together and usually connected via a linker.

Priorities for different applications differ but generally the shortest wavelength variants (blue fluorescent proteins, BFP) are suboptimal for imaging because of low brightness, fast bleaching and necessary excitation at short wavelengths [90]. Variants in the cyan emitting spectral range (CFP) exist however at high quality for a wide range of applications and a good protein is for example Cerulean [81]. Modern versions of GFP in the green emitting range are very good and standard reporters. Such improved versions are for example GFPmut2 [84], EGFP [76] or Emerald [90]. The longer wavelength yellow (YFP) variants are characterized by high brightness and modern variants include Venus [83] and Citrine [91]. Further extending the range to red wavelengths was however not possible by mutating wild type GFP. It could only be achieved after the discovery and cloning of naturally occurring red fluorescent proteins in corals and anemones [92, 93]. Unfortunately, red fluorescent proteins are often characterized by slow or even incom-

plete maturation [94, 95], probably also due to two required (slow) oxidation steps [94], and tend to oligomerize [96]. Extensive work of Tsien and coworkers [72, 97] resulted in strongly improved variants of the original DsRed (isolated from *Discosoma*), the most famous constructed variant being arguably mCherry. More recently, a derivative of a red fluorescent protein from *Entacmaea quadricolor*, named mKate2, was developed in the group of Chudakov [95, 98, 99] and promises high brightness and fast maturation. For a more complete overview on variants of fluorescent proteins and their quality, see [72] and [100]. Example spectra of suitable proteins for time-lapse imaging are displayed in Fig. 3.1A.

To image multiple processes in the same cell one needs to use spectrally distinct fluorescent proteins. With standard filters and full spectrum illumination (for example with a Xenon lamp) up to three different fluorescent proteins (cyan-yellow-red or green-red variants) can be simultaneously imaged. For different applications that number can vary: More colors can be imaged when including slowly maturing proteins [72] or when using narrow band (laser) excitation and filters. Using overlapping spectra and unmixing the multispectral images mathematically allows for simultaneous use of many overlapping fluorophores [101].

3.2.3. DIFFERENT REPORTER CONSTRUCTS

We are interested in employing fluorophores to report on the expression of native genes. The reporting can occur on different levels of the expression (transcription, translation), depending on the design of the construct. In the simplest case, the native gene and its promoter are not modified and the fluorescent gene is inserted under control of a copy of the promoter of interest (Fig. 3.1C upper row). The reporter may be inserted on a different location on the chromosome or provided on a plasmid (see for example the Alon plasmid reporter library [102]). The advantage of such a promoter reporter (also called transcriptional reporter) is that it minimally perturbs the native system and that it is relatively straightforward to construct. These reporters have proven very useful for high throughput comparisons of promoter strength and noise in constant and varying environments [102–104].

When investigating temporal fluctuations of expression on the single-cell level it is however desirable to have a more direct readout of expression of the native gene. We did not investigate whether in some cases a promoter reporter is sufficient for our purposes which might be worthwhile examining. But generally, such a reporter always suffers from the uncertainty to what extent a transient fluctuation is experienced by all identical promoters and to what extent only the local expression fluctuated. If the construct is located on a plasmid variation in plasmid copy number causes even more uncertainty. The most direct readout of native gene expression (and localization) can be achieved by translationally fusing native and reporter gene (Fig. 3.1C bottom row) [105]. Then, the reporter sequence is chromosomally inserted at the C-terminus (or N-terminus) of the native sequence and one joint mRNA is transcribed from the promoter. Both genes need to share one ribosome binding site (RBS), resulting in translation of both proteins connected together. Usually a linker of a few amino acids is inserted between the proteins to avoid interaction between them. Arguably, this method most accurately tracks expression fluctuations (and localization patterns) and was mostly used throughout this thesis.

The functionality of the fused protein of interest may be compromised and it should always be tested whether wild type performance for e.g. growth rate is still achieved.

An intermediate between these two types of constructs is the here termed transcriptional fusion (Fig. 3.1C middle row, note that nomenclature for types of constructs is not unified in literature): Native gene and reporter are transcribed into one mRNA but translated into independent proteins. Since low copy number of mRNA is thought to be the main source of noise [16, 21] the reporter is likely to track expression fluctuations well. This was confirmed for the *lac* enzymes for which a transcriptional and translational fusion was used in chapter 6. If localization instead of expression is of main interest and especially when no functional fusions exist [106] an alternative approach can be used: A translational fusion which still localizes is added to the cell, typically on a plasmid, additionally to the unmodified native gene.

3.3. QUALITY CRITERIA FOR FLUOROPHORES

Independent of the selected type of construct the fluorescent reporter needs to fulfill some quality criteria. The "[...] guide to choosing fluorescent proteins" of the Tsien lab [72] provides very good guidelines on desired properties and recommendable fluorophores (only the far-red variant mKate 2 was developed after publication of that guide). Which criteria are the most relevant depends on the type of application. Here, we summarize which properties are desired for imaging expression fluctuations *in vivo* with time-lapse microscopy (criteria based on [72]):

- **Maturation time**: Arguably the most important criterion for our applications. The maturation time should be short (roughly ≤10min) so that synchrony or time-lag between expression and growth fluctuations can reliably be detected (see also chapter 2.3.6).

- **Photostability**: Taking approximately six fluorescence images per cell cycle preferentially leads to small bleaching when compared to dilution of fluorescence by growth.

Further criteria are:

- **Brightness**: A fluorescence signals should be readily detectable also when transcribed from weaker promoters. The minimum required brightness depends on cellular autofluorescence in the specific wavelength range.

- **Monomeric protein**: The reporter is ideally monomeric to avoid artificial aggregation of fusion proteins [71].

- **Spectral separation**: Reporters for multi-color imaging should have minimal spectral cross-talk.

Several other criteria which were hardly restrictive for our application are [72]: Efficient expression at 37 °C, non-toxicity and insensitivity to the used environment. Any criteria for more complex applications such as photo-activation were ignored here.

3.4. TESTED PROTEINS AND EXPERIMENTAL DESIGN

We chose several modern fluorescent proteins based on literature search [6, 72, 99, 100, 107] to test how they score on the above criteria: mCerulean, Cerulean2.0, GFPmut2, mVenus, mCherry, mKate2 (see Table 3.1). Since in some projects we planned to image two different fluorescent proteins simultaneously, using only the in our lab well characterized and established GFPmut2 would not be sufficient. Therefore, we included red & green and cyan & yellow emitting reporters as potential pairs (the latter allowing for red as third reporter), see also Fig. 3.1. These combinations hardly cross-talk, thus fulfilling the last criterion of section 3.3. Proteins were either obtained as monomeric variant or made monomeric by introducing a A206K mutation [108], see also the second last criterion above. To test the requirements on maturation time, photostability and brightness each reporter gene was expressed in *E. coli* . We chromosomally inserted the reporter under control of a constitutive exogenous P_{N25} promoter at the *php* location, one color per strain. In the following sections the notation "GFP" shall always refer to "GFPmut2". Note that the yellow variant Citrine might have been interesting to test as well, though we will show in the following sections that mVenus already scores well on our experimental criteria.

We here briefly summarize the performed experiments and give more details in the sections below. All experiments were performed on the "standard" microscope (TE2000) equipped with the CMOS camera (Hamamatsu, see chapter 2.2). First, we obtained cross-correlations between expression and growth rate for all fluorophores. The reason for this specific test is that cross-correlations are our most frequently used observable and we aimed to exclude dependency of features on the reporter protein used. We thus tested whether the maximum correlation, its sign, time delays and shape of the cross-correlations were dependent on the specific fluorescent protein. Cross-correlations are also an indirect measure for maturation time because time delays suggest slow chromophore maturation.

Second, we measured maturation more directly for two example constructs. We grew

Fluorescent protein	Color (emission)	ASC number	Reference	Monomeric
mCerulean	cyan	ASC823	[81]	introduced
Cerulean2.0	cyan	ASC833	[107]	yes
GFPmut2	green	ASC835	[84]	introduced
mVenus	yellow	ASC825	[83]	introduced
mCherry	red	ASC824	[97]	yes
mKate2	(far-)red	ASC834	[99]	yes

Table 3.1: List of tested fluorescent proteins. Proteins for testing were selected based on literature search. GFP derived proteins were made monomeric in our lab by introducing a A206K mutation [108]. Note that in some critical fusions even "monomeric" proteins can aggregate [71]. ASC number refers to the Amolf strain collection index. We named the codon optimized Cerulean for *E. coli* from the Elowitz lab "Cerulean2.0". They also codon-optimized Venus which was not tested here but could be considered if the standard Venus would cause issues in a construct.

E. coli cells in a flow cell and blocked translation with an antibiotic [88, 109]. The remaining increase of fluorescence is due to maturation and can be used to extract the maturation rate.

Third, to test photostability we blocked translation and waited for proteins to mature. Then we bleached the cells with frequent illumination to obtain a characteristic bleaching rate. One needs to note that the obtained bleaching rates are specific to the type of lamp and filters used (see also below).

Several of these measurements automatically provided data on perceived brightness of each fluorescent protein, that is, its signal intensity above the camera noise background.

3.5. CROSS-CORRELATIONS

3.5.1. EXPERIMENT, ANALYSIS AND WHAT CAN BE TESTED

We started the comparison by obtaining cross-correlations between fluorescence and growth rate for all reporter proteins. The aim was to determine whether specific proteins could alter sign, strength or shape of the correlation. The experiment was performed as described in chapter 2 and cells were grown on M9 + 0.2mM uracil + 0.1% maltose. Multiple microcolonies per experiment and fluorescent protein were analyzed to increase statistical significance. The aim was to perform also repeat experiments on a different day but this was only possible for a subset of colors due to time constraints. We investigated the cross-correlation between protein production and growth as well as the cross-correlation between protein concentration and growth. The following questions could be tested:

- Are the cross-correlations (sign/shape) independent of the reporter? This is important to compare experiments which used different fluorophores as label (e.g. ribosomal rRNA and r-protein in chapter 7).
- Is the cross-correlation function for one protein reproducible in different experiments?
- Does maturation seem to happen fast? GFPmut2 is a fast maturing protein (≈5min) which we will show directly in section 3.6. Therefore, we consider its timing in cross-correlations relative to growth rate as "correct". If other colors show delays relative to GFP, it suggests slow maturation.

We used the correlations obtained for GFPmut2 as a reference for the shape: GFPmut2 matures fast (section 3.6) and thus does not introduce delays. Further, the correlations were typically reproducible and changed as expected upon modifying the environment [56] (chapter 6).

3.5.2. RESULTS

Fig. 3.2 and Fig. 3.3 show the cross-correlations between protein production and growth rate ($R_{p\mu}$) resp. protein concentration and growth rate ($R_{E\mu}$) for all tested fluorophores. We first focus on correlations between production rate and growth (Fig. 3.2). The shapes of the correlation curves for both Ceruleans, GFPmut2 and mVenus were quite similar:

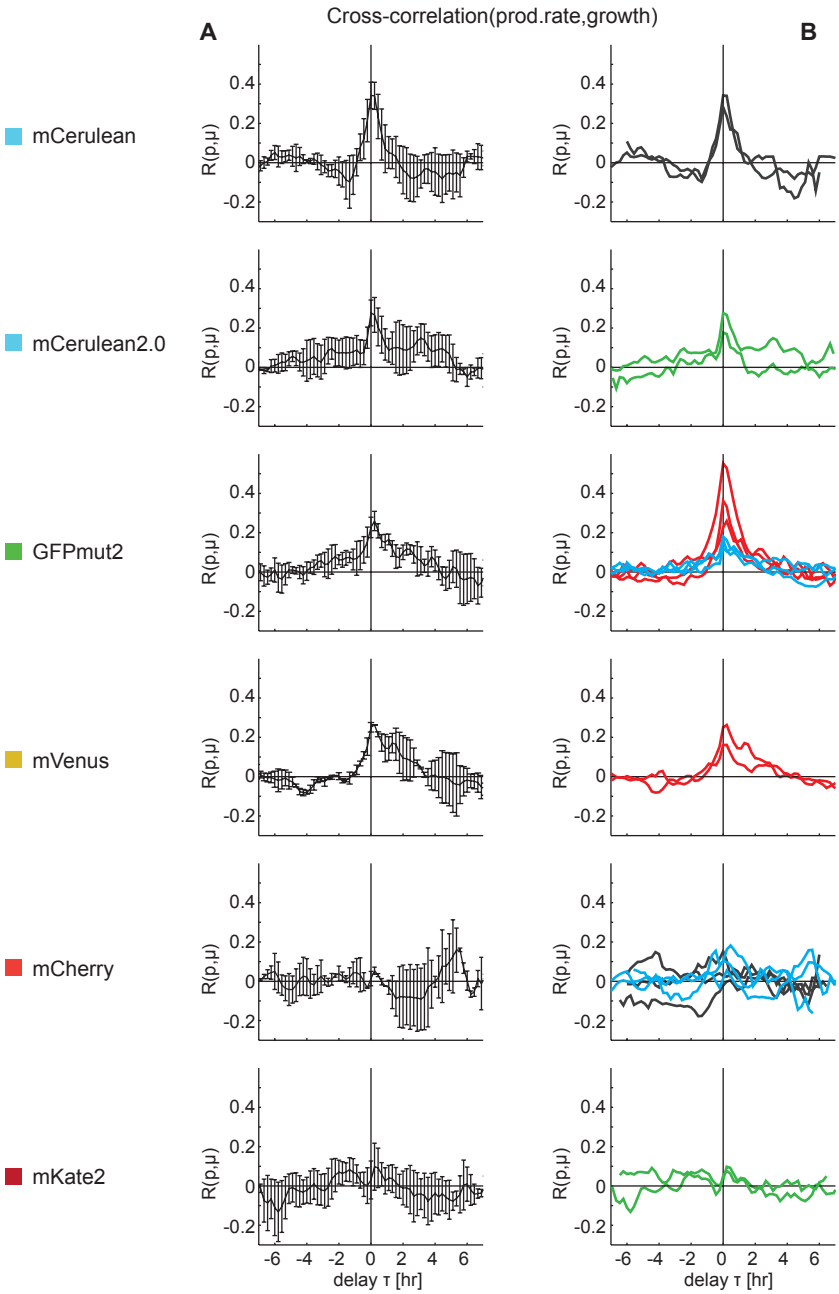


Figure 3.2: Cross-correlations $R_{p\mu}$ between production rate and growth for all tested fluorescent proteins. The colored squares indicate the spectral range. (**A**) Cross-correlation for one example dataset. Error bars are standard deviations obtained from four branch groups. (**B**) Cross-correlations for all measured datasets. Experiments were performed on four different days in total and each color corresponds to one day. The very large peak in one GFP correlation is likely an artifact of a slowly growing subpopulation. Growth rates were ≈ 0.85dbl/hr.

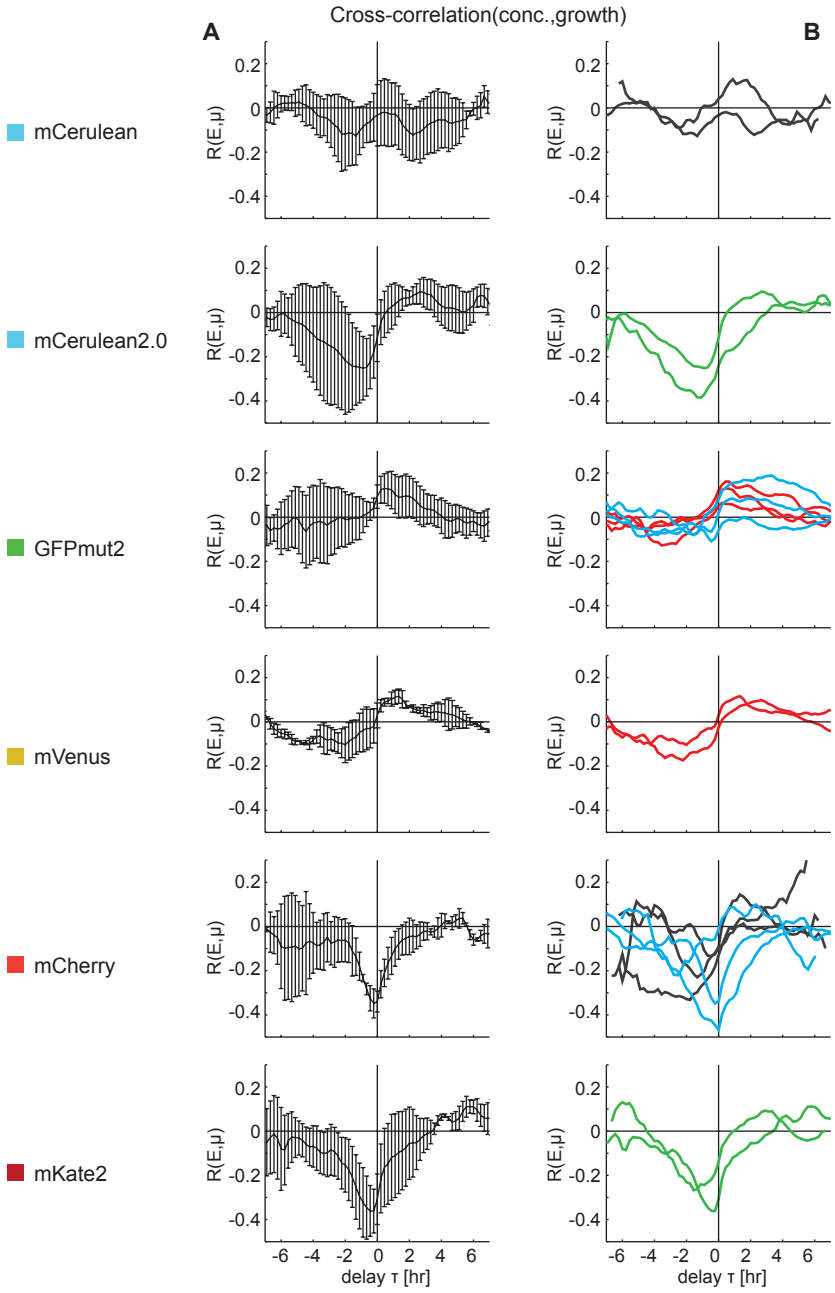


Figure 3.3: Cross-correlations $R_{E\mu}$ between concentration and growth for all tested fluorescent proteins. (**A**) Cross-correlation for one example dataset. Error bars are standard deviations obtained from four branch groups. (**B**) Cross-correlations for all measured datasets. Datasets (incl. the data from (**A**)) and color-code are identical to Fig. 3.2.

$R_{p\mu}$ peaked at zero time delay (τ) with a correlation coefficient of ≈0.2-0.3 and then decayed to zero within around two generations (≈2 hrs). GFPmut2 and mVenus might even show a small positive delay in the example datasets (Fig. 3.2A) but this shift is likely within the range of day-to-day variability (these two panels in (A) were taken on the same day). Contrary to that, $R_{p\mu}$ of mCherry and mKate2 were very different: The cross-correlations were very small and without well defined peak. Some of the mCherry cross-correlations appeared to peak at a negative time delay (one blue and one black line in Fig. 3.2B).

The concentration correlations $R_{E\mu}$ displayed more variability and were harder to interpret, but generally supported the conclusions obtained from $R_{p\mu}$ (Fig. 3.3): mCerulean, GFPmut2 and mVenus were all similar and their concentration was correlated very weakly with growth. The correlation was slightly negative (≈-0.1) at $\tau < 0$ and for an unknown reason slightly positive for $\tau > 0$. Again, and consistent with their very small $R_{p\mu}$ correlation, mCherry and mKate2 were different and displayed a very strong dilution mode. We note that also mCerulean2.0 showed a stronger dilution mode of which the reason is not clear and the experiment should be repeated to test whether the effect is real. Summarizing the results so far, we conclude that mCerulean, GFPmut2 and mVenus (potentially also mCerulean2.0) produced the same cross-correlations.

We next focused on the second point in the above list which is reproducibility of cross-correlations. To this end, we compared data from different microcolonies from the same experiment (i.e. from one day) to experiments from different days. The cross-correlations obtained for a specific fluorescent protein during one experiment were typically very similar for different microcolonies (Fig. 3.2B and 3.3B, lines of same color), with the only exception potentially mCherry. This similarity supported for example the above conclusion that $R_{p\mu}$ of mCerulean and mCherry, which were measured on the same day (Fig. 3.2B, black lines), were indeed different. The variation between results from different days was larger: While the broad features were conserved, the maximum correlation could change ($R_{p\mu}$ of GFPmut2 in Fig. 3.2B) as could the time lag ($R_{E\mu}$ of mCherry in Fig. 3.3B). Therefore, when a precise comparison of more detailed features is important it is advisable to measure the relevant strains simultaneously in one experiment.

The last point we checked was whether time lags ($\tau < 0$) of the peak in $R_{p\mu}$ hinted towards slow maturation times (Fig. 3.2). Both Ceruleans, GFPmut2 and mVenus were symmetric or even slightly biased towards $\tau > 0$. This suggests that they mature fast compared to our time resolution (≈15min between data points in Fig. 3.2) and are suitable to detect biological delays of that order and larger. Unfortunately, an analog interpretation of mCherry and mKate2 was not possible due to the absence of a correlation peak. This was surprising for mCherry because for different constructs (e.g. L31 and Icd fusion) it had regularly produced a spurious delay, which can also be seen in work of the Elowitz group (Fig. 4A,B in [107]). A reason might be the sensitivity of such delayed correlations: If production and growth are only weakly correlated (as for P_{N25}), then the delayed correlation should be even weaker because any other random fluctuation which occurs during the delay will decrease the correlation. If randomization is strong enough it could be that no delayed correlation is left and that $R_{p\mu}$ is approximately zero. Summarized, the $R_{p\mu}$ cross-correlations could hint towards a slower maturation time of mCherry and

mKate2, but the results are not very clear.

3.6. MATURATION TIME EXPERIMENTS

3.6.1. EXPERIMENT, ANALYSIS AND RESULTS

We directly determined maturation times for two fluorescent proteins, GFPmut2 and mCherry [56]. Cells were grown in a flow cell [44] on minimal medium containing 0.1% lactulose and 200 μM IPTG. We here describe the setup briefly, for further details see [44]: Similar to the static gel pad experiments, cells were applied onto a glass slide and covered with a, now thinner, polyacrylamide membrane. Instead of a microscope slide, a PDMS with an open channel on the bottom side was used to close the sample on top. The channel had two exits (punched holes) at the beginning and end. One exist was connected to a medium reservoir (filled syringe) via tubing, the other was connected to a waste bin. A pump flushed medium at constant speed from the reservoir through the channel. From the channel it diffused through the membrane to reach the cells. A valve was located before the PDMS and allowed to switch between different medium reservoirs.

When cells had grown into microcolonies of size 10-20, the medium was switched and a high concentration of the translationally inhibiting antibiotic chloramphenicol (≈100x minimal inhibitory concentration MIC) was added. Such a high antibiotic concentration immediately blocked protein production and growth, once cells were exposed. The remaining increase of total fluorescence is due to maturation of the already translated but still immature fluorescent proteins (Fig. 3.4). Following [109], we fitted the maturation curve with the function

$$M(t) = M_0 + I_0\left(1 - e^{-t/\tau_{mat}}\right) \quad . \tag{3.1}$$

Here, $M(t)$ is the amount of mature fluorescent proteins at time t, M_0 and I_0 are the amounts of mature and immature proteins at the time of antibiotics addition, t is the time since addition of antibiotics and τ_{mat} is the maturation time. The amount of mature proteins is measured in arbitrary fluorescent units. In practice, frequent illumination, which is necessary for high time resolution, led to bleaching of fluorescent proteins (see also section 3.7). Thus, the increase in fluorescence due to maturation is overlaid with a decrease due to bleaching. We therefore fitted an exponential decay function to the bleaching-dominated data at large times after the switch, extracted a bleaching rate and corrected fluorescence values with that rate. The fluorescence trace of each cell and its progeny was then fitted with the above maturation function. Averaging these rates resulted in the fitted maturation times 4.7 ± 0.5 min for GFPmut2 and 14.0 ± 2.4 min for mCherry (mean ± standard deviation). These numbers are in agreement with previously published data [97, 102, 110].

3.6.2. POSSIBLE ISSUES

There are some uncertainties concerning the determination of maturation times which we discuss here: First, after a medium switch the cells were not immediately exposed to antibiotics because the new medium first had to reach the PDMS channel (≈2 min),

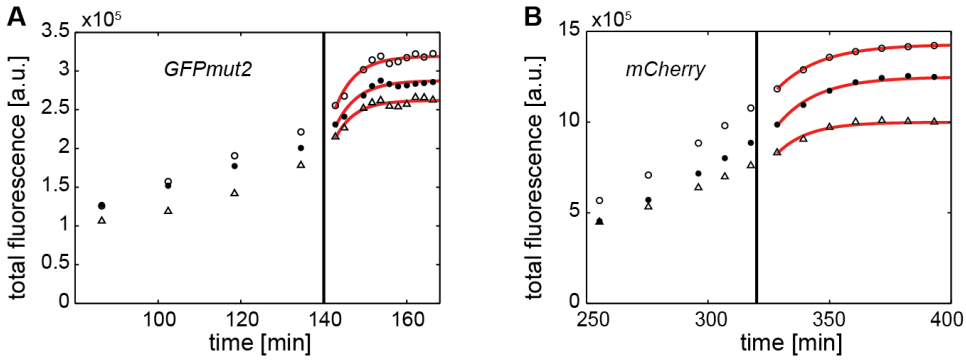


Figure 3.4: Determination of maturation times of GFPmut2 (A) and mCherry (B) . Cells were grown in a flow cell in minimal medium and a high concentration of antibiotics was added at the time indicated (black line). Protein production stopped and remaining increase of fluorescence is due to maturation. The bleaching-corrected curves were fitted by a saturating exponential, see main text. Each figure displays the total fluorescence time traces of three representative cells and their progeny.

and diffuse through the membrane (≈5 min in the absence of air bubbles). The onset time of antibiotics therefore needed to be estimated and we defined it as the time of growth cessation. However, excluding or including some data points at the edge of the fit window led to some variation in fit time and especially could increase the maturation time determined for mCherry. A decrease in that uncertainty could probably be achieved by increasing switching speed between media. This could be done by using flow cells with very thin membranes [111] to reduce diffusion time or by avoiding the membrane and gluing cells with polylysine [88] to the surface. Whether the slightly toxic polylysine interferes with protein dynamics would need to be tested.

Second, the maturation kinetics of mCherry are not perfectly understood but involve two oxidation steps [94]. Therefore, a sigmoidal fit [112] might be a better choice compared to an exponential fit. Our data does not allow to draw conclusions on the precise shape of the maturation curve.

Third, maturation depends on oxygen. We and other groups [112] have indications that differences in cellular metabolism, caused by changing growth medium or the used strain, may affect maturation rates (see also chapter 7.3.1 and Fig. 7.9). Presumably, internal oxygen concentration is altered, which then affects the rate of oxidation. It also cannot be excluded that external oxygen within the setup decreased over time. However, we aimed to keep that concentration high by providing a large reservoir in the sample chamber and using few cells. Taking all these points into account, we think that the determined values of mCherry are not quite clear because of sensitivity to analysis methods and might be larger. mCherry might therefore only be suitable for slow-growth experiments. The potential dependence of maturation rate on the environment adds additional ambiguity. Contrary to that, GFPmut2 matured clearly fast. Therefore, we considered it suitable for our time-lapse experiments and used its cross-correlation functions as a reference.

3.7. PHOTOSTABILITY EXPERIMENTS

3.7.1. EXPERIMENT AND ANALYSIS

To measure photostability, we pregrew cells as described in chapter 2 and used M9 + 0.2mM uracil + 0.1% maltose as growth medium. On the next day, growth and protein production was stopped by adding 1mg/ml of the translational inhibitor spectinomycin (100-500x MIC). The acryl gel was immersed into medium with added antibiotics. Before starting microscopy we waited for two hours to allow all fluorescent proteins to mature. Then, the sample chamber was assembled as described in chapter 2, but cellular density was chosen higher (at least OD=0.2 or further concentrated by spinning down). We searched for positions with ca 10-50 cells in the field of view, then acquired a phase contrast and fluorescence movie at maximum frequency (ca. 300 images/hour). Fluorescent images were taken every loop with illumination time 200ms and refocusing was reduced to every fifth loop to increase image frequency. After roughly 300 images fluorescent proteins were sufficiently bleached and we restarted movie acquisition for a new position.

The cells were segmented and tracked and time dependent total fluorescence was determined for each cell. We confirmed that choosing total fluorescence (sum of all pixels) or fluorescence concentration did not have any influence on determined values. As result we obtained for every cell a time-dependent total fluorescence curve $F(t_{cum})$ (Fig. 3.5A, fluorescent protein is mKate2). Here, t_{cum} is the summed (cumulative) illumination time experienced by a cell since the beginning of the experiment. We found that the decay of individual fluorescence traces appeared deterministic and exponential but the decay rate seemed to vary widely for the same protein. The reason is that the field of view is illuminated unevenly (Fig. 3.5B). Therefore, cells in the center (light gray trajectories in Fig. 3.5A) were illuminated stronger and as consequence bleached faster than cells located at the corners (dark gray trajectories). We corrected for the uneven intensity by introducing effective illumination times which were proportional to the locally applied light intensity. Since a microcolony in a growth experiment is typically located in the center quadrant of the chip (Fig. 3.5B white square), we normalized to the average light intensity in that quadrant as "standard intensity". Then, bleaching curves of individual cells nicely collapsed (Fig. 3.5C), which could also be expected for a photochemical (non-biological) process.

We fitted an exponential decay function

$$F(t_{cum}) = F_0 \cdot 2^{-t_{cum}/T_{1/2}} \tag{3.2}$$

to each trajectory. Here, $T_{1/2}$ is the characteristic time after which half of the fluorescent molecules have been bleached and F_0 accounts for the initial fluorescence value. Using the fitted $T_{1/2}$ of all cell trajectories, we then calculated mean and standard deviation of half-times (red curve in Fig. 3.5C).

3.7.2. RESULTS

Applying this analysis to all tested fluorescent proteins we found that bleaching characteristics were always well described by single exponential functions, which is a priori not necessarily the case [72] (Fig. 3.5D). A double-exponential could not be excluded

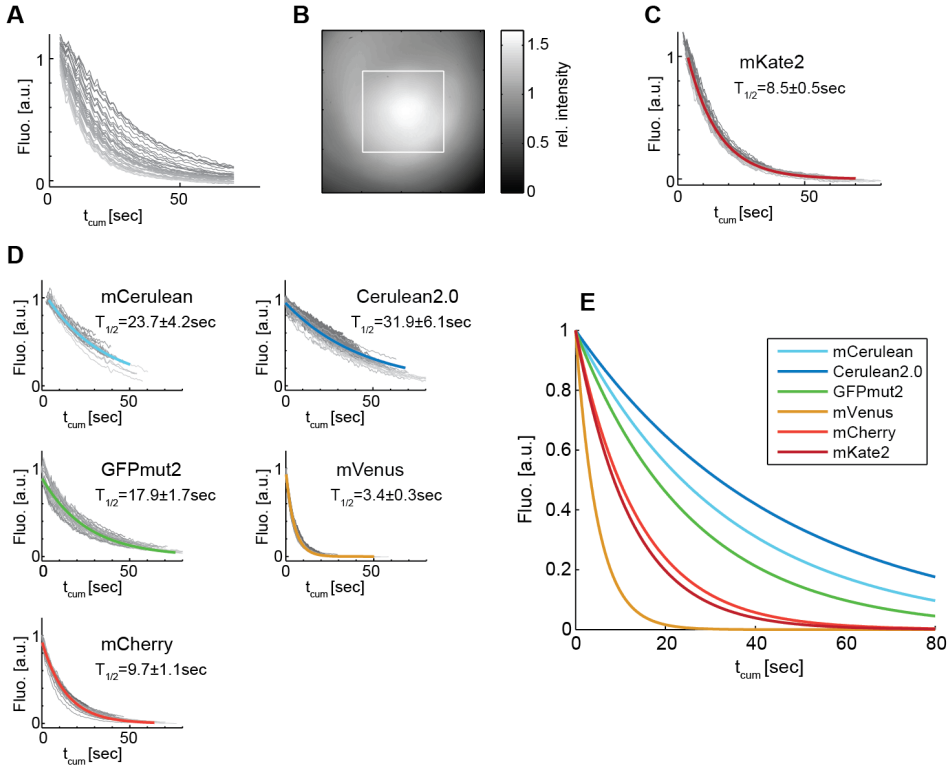


Figure 3.5: Bleaching curves of fluorescent proteins. (**A**) Total cellular fluorescence as a function of integrated illumination time t_{cum} for mKate2. Each trace is one cell and initial fluorescence values are normalized to one. The decays appear single-exponential but with different time constants. Brighter colored trajectories are from cells located more centrally in the field of view and decay faster. (**B**) Distribution of illumination intensity on the camera chip in the red channel. The illumination is uneven and strongest in the center. We normalized illumination by introducing effective illumination times which correspond to a theoretical illumination with a center-quadrant intensity (white square). (**C**) After normalization, the bleaching curves nicely collapse onto one curve. The red curve is obtained by averaging fitted decay rate and initial value of all cells. (**D**) Same as (**C**) but for all other proteins tested. All curves can be fitted with single-exponentials. The collapse of data for green and cyan proteins can be improved by normalizing not with the first (noisy) fluorescence value but a later value. The large standard deviations for cyan variants could be due to insufficient correction for uneven illumination. (**E**) Average bleaching curves of all tested fluorophores. $T_{1/2}$ of the Cerulean versions are not significantly different (Table 3.2). Data sizes are given in caption of Table 3.2.

for e.g. GFP and mCherry traces. We found that photostability varied widely for different fluorescent proteins [72] and $T_{1/2}$ ranged from ca. 3 seconds for mVenus to >20 seconds for both Cerulean variants (Fig. 3.5D,E and Table 3.2). To estimate the impact of bleaching in time-lapse experiments, we calculated for typical experimental settings the fraction of bleached proteins within one cell-cycle (6 images with 100ms il-

Fluorescent protein	Half-time $T_{1/2}$ [sec] (mean ± stddev)	Fraction bleached per cell-cycle (6·100ms)	Fraction bleached per cell-cycle (adjusted illum. times)	Relative signal intensity (identical illum. times)
mCerulean	23.7 ± 4.2	1.7%	1.7% (100ms)	1
Cerulean2.0	31.9 ± 6.1	1.3%	1.3% (100ms)	1
GFPmut2	17.9 ± 1.7	2.3%	0.7% (28ms)	3.5
mVenus	3.4 ± 0.3	11.6%	1.9% (15ms)	6.5
mCherry	9.7 ± 1.1	4.2%	0.9% (21ms)	4.9
mKate2	8.5 ± 0.5	4.8%	1.1% (22ms)	4.6

Table 3.2: Bleaching rates of fluorescent proteins. $T_{1/2}$ is the time after which fluorescence intensity has decreased to 1/2. Mean and standard deviation were obtained from the distribution of fitted individual exponential decay functions to each cell. The fraction of proteins (present at cell birth) which are bleached throughout one cell cycle was estimated in two ways: First, acquisition of six images with 100ms illumination time. Second, acquisition of six images but adjustment of illumination times such that each fluorophore has the same signal strength above background (see right column). Illumination for mCerulean was kept constant and the illumination times for other colors were adjusted accordingly. Then, the unstable but bright mVenus performs similarly well as other colors. Right column: Measured signal intensity above background for identical illumination times, normalized to mCerulean intensity. Number of measured cells: mCerulean: n=22, Cerulean2.0: n=56, GFPmut2: n=33, mVenus: n=28, mCherry: n=35, mKate2: n=46. We note that the GFP filter transmits less light, resulting in an apparent larger photostability. Using the more modern 49002 filter (Chroma) would increase bleaching rate of GFP by very roughly 50%. That effect would be overcompensated by shorter required illumination times, resulting in even smaller bleaching for a given signal-to-noise ratio.

lumination each, Table 3.2). With these settings, the fraction of initially present proteins that would be bleached after one cell cycle ranged from <2% (both Cerulean versions) to >10% (mVenus). However, these values are biased because the much brighter mVenus needs to be illuminated shorter than e.g. mCerulean to obtain the same signal to noise ratio. Therefore, we next adjusted example illumination times such that all proteins would be imaged with approximately the same signal-to-noise ratio as mCerulean. Then, the different proteins were bleached by much more similar factors (each <2%), and GFP was bleached the least. Obviously, increasing illumination times would increase the fraction of bleached proteins. To assess the relevance of bleaching in a time-lapse experiment, the bleached fraction needs to be compared to the 100% increase of proteins within one cell cycle (doubling of size). Therefore, 2% is a small loss and all fluorescent proteins can be considered suitable for time-lapse imaging. However, for longer illumination times, which are needed for low-copy proteins, bleaching can become substantial and should be corrected for during analysis.

Importantly, the determined bleaching rates are to some extent specific to the setup used, especially the light source and filters. Different Xenon lamps will probably produce similar bleaching characteristics. But a different source than a Xenon lamp, for example a higher-intensity laser illumination, can result in substantially different bleaching

curves [72]. The choice of filter determines the excitation intensity and how much of the emitted light is detected. A better filter (higher transmission of desired wavelengths) increases the bleaching rate per second but it decreases the required illumination time for a given signal-to-noise ratio. The net effect is a reduction in photobleaching. While our cyan, yellow and red filters belong to the best filters available, the gfp filter could be upgraded to decrease bleaching of GFP proteins even further, if necessary.

3.7.3. A NOTE ON FLUOROPHORE BRIGHTNESS

We used the signal intensities from the beginning of the bleaching experiment to compare apparent brightness of different proteins in our setup. For medium-high expression (e.g. promoter P_{N25} , induced P_{lac}) all tested colors were by far sufficiently bright and brightness is then not a relevant selection criterion. For low expression, however, brighter proteins are advantageous because illumination times can be kept shorter. The signal intensity should then also be compared to cellular autofluorescence: Autofluorescence is highest at short wavelengths (cyan) and almost not detectable at long wavelengths (red).

We here only focus on the measured signal intensity on the camera chip, which is the final result of many contributing factors: intrinsic fluorophore brightness, the excitation spectrum and intensity of the lamp, filter spectra, camera sensitivity in the specific wavelength range and expression efficiency of a specific protein. Here, intrinsic fluorophore brightness is defined as the product of extinction coefficient and quantum yield and describes the efficiency of a protein to absorb and re-emit a photon. As summarized in [72], mVenus is the intrinsically brightest protein, usually followed by green, cyan and red proteins.

Assuming similar expression efficiency we found that mVenus yielded the highest signal intensity in our setup (Table 3.2, right column). It was followed by mKate2 and mCherry, then GFPmut2 and finally both similarly bright Cerulean versions (however other experiments indicated that Cerulean2.0 might be brighter). As mentioned, detected GFP brightness could be increased by exchanging the filter. The red emitting proteins were quite bright and could be suitable for low expression due to the almost absence of autofluorescence in the red range. However, their slow maturation (see above) renders them less suitable for some time-lapse experiments. We note that expression efficiency can change for different reporters and constructs. Therefore, these recommendations should provide a good start for designing constructs but in some cases brightness might need to be tested for the individual case.

3.8. CONCLUSIONS

To conclude, we found that the proteins mVenus, GFPmut2 and mCerulean (and likely Cerulean2.0) produced similar and reproducible cross-correlations between expression and growth rate. Cross-correlations of the red emitting proteins mCherry and mKate2 appeared less well defined and less reproducible and therefore these proteins were less suitable in the tested conditions. Photostability of all measured proteins was sufficiently high and rather comparable when adjusting for relative signal brightness. But if photostability needs to be optimized, for example because of long illumination times, then

GFPmut2 was the most stable protein for a fixed signal intensity. When high brightness for a fixed illumination time is important mVenus was the best choice. The red proteins are both rather photostable and bright (plus absence of red autofluorescence) and are therefore good choices when dynamics are of less importance.

GFPmut2 is probably the most suitable protein for a single-label strain: It produces good cross-correlations, matures fast (directly measured), has sufficient brightness and high photostability. However, in a dual-color strain it could only be combined with red emitting proteins, which appear less suitable. Before the beginning of this characterization study, we had measured GFPmut2 & mCherry dual-label strains in some experiments (chapter 6). We found that in slow growth conditions the results for cross-correlations were reasonable. For double-label experiments, the best choice appears to be a combination of mVenus and mCerulean or Cerulean2.0. They are similarly good as GFPmut2, except for lower photostability of mVenus (which is likely of little importance) and dimmer fluorescence of Cerulean. At the beginning of designing a new reporter strain one should consider whether a second fluorescent protein might need to be added later. Then, it is always the best choice to start with cloning mVenus (or Cerulean) instead of GFPmut2. One note of caution needs to be added: Cerulean is excited at short wave lengths and while we observed constant growth rate at short illumination times (≈70ms), some cells grew slower at the end of the experiment for longer illumination (≈200ms). Compared to other fluorophores, Cerulean also requires relatively long illumination times (Table 3.2) and cells display significant autofluorescence in this spectral range. Therefore, it may be advisable to use mVenus for low-copy proteins and, in a dual-label strain, Cerulean for the more abundant proteins (for which shorter illumination is sufficient). To close, the results presented here should be seen as general guidelines for selecting fluorescent proteins. They cannot make testing fusion functionality or growth burden on the individual construct basis superfluous.

The experiments on cross-correlations in this chapter were performed jointly by Martijn Wehrens and the author.

4

INTERDEPENDENCE OF GENE EXPRESSION AND GROWTH RATE FLUCTUATIONS IN BACTERIA

Fluctuations in gene expression and growth rate of single cells can have important consequences for cellular function and fitness. While average gene expression and growth rates obtained in different environments are coupled, it is not known how their fluctuations (noise) are related, nor whether different environments play a role in that relation. To study this question, we use time-lapse microscopy and measure growth and expression fluctuations of E. coli cells at various mean growth rates, two different expression levels and with different promoters. We find that while growth and expression fluctuations are modestly correlated ($0 \leq R \lesssim 0.5$), the intensities of growth and expression noise are strongly correlated ($R > 0.9$) and scale linearly. Noise intensities typically decrease with increasing average growth rate. Datasets with increased noise levels, that deviated from this trend, still displayed the same scaling between noise intensities. By measuring expression of two protein reporters under identical control, we show that global (extrinsic) noise increases around six times stronger with increasing growth noise than local (intrinsic) noise, but the dependence of the latter is not negligible. We further develop a linear noise model and show that fluctuations in a cell can be described by a set of independent noise sources, whose intensity is, remarkably, set jointly by one global parameter. This model can predict changes occurring upon a decrease in expression level and explain the different expression noise observed in transiently faster or slower growing cells. These results are a step towards a better understanding of the structure of cellular noise.

4.1. INTRODUCTION

On the single-cell level, variability in genetically identical cells exposed to the same environment is ubiquitous: the expression of genes displays random fluctuations in time (noise) [5–8, 113], which has important consequences for signal transmission fidelity and cellular phenotype [9, 11, 12, 15, 26, 27, 114]. The growth rate of cells also fluctuates in a constant environment, which may have direct consequences for cellular fitness [32, 56, 115]. Average gene expression in a population of cells was found to depend on its average growth rate, when measured in different steady states [1, 116–118]. Since cellular growth rate reflects the production of biomass, of which a major fraction are proteins, also fluctuations in expression and growth rate should not be separate phenomena, but must be inherently connected. However, how they are related remains incompletely understood.

Gene expression noise is thought to be the result of many effectors such as stochasticity of biochemical reactions, fluctuations in concentration of RNA polymerase, ribosomes or transcription factors, as well as progression through the cell cycle [5, 7, 16, 119, 120]. Expression noise was found to be inversely related to average expression level [6, 21] and, in yeast, to be negatively correlated with average growth rate [121]. For the latter, a major cause was shown to be a changing distribution in a population of cells into different cell-cycle stages [121]. Growth fluctuations may have many origins, among which are fluctuations in concentration of limiting enzymes [56]. It was shown that these fluctuations can propagate forward through metabolic pathways and cause growth fluctuations, and that growth fluctuations can in turn propagate back again to affect enzyme concentration. Growth noise may be dependent on the environment as well [56, 115]. In this study, we aim to address the question how the fluctuations in these two important cellular outputs, expression of proteins and volume growth, are related and how changes in one may be connected to changes in the other.

To this end, we measure growth end expression fluctuations of hundreds of *E. coli* cells using time-lapse microscopy. By growing cells on various carbon sources, we explore a wide range of average growth rates, which we find to affect noise intensities. Additionally, we measure gene expression from different promoters and at two induction levels. We find that the intensities of gene expression and growth noise are strongly correlated, they scale linearly and typically depend inversely on the average growth rate. In contrast to that, fluctuating expression and growth time traces are only modestly correlated. The correlation coefficient of the fluctuations is found to be a unique and non-linear function of the intensity of growth noise. We then use a dual-reporter strain that expresses two protein reporters under control of identical promoters to separate global (extrinsic) and local (intrinsic) expression fluctuations. Extrinsic noise is found to increase around six times stronger with increasing growth noise than intrinsic noise, but the dependence of the latter is not negligible. We further show that a linear noise model can describe our results and predict the effects of changing average expression level. A central result of modeling is that the magnitudes of different sources of cellular noise vary jointly and can thus be described by using only one global noise intensity parameter. We further find that in a population of cells, bacteria that transiently grow faster due to stochastic variations, display lower expression noise levels. Finally, we show that growth fluctuations are mostly only coupled to extrinsic expression fluctuations but

hardly to intrinsic fluctuations. These results contribute to a better understanding of systems-level properties of cellular noise, including inherent constraints and relations between different sources of noise. In the long term, this may contribute to better understanding and predicting bacterial behavior.

4.2. RESULTS

4.2.1. EXPRESSION AND GROWTH FLUCTUATIONS DEPEND ON THE GROWTH MEDIUM.

To study growth and expression fluctuations, we acquired time-lapse movies of *E. coli* cells as they grew into microcolonies of several hundred cells (Methods 4.4). We supplied one of various carbon sources to explore different growth rates. By using automated image analysis and cell length measurement, we determined instantaneous growth rate $\mu(t)$ of individual cells at sub-cell-cycle resolution. We measured the rate of gene expression $Y(t)$ with a chromosomally encoded fluorescent protein (YFP) under control of a P_{lac} promoter (Fig. 4.1A) [7]. Expression rate fluctuates quasi-periodically due to gene duplication during cell-cycle progression [119], and growth rate typically displays quasi-periodic fluctuations as well. We subtracted these cell-cycle trends in both protein production rate and growth rate (chapter 2.3.4). Explicit notation of time dependence will be omitted and we will write μ and Y. We found that both protein expression and growth rate fluctuated over time (Fig. 4.1B,C) [56]. To quantify these fluctuations and their interrelation, we determined their noise intensity η_μ^2 and η_Y^2 and correlation $R(Y,\mu)$. Noise intensity was defined as variance divided by the squared mean (σ^2/μ^2, which is the squared coefficient of variation CV^2). Fluctuations in protein expression and growth rate were positively correlated (Fig. 4.1D and 4.6, ref. [56]) and therefore temporarily faster growing cells produced on average more proteins (Fig. 4.1D and 4.7).

To understand what factors determine the relation between growth and expression fluctuations, we aimed to alter the cellular noise by growing cells under different conditions. We used different carbon sources that resulted in growth rates from 0.2 to 1.8 dbl/hr, as well as two different induction conditions resulting in either high or low expression levels. Additionally, we measured expression fluctuations from different, constitutive promoters (P_λ , P_{N25}) (see Methods 4.4 and Fig. 4.1E). We indeed found that the noise properties were significantly affected. The intensity of growth noise varied largely, ranging from around 0.01 to 0.09, and expression noise varied from 0.08 to 1.6 and was inversely dependent on average expression rate [6, 8, 19, 21]. The correlation $R(Y,\mu)$ was also affected, and ranged from 0 to around 0.5 (Table 4.1). Finally, we surmised that the noise intensity could be affected by using high fluorescence illumination exposure times, as this is known to result in oxidative stress and reduced mean growth rates [122]. Indeed we found that the noise intensities increased under these conditions (Fig. 4.1E). Thus, both the expression and growth noise depended strongly on the growth conditions.

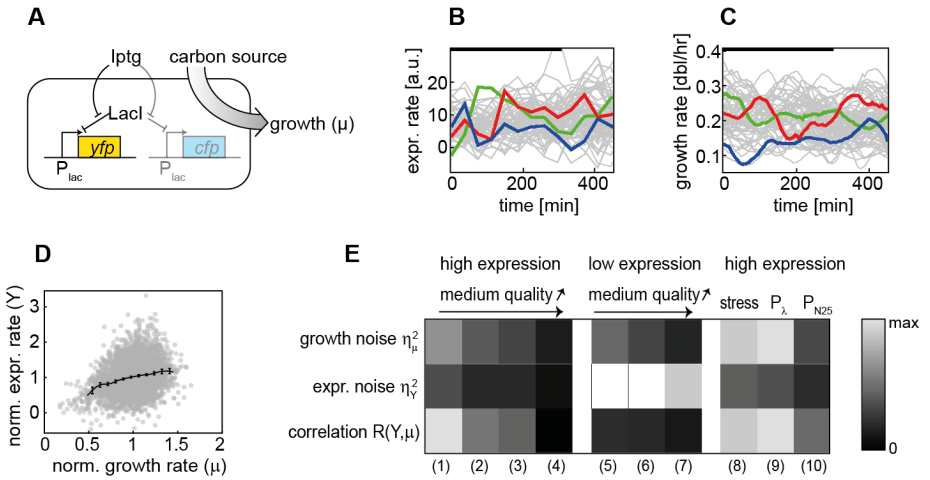


Figure 4.1: Fluctuating growth rate and protein production rate, as well as their noise intensity and correlation. (**A**) Reporter strain and adjustable parameters. Expression of the fluorescent reporter YFP is controlled by a P_{lac} promoter. A second reporter (CFP) under control of an identical copy of the P_{lac} promoter is needed for later analysis. LacI proteins repress expression but can be deactivated by addition of the gratuitous inducer Iptg. The environment and thus growth rate is changed by altering the supplied carbon source. (**B**) Time traces of fluctuating YFP expression rate and (**C**) growth rate for growth on acetate in the presence of Iptg (μ=0.2dbl/hr, column (1) in (**E**)). Three example lineages are highlighted. The black bar denotes the average interdivision time. (**D**) Correlation of protein production rate and growth rate for growth on maltose in the presence of Iptg (μ=0.6dbl/hr, R=0.28, column (2) in (**E**)). Variables are normalized by their mean. The line denotes the conditional expectation $\langle Y|\mu\rangle$, that is the average production rate for a given growth rate, and was obtained with kernel density estimate (Methods section 4.4.2). Error bars are standard error of the mean (SEM), displayed for a subset of data points, and were obtained by bootstrapping (Methods). (**E**) Noise intensity and correlation of expression and growth fluctuations for multiple tested environments. For coloring, η^2_μ and $R(Y,\mu)$ values were normalized by the maximum of each row. For η^2_Y, which varies over a large range, $0.6\cdot\max$ was used for normalization, therefore η^2_Y in (5) and (6) is saturated on the used color scale (displayed in white). Growth media: (1),(5) M9+0.18% acetate (μ≈0.2-0.3dbl/hr), (2),(6) M9+0.1%maltose (μ≈0.6dbl/hr), (3) M9+0.1%lactose (μ≈0.6dbl/hr), (4),(7) EZ defined rich + 0.2%glycerol (μ≈1.8dbl/hr). To reach high expression in (1),(2),(4), 200 μM Iptg was added (induction in (3) directly by lactose). For definition of high and low expression, see Methods 4.4. Additional experiments: (8) M9+0.1%maltose+200μM Iptg with high illumination intensity that leads to a decrease in growth rate and hence is toxic to the cell, (9) M9+0.1%maltose and strong, constitutive P_λ-promoter, (10) M9+0.1%lactose and constitutive P_{N25}-GFP (expression rate here also denoted with Y). Replicates are not displayed in this overview. For values, see Table 4.1.

4.2.2. EXPRESSION AND GROWTH NOISE INTENSITY ARE STRONGLY CORRELATED AND SCALE LINEARLY.

We found that η^2_μ and η^2_Y typically decreased with increasing average growth rate (Fig. 4.8) [56, 115, 121]. Thus, $\langle\mu\rangle$ was in general a fairly good predictor for expression and growth

noise. Some datasets, however, did not follow this trend (Fig. 4.1E, and Fig. 4.8 gray and dark blue symbols). Therefore, we next tested whether growth and expression noise might be more directly related to each other by plotting expression noise intensity against growth noise intensity (Fig. 4.2A). We first focused on experiments with high expression (induced P_{lac} and constitutive promoters; for classification, see Methods 4.4). Surprisingly, the expression and growth noise intensities collapsed onto one single line. Indeed, the data displayed a high correlation coefficient of 0.94. That means that expression noise and growth noise intensity determined each other. The data points from the constitutively expressed proteins also obeyed the scaling (Fig. 4.2A, square and diamond symbols), suggesting that the scaling is independent of the specific promoter used. This observation was notable for two reasons: First, one could reason that the scaling between the noise intensities is explained by the observed correlations between the growth and expression signals (Fig 4.1D, Fig. 4.6), which reflect the fact that faster growing cells also produce proteins faster. However, these correlations were small ($0 \leq R \lesssim 0.5$) in comparison to the ones observed here, which shows that they cannot directly explain the observed scaling. Second, the linearity of the relation is noteworthy. Many cellular components are required for growth and thus growth fluctuations may result from a complex combination of their fluctuations. This could affect the resulting noise, lead to buffering or enhancement and, for example, have led to a more complex nonlinear relation between growth and expression noise.

To better understand the origins of the scaling, we decomposed noise in protein expression further: Fluctuations in gene expression can be separated into two elementary categories, "extrinsic" and "intrinsic" noise [7, 19]. Extrinsic noise refers to fluctuations that affect multiple genes simultaneously, and are thought to arise from variation in the concentration of transcription factors, ribosomes, etc. Intrinsic noise refers to fluctuations that affect one particular gene, which are thought to arise from stochasticity of chemical reactions. Because these noise types are fundamentally different, we wondered whether they would relate differently to growth noise. Hence we measured expression C from a second, chromosomally encoded, fluorescent reporter gene (*cfp*), additionally to YFP expression (Fig. 4.1A). Both genes were under control of P_{lac} -promoters (or P_{λ} -promoters) and located equidistant to the origin of replication [7]. Then, joint fluctuations in protein expression rate determine extrinsic noise η^2_{extr} , while differences in fluctuations determine intrinsic noise η^2_{intr} . Mathematically, this can be written as [7, 19]:

$$\eta^2_{extr} = \frac{Cov(Y,C)}{\langle Y\rangle\langle C\rangle} \quad , \qquad \eta^2_{intr} = \frac{Var(Y-C)}{2\cdot\langle Y\rangle\langle C\rangle} \quad . \tag{4.1}$$

Angular brackets denote averaging over all data points (time points and cells). The sum of both noise terms equals total expression noise:

$$\eta^2_E = \eta^2_{extr} + \eta^2_{intr} \quad . \tag{4.2}$$

Here, E denotes expression rate (either Y or C). Consistently, the noise properties of the C signal were similar to those of the Y signal (Table 4.1).

We found that both extrinsic and intrinsic noise increased linearly with growth noise, with the extrinsic noise showing a strong and the intrinsic noise showing a weak dependence (Fig. 4.2B). The dependence on growth noise, as quantified by the slopes of linear

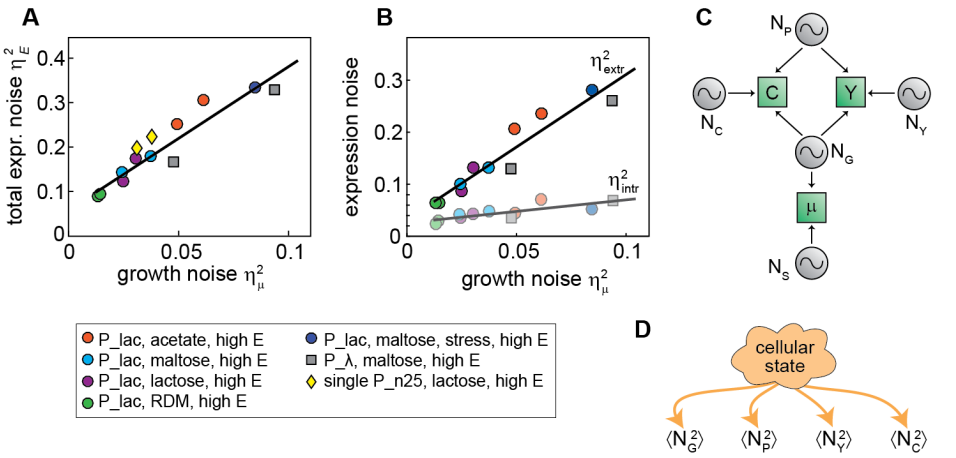


Figure 4.2: Linear scaling of noise intensities. (**A**) Total expression noise and (**B**) extrinsic resp. intrinsic noise in dependence of growth noise. Lines are linear fits of the model. "High E" in the legend refers to high expression level (extrinsic noise dominant). η^2_E is the average of the very similar η^2_Y and η^2_C . Growth of cells expressing YFP from P_λ promoters (gray squares) was variable and sensitive to preparation steps. This produced differences in noise between replicates, which could be used to explore a larger range of noise intensities. (**C**) Linear noise model. Specific and common noise sources (circles) act on the observables Y, C and μ (squares). N_P can not only include transcription factors but also components relevant for the expression of all proteins, such as amino acid or RNA polymerase concentration. Noise sources vary in intensity (except for N_S) and can additionally have a constant-intensity noise floor. Shared noise sources can transmit to connected observables with different efficiencies. For full model description, see appendix 4.5.1 and Fig. 4.9A. (**D**) Magnitude of different noise sources vary jointly and can thus be described by only one shared parameter.

fits, was around six times stronger for extrinsic noise than for intrinsic noise. The correlations were strong in both cases (0.96 and 0.81, respectively). This suggested that the intensity of extrinsic expression noise sources is strongly correlated with the intensity of sources of growth noise. The fact that the extrinsic noise depends more strongly on the growth noise than the intrinsic noise may be expected, as factors that cause fluctuations in growth rate should reflect a global property of the cellular state and thus should affect both genes. Consistently, the intrinsic noise depends only weakly. However, having a dependence is notable, as intrinsic noise is often considered to be purely local and hence not correlated with other cellular processes. It could suggest that local stochastic events do also depend on fluctuating concentrations of global factors, or that the two seemingly identical genes respond differently to global noise sources.

4.2.3. ONE GLOBAL PARAMETER SETS THE INTENSITY OF DIFFERENT CELLULAR NOISE SOURCES.

The linear relations between expression noise and growth noise suggested that the data could be explained by a model in which noise sources couple linearly to expression and growth. If so, such a model could also show which constraints the data imposes on such models. We thus developed a linear noise model that contained different independent sources of noise, which can either act on one observable only or on multiple observables simultaneously (Fig. 4.2C) [6, 27, 56]. For the latter, we distinguished between global fluctuations and fluctuations only affecting protein expression (which could for example be amino acid concentration or transcription factors). Fluctuations from sources that act on two or more observables cause correlations between them. For details on the model, see appendix 4.5.1.

A priori, the intensity of different noise sources may be either constant or depend on the environment, but together, they should reproduce the environment-dependence of growth and expression noise. To determine which sources vary and how they vary, we fit the model to the experimental data. As an example, consider the implications of a varying intensity of global noise, $\langle N_G^2 \rangle$. Then, the growth noise, as well as the covariance $Cov(Y,\mu)$, would change linearly with $\langle N_G^2 \rangle$. As consequence, the covariance should be a linearly increasing function of growth noise (see eq. 4.16). Indeed, the experimental $Cov(Y,\mu)$ plotted against η_μ^2 collapsed onto an increasing line (Fig. 4.10A), thus supporting the structure of the linear noise model, and the intensity-variation of the global noise source.

Using similar arguments, we found that the intensity of almost all noise sources changed with the environment. We only keep the growth-specific noise $\langle N_S^2 \rangle$ constant as it is not constrained by the data. Several noise sources possessed, additionally to the intensity-varying contribution, a non-zero noise floor of constant intensity. This suggested the existence of a minimum noise limit in cells, and was visible as vertical and horizontal shifts in the linear noise relations (Fig. 4.2A,B, 4.9B, appendix 4.5.1). Surprisingly, the variation in noise intensities is concerted, meaning that doubling the intensity (above the noise floor) of one noise source is accompanied by a doubling of intensity of all other noise sources (Fig. 4.2D). Therefore, the changes in all noise sources could be described by one parameter. This parameter quantifies the amount of "chaos" in the cell. For a detailed reasoning why noise intensities are coupled and why simpler models would not capture the data, we refer the reader to the appendix 4.5.1 (see also Fig. 4.9).

4.2.4. NOISE AT LOW EXPRESSION LEVEL SCALES DIFFERENTLY BUT LINEARLY, WHICH CONFIRMS THE MODEL.

To test for the generality of the model, we investigated whether it could explain changes occurring upon a decrease in gene expression level. Therefore, we grew cells in the absence of Iptg, which decreased expression rate from the P_{lac} -promoters by a factor of ≈ 30. Consequently, expression noise increased roughly by a factor of eight, with the main contributor being an increase in intrinsic noise (Table 4.1, see also Fig. 4.1E) [6, 19, 21]. Unexpectedly, also extrinsic noise increased significantly (around 5-fold), resulting in a significant contribution of both noise categories to expression noise (Ta-

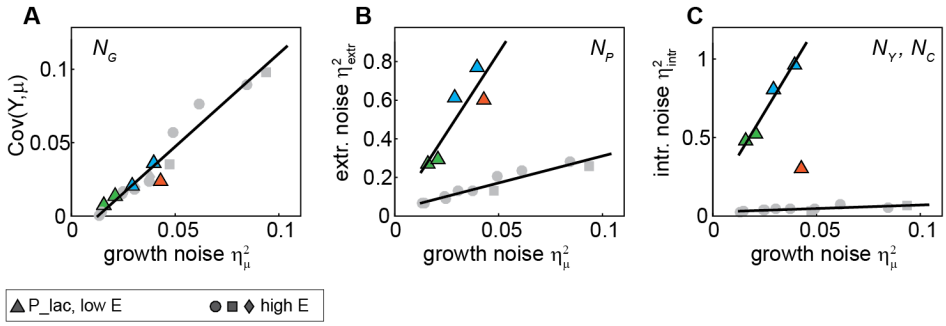


Figure 4.3: Influence of mean expression rate on scaling properties. Colored triangles: low expression, color code for growth media as in Fig. 4.2; gray symbols: high expression (taken from Fig. 4.10A and Fig. 4.2A,B). (**A**) $Cov(Y,\mu)$ in dependence of η^2_μ . Global noise, $\langle N^2_G \rangle$, determines the slope of the line and is independent of expression level. (**B**) η^2_{extr} in dependence of η^2_μ . Protein expression noise $\langle N^2_P \rangle$ mostly determines line characteristics (slope and intercept) and is thus increased. (**C**) η^2_{intr} in dependence of η^2_μ . Local noise sources are strongly increased. In (**C**) , the orange data point (growth on acetate, μ=0.2dbl/hr) was excluded from the fit because the low image acquisition frequency (every 36min, due to the low growth rate) likely missed part of the fast intrinsic fluctuations [5] and therefore underestimated η^2_{intr} .

ble 4.1). The increase in extrinsic noise is for example visible as simultaneous bursting in Y and C expression rate (Fig. 4.11). It suggests that, next to stochastic (un)binding events of repressor molecules [8, 113], also the concentration of repressors significantly affects gene expression noise [7].

The model predicts that the $Cov(Y,\mu)$ should still obey the same scaling (Fig. 4.10A) at these lower expression levels, as this covariance is determined by N_G, while changes in induction of a gratuitous protein should affect only N_P, N_Y and N_C. This was indeed the case for our experimental data (Fig. 4.3A, Fig. 4.10C).

Second, the experimental data showed that intrinsic noise strongly increased at low expression. Within the model, this corresponds mainly to an increase in N_Y and N_C. Growth and expression noise intensity should again display a linear relation, but now being steeper and potentially upshifted, depending on whether the part of the noise source with varying intensity or the noise floor is changed (appendix 4.5.1). Third, the measured increase in extrinsic noise implies an increase in N_P. As a result, the extrinsic noise as function of growth noise is expected to increase in steepness (and potentially offset), but less extreme. Indeed, we found that the experimental data followed exactly that prediction and collapsed onto steeper lines (Fig. 4.3B,C). Overall, these correspondences between model and data supported the structure of the model.

4.2.5. CORRELATION BETWEEN GROWTH AND EXPRESSION FLUCTUATIONS DEPENDS ON THEIR NOISE INTENSITIES.

We showed that noise intensities of growth and expression fluctuations are strongly correlated (R = 0.94), but that at the same time the fluctuations themselves are modestly

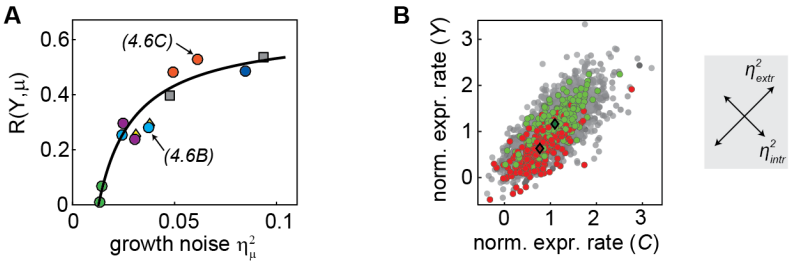


Figure 4.4: Correlation of fluctuations. (**A**) $R(Y,\mu)$ in dependence of growth noise. Black curve: Calculated relation using the line fits of the model (eq. 4.22). Numbers refer to figures of corresponding scatter plots. Color code and symbols as in Fig. 4.2. (**B**) Coupling of growth fluctuations to extrinsic and intrinsic fluctuations. Subsets with fastest (slowest) growth rates are colored in green (red), and their averages are marked as diamonds. Data size is 5% of all data for each subset. (Gray box) Scattering of data points along indicated directions is a measure for extrinsic and intrinsic noise. Expression rates were normalized by their average.

correlated ($0 \leq R \leq 0.53$). To better understand this apparent contradiction, we turned to the model. When plotting the correlation $R(Y,\mu)$ against growth noise, we found that data points followed a monotonously increasing non-linear curve (Fig. 4.4A). Thus, the larger the fluctuations in a cell, the more strongly growth and expression fluctuations are also coupled. Because correlation is defined as $R(Y,\mu) = Cov(Y,\mu)/\sqrt{\eta_Y^2 \eta_\mu^2}$, its dependence on growth noise can be readily computed from the line fits for $Cov(Y,\mu)$, η_Y^2 , and η_μ^2 by the model (Fig. 4.4A, black line; eq. 4.22 in appendix). The rather small $R(Y,\mu)$ can be explained with the structure of the model noise sources: Only the global noise source N_G causes correlations in growth and expression fluctuations, while all other, independent noise sources cause incoherent fluctuations, thus limiting the correlation. In the low-noise limit, which is for example realized in rich growth medium, N_G vanishes, which results in uncorrelated fluctuations of growth and gene expression (because other noise sources do not vanish, see appendix 4.5.1). Using the analytical expression for the curve (eq. 4.22), we found that the correlation increases linearly with growth noise for low noise intensities (Fig. 4.4A). Interestingly, in the high-noise regime the correlation saturates at ≈ 0.6, that is, significantly below one. Thus, the model suggests the existence of a limit for correlations $R(Y,\mu)$ that cannot be exceeded in cells.

We wondered to what extent the growth fluctuations are coupled specifically to extrinsic, respectively intrinsic, expression fluctuations. Since intrinsic noise is thought to result only from stochastic fluctuations in local reactions, one would expect these fluctuations to be independent of growth fluctuations. Indeed, when plotting two subsets of cells with temporary higher or lower growth rate, we found that these clouds were separated along the axis associated with extrinsic fluctuations (Fig. 4.4B, Fig. 4.12A-C).

The correlation between growth and extrinsic (intrinsic) expression fluctuations cannot be quantified directly. The reason is that for individual data points it cannot be determined what part of the fluctuation is extrinsic or intrinsic - only the noise intensities (η^2_{extr} , η^2_{intr}) are experimentally accessible. Therefore, we used a (co)variance decom-

position approach, which could separate the (accessible) expression noise into growth-related and unrelated fluctuations [123–125]:

$$\eta^2_{extr} = Cov\left(\langle Y|\mu\rangle, \langle C|\mu\rangle\right) + \left\langle Cov\left(Y, C|\mu\right)\right\rangle = \eta^2_{extr,E} + \eta^2_{extr,U} \tag{4.3}$$

$$\eta^2_{intr} = \tfrac{1}{2} Var\left(\langle (Y-C)|\mu\rangle\right) + \tfrac{1}{2}\left\langle Var\left((Y-C)|\mu\right)\right\rangle = \eta^2_{intr,E} + \eta^2_{intr,U} \tag{4.4}$$

For convenience, Y, C, and μ were here normalized to an average of one. The notation $\langle ...|\mu\rangle$ indicates that the conditional average is calculated for a given growth rate. Similarly, $Cov\left(Y, C|\mu\right)$ and $Var\left((Y-C)|\mu\right)$ denote conditional covariance and variance. Theoretically, variables need to be conditioned not only on current growth rate but its full history [123, 124]. To test the relevance of history in this case, we compared conditioning on only current growth rate to additionally considering one past growth data point. The results were not affected and we therefore omitted growth rate history.

The first term in eq. 4.3 is the covariance "explained" by growth rate fluctuations (denoted $\eta^2_{extr,E}$) because all other sources of variability were already averaged out by the inner brackets. Argumentation and notation for intrinsic noise are similar. Note that we adopted the typically used terminology that growth rate "explains" expression noise, but emphasize that this analysis concerns correlation, not causation.

By decomposing extrinsic expression noise (eq. 4.3), we found that growth fluctuations coupled in part to extrinsic noise. The explained fraction, defined as $F_{extr,E} = \eta^2_{extr,E}/\eta^2_{extr}$, ranged from 0 to 0.3 in the different experiments (Fig. 4.12D). To get a more intuitive understanding for $F_{extr,E}$, we linearized eq. 4.3 and thus obtained a simple relation between explained fraction and correlations (appendix 4.5.2):

$$F_{extr,E} = \frac{R(Y,\mu)R(C,\mu)}{R(Y,C)} \quad . \tag{4.5}$$

The fractions obtained with this linearized equation were the same as with the general equation (Fig. 4.13A), which further supported our restriction to a linear noise model.

For intrinsic noise, we found that the coupling to growth fluctuations was close to zero. However and unexpectedly, in several instances the explained fraction was, although low, still significantly different from zero (Fig. 4.4B: fraction 0.030, $p < 0.001$, Fig. 4.12C: fraction 0.026, $p = 0.01$. For p-value determination see Methods 4.4.2) The reason is that Y expression coupled slightly stronger to growth fluctuations than C expression (Fig. 4.7, Table 4.2), and as a consequence the central term in intrinsic explained noise, $\langle (Y-C)\,|\mu\rangle$ (eq. 4.4), is non-zero. This means, that the measured intrinsic noise is not only local noise, but a small fraction also results from different coupling to common noise sources ($\leq$ 15%, see appendix 4.5.1). Biologically, the small difference in Y and C expression fluctuations might be due to differences in local DNA accessibility (for example influenced by fluctuations in expression frequency from adjacent genes) or differences in the amino acid composition of the otherwise very similar proteins.

4.2.6. TRANSIENTLY FASTER GROWING CELLS DISPLAY LOWER EXPRESSION NOISE.

While gene expression noise and growth noise were strongly correlated, expression noise was also correlated to the mean growth rate, though less strongly (Fig. 4.8B). This decrease might be related to changes in chemical composition and operation of the cell,

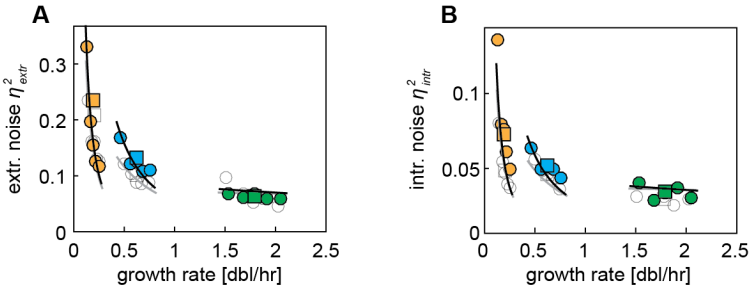


Figure 4.5: (A) Extrinsic noise and (B) intrinsic noise decrease with increasing growth rate. Squares: Population averages (also plotted in Fig. 4.8B). Circles: Data of each population was binned into five subsets according to growth rate. Black curves were calculated from the model. Color code as in Fig. 4.2, replicates are shown in light gray. Expression noise decreases with increasing growth rate for the different population averages, as well as for fluctuations within a microcolony.

availability of precursors or cell size. We wondered what the implications for expression noise were when the growth rate changed spontaneously rather than due to altered external conditions. To this end, we binned the full dataset of a microcolony into subsets with different growth rates and then determined expression noise for each subset. Note that since growth rate fluctuates faster than the cell cycle [56], different data points of the same cell may belong to different bins.

We found that within a microcolony, both extrinsic and intrinsic noise decreased with increasing growth rate (Fig. 4.5). In other words, transiently faster growing cells displayed lower-than-average expression noise. Thus, the same qualitative trend for expression noise holds for different colony-average growth rates and random growth variations within a colony (Fig. 4.5, squares, respectively circles). By applying our model (black lines), we could explain the decrease in expression noise during random growth upfluctuations: Transiently faster growing cells produce on average more proteins, but the model suggested that variance of expression is unaffected (appendix 4.5.1) and therefore expression noise (variance divided by squared mean) is decreased.

On a quantitative level, differences between the various decreasing "expression noise versus growth" functions ("η^2_E vs. μ ") could be discerned. The following argumentation holds for both extrinsic and intrinsic noise individually, as well as for the total expression noise. First, the range of expression noise intensities covered in slowly growing populations (orange circles) was much larger than at faster mean growth rate (green circles). In the model, the steepness of the η^2_E vs. μ curve depends positively on the magnitude of the global noise source and thus positively on growth noise (appendix 4.5.1). Since slower growing microcolonies typically displayed larger growth noise (Fig. 4.8A), the η^2_E vs. μ function decreased more steeply and thus a wider η^2_E range was covered. Second, η^2_E vs. μ curves obtained from single-cell fluctuations did not collapse onto a curve that connected the population averages. This has an interesting implication: Cells that by chance grow at the same speed, but belong to different colonies with different mean growth rates, are predicted to be distinguishable by their expression fluctuations.

Expression noise intensity of the transiently faster growing cells (belonging to the colony with low average growth) is predicted to be lower.

4.3. DISCUSSION AND CONCLUSIONS

We found that fluctuations in gene expression and growth rate are inherently related and that their noise intensities scale linearly. For gene expression or metabolic networks that contain well defined functional links, such a coupling and in that case even propagation of fluctuations was previously shown [26, 27, 56]. Here, we found that noise in different cellular parameters is generically connected, also in the absence of a direct causal link.

Noise intensities were strongly correlated, while fluctuations in gene expression and growth were found to be only modestly correlated. This underlined the importance of clearly distinguishing between fluctuating time traces and the magnitude of these fluctuations. Our results implied that, in a model, cellular noise can be grouped into few different noise sources that all fluctuate non-synchronously, but whose noise magnitudes vary jointly. Thus, one single parameter could describe the "level of chaos" in a cell. Biologically, such a noise structure could for example arise when different pathways compete for the same limiting resources or enzymes [126, 127]. Up-fluctuations in one flux could entail down-fluctuations in different pathways that require the same biomolecules. Then, magnitude of such fluctuations may be coupled, while the up-fluctuations themselves are not synchronous. Alternatively, de facto non-synchronous noise sources could arise when consequences of different fluctuating components (ribosomes, ATP, ...) become apparent at variable time delays. For example, a random fluctuation in amino acid pool may immediately affect protein expression (N_P in the model), which could then alter rates of biosynthetic fluxes at a delay (once enzyme concentrations are significantly modified) [56]. If the latter affects growth and expression (N_G in the model), both sources of fluctuations would be coupled but not in synchrony. It would be interesting to experimentally test these proposed scenarios in future research.

Besides scaling, our results suggested the existence of a non-zero minimum noise floor, which appears to be present even in the most favorable growth environment. This would mean that reducing noise in cells even further is either not possible or would be too costly. Indeed, theory studies showed that suppression of molecular fluctuations is extremely costly [128] and that signaling accuracy is limited by the finite pool of resources [129].

Noise intensities of growth and expression rate changed systematically with the environment and typically decreased with increasing average growth rate. While this observation appears in line with recent results from yeast [121], we note that that study identified the varying fraction of cells in different cell-cycle stages (pre/post chromosome replication) as major source of the changing noise level. This effect should be minor in our data because we had subtracted the quasi-periodic cell-cycle trend from the protein production rate. Subtracting the trend should cancel the effect of varying chromosomal copy number at the beginning and end of a cell-cycle, at least to a significant extent. Besides, the cell-cycle in yeast is structured differently than in bacteria, and possesses more well defined distinct phases that display differences in gene expression patterns. Our results suggested that also other, unknown factors can cause an inverse relation between

noise intensity and average growth rate.

In literature, growth noise was both reported to depend and be independent of average growth rate [32, 56, 115]. The reason for this discrepancy is not known but might be a result of the different used strains. We would argue that since proteins are required for cellular growth and protein noise is environment-dependent [56, 121], a scenario in which growth noise changes as well is likely. Interestingly, whether noise in interdivision times changes with mean growth rate [33] cannot be directly inferred from the dependence of growth noise on mean growth rate [115]. Growth noise levels at lower average growth rate might be increased due to the on average smaller size of cells [1], because in a smaller volume the ability to average over molecular fluctuations is decreased [130]. Alternatively, the risk of transiently experiencing shortages of required biomolecules (such as amino acids) might be increased in poorer growth conditions since they are then not provided with the medium. In addition, lower concentrations of for example RNA polymerase at lower growth rates could increase extrinsic and intrinsic expression noise [120, 121]. Extrinsic noise can also already be affected if only noise intensity but not average concentration of RNA polymerase varies [120]. Intrinsic noise in particular may also increase due to the on average lower number of chromosomal copies at slow growth and therefore transcription from fewer gene copies [1]. We remind that growth rate dependence of the average copy number per cell (averaged over a cell's interdivision cycle) is a different aspect than a changing partitioning of a population of cells into different cell-cycle stages (measured at one time point) [121].

In our study, varying the average growth rate was predominantly a means to modify noise intensities, and we found that the tight coupling of expression and growth rate occurred actually on the level of noise intensities. Datasets with increased noise levels, that deviated from the typical relation between noise intensities and average growth rate, collapsed onto the same linear relation between growth and expression noise.

Both gene expression and growth noise can have important consequences for cellular function, phenotype and fitness [9, 12, 15, 30]. However, many aspects of fluctuations in these parameters are unknown. In particular, we know little about the molecular mechanisms that underlie the coupling between these fluctuations, the effect of the environment on noise intensities, and possible constraints on the magnitude of the fluctuations. Especially growth fluctuations have only recently been investigated, despite their importance for cell proliferation. This study aimed to better understand some of these inherent relations between fluctuations. Learning more about properties of cellular noise will help to better understand bacterial behavior. Active manipulation of noise intensities [130] or noise transmission [56] would be an interesting next step.

4.4. MATERIALS AND METHODS

A detailed description of experiments and image analysis methods can be found in chapter 2. Here, we focus on describing methods and analysis tools specific to this study and only briefly summarize the common methods.

4.4.1. EXPERIMENTS AND IMAGE ANALYSIS

BACTERIAL STRAINS

In the experiments we used three different strains, which are all derivatives of *E. coli* strain MG1655. MG22 [7] contains the fluorescent genes *yfp* and *cfp* under control of P_{lac} promoters, located at *intC* and *galK*, which are at opposite halves of the chromosome and equidistant from the origin of replication. M22 [7] is the same as MG22, except for that expression of *yfp* and *cfp* is controlled by strong constitutive P_{λ} promoters. ASC659 contains *gfp* under control of a constitutive P_{N25} promoter at the *che* location. The strain was constructed using the Datsenko & Wanner protocol [58]. If not specified further, strain MG22 was used in the displayed plots, for example in scatter plots (Fig. 4.1D, or 4.4B) and conditional expectation plots (Fig. 4.7).

Growth rate of different M22 cells varied, was sensitive to preparation steps and fluctuations were typically large. Presumably, high noise levels are due to the strong expression of gratuitous proteins, which could unbalance metabolism. Since expression of both reporters was strongly positively correlated, we excluded direct titration effects, because competition for resources should decrease correlation between expression rates of the proteins.

EXPERIMENTS

For microscopy, cells were grown in either M9 minimal medium or rich defined medium (MOPS EZ rich defined medium from Teknova). M9 was supplemented with 0.2mM uracil and one of various carbon sources (0.18% acetate, 0.1% maltose, 0.1% lactose). The rich medium contained 0.2% glycerol as carbon source. When full induction in strain MG22 was required, 200μM Iptg was added to the growth medium. For a full list of experiments, see Table 4.1.

Cells were grown and prepared for microscopy as described in chapter 2. Movies of growing microcolonies were acquired on the Nikon TE2000-E inverted microscope (chapter 2.2). Imaging frequency was on average 55 phase contrast and 8 fluorescence images per cell cycle and fluorescence illumination time was kept as short as possible. We confirmed that cells in the colony center grew on average at the same speed as at the edge of the colony.

We call a dataset to be at "high expression" when extrinsic noise dominates over intrinsic noise (Table 4.1). These experiments correspond to expression in the extrinsic noise limit (floor), depicted for example in Taniguchi et al. [6]. Next to high expression, we consider one "low expression" level, which is the basal expression level from the (truncated) P_{lac} -promoters. Here, intrinsic noise in protein production rate was larger than extrinsic noise (Table 4.1).

IMAGE ANALYSIS

Images were analyzed offline as described (chapter 2). Briefly, cells were segmented and tracked and cell length as well as instantaneous growth rate was determined. Cellular length can be used as measure for cell size because cell width is constant throughout the cell cycle (Fig. 5.12, [32]. Fluorescence images were corrected and concentration and production rate was calculated for each cell. Both protein production rate and growth rate were corrected for quasi-periodic cell-cycle fluctuations.

4.4.2. DATA ANALYSIS

CORRELATIONS

For analysis, we used a time window of the experiment in which population average growth rate and protein concentration was roughly constant. Few cells stopped growing or filamented and where additionally removed from the dataset. The linear noise-scaling (e.g. Fig. 4.2, 4.3) was robust to taking the full time window of the experiment. Naturally, then absolute noise intensities increased. Data sizes for each experiment can be found in Table 4.1 (ca 8 data points belong to one cell with completely observed cell cycle).

We used standard (Pearson) correlations to quantify the coupling of fluctuations in two signals. This requires that the fluctuations occur simultaneously and not with a time-delay relative to each other. Such delays were observed in cells [27, 56] and we therefore calculated cross-correlations (chapter 2.3.6) to test whether the maximum correlation would occur at a delay. We found that the cross-correlation of protein production rate and growth rate peaked at a delay of zero, which means that the two signals fluctuated simultaneously (Fig. 4.14). We also confirmed that fluctuations in production rates Y and C were simultaneous, which was expected because they are under control of identical promoters.

NOISE DECOMPOSITION AND CONDITIONAL EXPECTATION VALUES

To determine the conditional expectation values and (co)variances in the noise decomposition equations (eqs. 4.3, 4.4), we used kernel density estimate (KDE) (MatLab script for one and two-dimensional KDE published by [131]). By applying KDE, we directly obtained continuous functions for the probability distribution of single variables (e.g. $P(Y)$ and $P(\mu)$), as well as for joint distributions (e.g. $P(Y,\mu)$). All conditional probability distributions such as $P(Y|\mu) = P(Y,\mu)/P(\mu)$ could then be calculated from these distributions. Then, conditional expectation values could directly be calculated. As example, the conditional expectation $\langle Y|\mu\rangle$ depicted in Fig. 4.1D, which is defined as $\sum P(Y|\mu)\cdot Y$, was calculated by summing over a fine grid of Y-values (>200 points) at fixed μ. Calculation of $Cov(...|\mu)$ and $Var(...|\mu)$ was similar.

Displayed error bars for $\langle Y|\mu\rangle$ (for example Fig. 4.1D, same for $\langle C|\mu\rangle$) are standard errors of the mean and were obtained by bootstrapping. For each resampled set of the experimental data we determined the probability distributions with KDE and then the conditional expectation function $\langle Y|\mu\rangle$. The distribution of conditional expectation values was then used to compute its standard error at each Y-value.

To test robustness of our results to the chosen analysis method, we also determined the noise decomposition (eqs. 4.3 and 4.4) with an alternative binning method. Here, the data (Y, C, μ-triplets) was binned according to growth rate into 15 bins, with equal number of data points each. Then, averages were calculated for each bin, which yielded for example $\langle Y|\mu\rangle$, defined at 15 discrete μ-values (similar calculation for (co)variance). The chosen number of bins was a compromise between covering growth rate in sufficiently fine steps and still having enough data points per bin to calculate a meaningful average/(co)variance. We found that the choice of method, KDE or binning, did not affect the results (Fig. 4.13B).

SIGNIFICANCE TEST

The p-values for the explained intrinsic fraction (section 4.2.5) were determined by using Monte Carlo sampling. A non-zero explained fraction suggested that the coupling of the two reporter proteins to growth rate differs (eq. 4.4). Therefore, we tested the null-hypothesis that the correlation of YFP and CFP production rate with growth rate is identical. We created 10.000 sample datasets with each the size of the experimental dataset. For sample generation, Y, C and μ were drawn from Gaussian distributions. Standard deviations of the distributions were taken from the experimental dataset and correlations $R(Y,\mu)$ and $R(C,\mu)$ were kept equal and constant (we tested taking the average or either one of the experimental values). For each sample, the explained fraction of intrinsic noise was calculated. Then, the p-value was determined as fraction of samples that yielded an explained fraction of at least the measured value.

At high expression, the null-hypothesis was rejected ($p \leq 0.01$). Therefore, the non-zero explained intrinsic fraction was not simply an artifact of small data size. At low expression, the null-hypothesis could not be rejected ($p = 0.48$).

4.5. APPENDIX

4.5.1. NOISE MODEL

CONSTITUTING EQUATIONS

In the model, protein production rates $C(t)$, $Y(t)$ and growth rate $\mu(t)$ are the sum of different, partly shared and partly specific, random fluctuations $N_Q(t)$. Q shall refer to the indexes for "global" (G), "growth specific" (S) etc, see Fig. 4.2C and Fig. 4.9A. With one exception (growth specific noise $N_S(t)$ was set to have constant intensity, see also analysis below), each of these noise sources is the sum of two types of fluctuations: One part, $\nu_{Q0}(t)$, has a constant noise intensity. A second part, $\nu_Q(t)$ varies in noise intensity. Both components are required to reproduce the data (see fitting below). All parts of all noise sources are defined as independent zero-mean Ornstein-Uhlenbeck processes and thus $\langle \nu_{Q1}\nu_{Q2}\rangle = \delta_{Q1,Q2}\langle \nu_{Q1}^2\rangle$ (δ is the Kronecker delta). That means that neither fluctuations from different sources nor fluctuations from the constant and varying part of the same source are coherent.

The random fluctuations with variable intensity, $\nu_Q(t)$, are transmitted with efficiencies a_{ZQ} to an observable Z (Z being Y, C, or μ). The constant-intensity noise sources $\nu_{Q0}(t)$ shall have a transmission efficiency of one because any deviating efficiencies can be directly incorporated into most of the noise amplitudes by rescaling them. Therefore, fluctuations in an observable Z, which are caused by a source Q, are:

$$N_{ZQ}(t) = a_{ZQ}\nu_Q(t) + \nu_{Q0}(t) \quad . \tag{4.6}$$

Then, the fluctuations in growth rate and protein production rate can be written as:

$$\mu(t) = a_{\mu G}\nu_G(t) + \nu_{S0}(t) + \langle\mu\rangle \quad , \tag{4.7}$$

$$Y(t) = a_{YG}\nu_G(t) + a_{YP}\nu_P(t) + a_{YY}\nu_Y(t) + \nu_{P0}(t) + \nu_{Y0}(t) + \langle Y\rangle \quad , \tag{4.8}$$

and analog for $C(t)$. Again, we will omit explicit notation of time dependence and will

write Y, C and μ. Observables (Y, C, and μ) shall each be normalized to an average of one, and thus noise and variance are equivalent (for example: $\eta_\mu^2 = Var(\mu)$).

We can now calculate noise and covariances of all observables:

$$\eta_\mu^2 = a_{\mu G}^2\langle \nu_G^2\rangle + \langle \nu_{S0}^2\rangle \tag{4.9}$$

$$\eta_Y^2 = a_{YG}^2\langle \nu_G^2\rangle + a_{YP}^2\langle \nu_P^2\rangle + a_{YY}^2\langle \nu_Y^2\rangle + \left(\langle \nu_{P0}^2\rangle + \langle \nu_{Y0}^2\rangle\right) \tag{4.10}$$

$$Cov(Y,\mu) = a_{\mu G}a_{YG}\langle \nu_G^2\rangle \tag{4.11}$$

$$\eta_C^2 = a_{CG}^2\langle \nu_G^2\rangle + a_{CP}^2\langle \nu_P^2\rangle + a_{CC}^2\langle \nu_C^2\rangle + \left(\langle \nu_{P0}^2\rangle + \langle \nu_{C0}^2\rangle\right) \tag{4.12}$$

$$Cov(C,\mu) = a_{\mu G}a_{CG}\langle \nu_G^2\rangle \tag{4.13}$$

Additionally, we obtain for extrinsic and intrinsic noise:

$$\eta_{extr}^2 = Cov(Y,C) = a_{YG}a_{CG}\langle \nu_G^2\rangle + a_{YP}a_{CP}\langle \nu_P^2\rangle + \langle \nu_{P0}^2\rangle \tag{4.14}$$

$$\eta_{intr}^2 = \tfrac{1}{2}\left((a_{YG}-a_{CG})^2\langle \nu_G^2\rangle + (a_{YP}-a_{CP})^2\langle \nu_P^2\rangle + a_{CC}^2\langle \nu_C^2\rangle + a_{YY}^2\langle \nu_Y^2\rangle\right) + \tfrac{1}{2}\langle \nu_{Y0}^2\rangle + \tfrac{1}{2}\langle \nu_{C0}^2\rangle \tag{4.15}$$

Note that we set $\nu_{G0} = 0$ in equations 4.7-4.15 based on the experimental data: $\langle \nu_{G0}^2\rangle \neq 0$ would be one contribution to the covariance $Cov(Y,\mu)$ and the covariance would therefore always be positive. However, in the experiments the value zero for $Cov(Y,\mu)$ was in fact measured (not only extrapolated, see Fig. 4.10).

MODEL FITTING

Solving eq. 4.9 for $\langle \nu_G^2\rangle$ and inserting into eq. 4.10-4.15 expresses all noise intensities and covariances as function of growth noise. For example, we obtain for $Cov(Y,\mu)$ (see Fig. 4.10A,B):

$$Cov(Y,\mu) = \frac{a_{YG}}{a_{\mu G}}\left(\eta_\mu^2 - \langle \nu_{S0}^2\rangle\right), \tag{4.16}$$

which is a linear function of growth noise. The coupling constants set the slope of the line, while the axis intercept is set by the constant noise floors (see also Fig. 4.9B for a different example).

For fitting the model parameters, we first plotted all noise intensities and covariances of eq. 4.10-4.14 and fitted them with linear functions to obtain axis intercepts and slopes (Fig. 4.2A,B, 4.10A). We did not include the data of strain ASC659 (P_{N25} -*gfp*) in the fitting because due to the single reporter no data on η_{extr}^2 and η_{intr}^2 was available. The fact that noise values for that strain still collapsed onto the same lines thus supported our suggested noise structure.

On first sight, the model contains many free parameters, but several of the parameters can be fixed: The normalization of noise sources with varying intensity (ν_Q) is arbitrary because only the product of (squared) transmission coefficients a_Q and noise intensity $\langle {\nu_Q}^2\rangle$ is "felt" by Y, C and μ. We therefore define $a_{\mu G} = 1$. This sets the scale of $\langle \nu_G^2\rangle$ to be the same as η_μ^2 (eq. 4.9). We show in the next paragraph that, for reproducing the data, the magnitude-varying components of the different noise sources have to vary concertedly in intensity (growth-specific noise is underdetermined and an exception, see below). Thus, all of their magnitudes can be described jointly by one global

parameter. Then, without loss of generality the scale (normalization) of the other noise sources can be defined as:

$$\langle \nu_G^2 \rangle = \langle \nu_P^2 \rangle = \langle \nu_Y^2 \rangle = \langle \nu_C^2 \rangle \quad . \tag{4.17}$$

Generality is not lost because the corresponding transmission coefficients are free parameters.

Still, the system (eq. 4.9-4.15) is underdetermined by one since eq. 4.15 is not independent but a linear combination of eq. 4.10, 4.12 and 4.14. To solve the system, we assumed that the variances resulting from the intensity-varying intrinsic noise sources for CFP and YFP are the same. That means that the coupling constants are identical: $a_{YY} = a_{CC}$. This assumption is reasonable because expression of both genes is controlled by identical promoters. We write the common coupling constant as a_{XX} and the noise source as ν_X.

Now, all model parameters can be determined. Going back to the example of the $Cov(Y,\mu)$ vs. η_μ^2 data (Fig. 4.10A,B), a linear fit yields $\langle \nu_{S0}^2 \rangle$ as x-axis intercept and a_{YG} as slope. Analogously, and by using eq. 4.17, all parameters of eq. 4.9-4.15 can be determined. The resulting fit values can be found in Table 4.2. We note that while the actual values of the fit parameters are of interest, the main result of this study is an insight into the structure or topology of noise sources: (Almost) all sources of fluctuations need to vary in intensity and, especially, the intensity-variation is concerted.

SIMPLER MODELS AND ALTERNATIVES TO THE CONCERTED VARIATION OF NOISE SOURCES

For several noise sources (N_P, N_C, N_Y, N_S) we allowed for a noise floor of constant intensity, next to a intensity-varying part. Mathematically, these noise floors can shift the linear relations between noise intensities along both x-axis and y-axis. We had measured non-zero axis intercepts (Fig. 4.2A,B and 4.10), which implied the existence of such noise floors within the model. This suggests that cells are subject to a non-zero minimum noise level below which they cannot get, even in the most favorable environment.

Most of the noise sources in the model (N_G, N_P, N_C, N_Y) contained a component that varied in magnitude. We tested the consequences if one or more of these components were forced to be zero (that is, the respective noise source then had constant intensity). When forcing the intrinsic noise sources (N_C, N_Y) to have constant intensity, then the model could not capture the measured scaling of intrinsic noise any more (Fig. 4.9C-E). If the shared protein source (N_P) was additionally forced to be constant, then not even extrinsic noise could be fitted any more (Fig. 4.9F-H). The global noise source N_G clearly could not have constant intensity because of the variation measured for $Cov(Y,\mu)$ (Fig. 4.10).

One could also consider using a simpler model with less noise sources, for example including only one extrinsic source of fluctuations. However, we just showed that the data could not even be fitted any more if the number of noise floors or intensity-changing sources was restricted. Thus, fully omitting a noise source would lead to even worse fit results. In principle, one could also think of noise sources acting only on one protein and growth (for example Y and μ). However, such sources are not likely, given that both C and Y denote protein expression and due to the symmetry of the construct.

The most striking feature of the model is the concerted variation of the different noise intensities (eq. 4.17). We here argue that, if variation was not concerted, then data points would scatter randomly and not collapse onto lines. Consider a set of experiments with the same amount of global noise $\langle N_G^2 \rangle$, then growth noise η_μ^2 is the same in all of these experiments (eq. 4.9). But if the magnitude of protein noise $\langle N_P^2 \rangle$ can for example vary independently, then η_Y^2 (η_E^2) varies arbitrarily for fixed growth noise (eq. 4.10). Thus, these data points would not lie on the line of Fig. 4.2A (that is, collapse onto one point because growth noise is constant) but randomly scatter around it. This, however, contradicts the experimental data. Similar arguments can be made for each noise source.

VARYING THE INTENSITY OF THE GROWTH SPECIFIC NOISE SOURCE N_S

We set the variance of growth specific noise $\langle N_S^2 \rangle$ to be constant because we aimed for the simplest model to describe the data. In this section we test the effects of permitting this noise source to vary as well in intensity. The biological and experimental origins of $N_S(t)$ are not very clear and could for example contain fluctuations in efficiency of cell wall insertion, osmotic pressure, but also uncertainties in cell segmentation. Therefore, independent variation and probing of $\langle N_S^2 \rangle$ is likely impossible. Given that, it is not possible to know how/whether this noise source varies in intensity. However, since all other noise sources varied concertedly, we considered this the most likely mode of variation for $\langle N_S^2 \rangle$. Therefore, to investigate the consequences of varying $\langle N_S^2 \rangle$, we assumed the noise intensity to vary jointly with other noise sources. Then, growth rate reads as

$$\mu(t) = a_{\mu G} \nu_G(t) + a_{\mu S} \nu_S(t) + \nu_{S0}(t) + \langle \mu \rangle \tag{4.18}$$

and growth noise intensity is

$$\eta_\mu^2 = a_{\mu G}^2 \langle \nu_G^2 \rangle + a_{\mu S}^2 \langle \nu_S^2 \rangle + \langle \nu_{S0}^2 \rangle \quad . \tag{4.19}$$

Analogously to eq. 4.17 and using that noise sources vary concertedly, we set

$$\langle \nu_G^2 \rangle = \langle \nu_S^2 \rangle \quad . \tag{4.20}$$

The relative strength of the two noise sources acting on growth is then set by the transmission coefficients. For a vanishing intensity-varying growth noise, we had set $a_{\mu G} = 1$, which fixed the scale of $\langle \nu_G^2 \rangle$, and $a_{\mu S} = 0$, see also above and main text. Using this, and inserting eq. 4.20 into eq. 4.19, we obtain the relation

$$a_{\mu S}^2 + a_{\mu G}^2 = 1 \quad . \tag{4.21}$$

Every $0 \leq a_{\mu G} \leq 1$ then defines a version of the model in which the growth specific noise source varies in a different range and thus the relative contributions of growth specific and global noise source to η_μ^2 change. For each $a_{\mu G}$ this more general model can then be fitted to the experimental linear relations (Fig. 4.2, Fig. 4.10) by using eqs. 4.10-4.14 and eq. 4.19. The coupling constants are obtained as function of $a_{\mu G}$ (Fig. 4.15). We found that the fitted transmission efficiencies to common noise sources depended on the chosen model, specified by $a_{\mu G}$, while transmission efficiencies from intrinsic sources were hardly affected.

Why fit values for transmission efficiencies change can be understood intuitively for the example of $Cov(Y,\mu)$, plotted in dependence of growth noise: The product $a_{\mu G}a_{YG}$ determines the slope of this function. Since the experimental slope is fixed, a decrease of $a_{\mu G}$ in the model is compensated by an increase in a_{YG}. We emphasize that the arbitrariness of coupling constants is purely on the level of the theoretical model. It does not affect fitting precision or predictions for any observable statistic such as noise, correlations or explained fractions.

LOW EXPRESSION: PREDICTIONS AND FITTING OF THE MODEL

When lowering expression level, both noise floor (ν_{Q0}) and intensity-varying component ($a_{ZQ}\nu_Q$) of the different noise sources could be affected. An increase in noise floor would lead to an up-shift of the expression noise lines (see e.g. $\langle\nu_{P0}^2\rangle$ in eq. 4.10). An increase in the intensity-varying component is reflected in larger coupling constants because we fixed the scale of the noise sources (eq. 4.17). Larger coupling constants would increase the steepness of the expression noise vs. growth noise lines (see e.g. a_{YP} in eq. 4.10).

The model predicts that covariances and noise obtained from low expression data also collapse onto lines when plotted against growth noise. This prediction can be directly made from the structure (and concerted noise variation) of the model - if data would not collapse onto lines, then the model structure would have been wrong. Additionally, we can predict that lines can be both up-shifted (increased noise floor) and steeper (increased coupling constants).

Plotting $Cov(Y,\mu)$, η_{extr}^2 and η_{intr}^2 against growth noise (Fig. 4.3) reveals properties of N_G, N_P and N_Y (N_C) for the following reasons: Since the first relation is unaffected by expression level, we can conclude that N_G is not changed. Then, considering the equation for extrinsic noise (eq. 4.14), we find that all changes in offset and slope must be a result of changing N_P (that is, ν_{P0} and coupling a_{YP} and a_{CP}). Finally, the intrinsic noise line is affected by both N_P and the local N_C and N_Y sources. However, we know (see main text) that at low expression intrinsic noise increased much stronger than extrinsic noise. Thus, the increase in slope and offset of η_{intr}^2 vs. η_μ^2 is largely due to changes in N_Y and N_C. Note that we here present this approximation only to provide an intuition for Fig. 4.3C. For fitting, we used the full equation for intrinsic noise.

We found that η_{extr}^2 and η_{intr}^2 collapsed onto lines, which confirmed the model. Then, by fitting the data, we extracted the increased noise floors and coupling constants for N_P, N_Y and N_C. The coupling constants to the intensity-varying component of intrinsic noise was roughly increased by $\sqrt{60}$, which increased the relative contribution of that noise component by a factor of ≈ 60, compared to high expression. Coupling constants to the shared protein source (intensity varying component) were increased by roughly $\sqrt{10}$, increasing the relative contribution of that noise source by a factor of ≈ 10. Noise floors of these sources were up-shifted as well, but less strongly. The here mentioned factors are however dependent on whether growth specific noise source (N_S) is set to have constant intensity. If we allow its magnitude to vary, we find an increase of > 10 for the changed contribution of shared protein noise and still roughly 60 for the local sources.

One consequence of the increase in protein specific noise sources is that at low expression correlations between expression and growth were very small (Fig. 4.6A,C). The

reason is that the relative contribution of N_G, which is the only source that couples expression and growth, decreased strongly compared to contribution of the other, expression dependent, noise sources.

INTRINSIC NOISE

Intrinsic noise is by definition local or gene specific [19] and is therefore in the model described by the local noise sources N_Y and N_C (Fig. 4.2C). N_Y and N_C cause differences in Y and C. Contrary to that, the common noise sources N_G and N_P should theoretically affect both reporter genes in the same way due to the symmetry of the construct. However, we measured a difference in the coupling constants of the two reporter genes towards the common noise sources, with YFP coupling stronger than CFP to the shared noise sources (Table 4.2). This difference in coupling constants caused for example a non-zero fraction of intrinsic noise that is captured by growth noise. Of the experimentally measured intrinsic noise at high expression, this coupling difference accounts for 15% of the intrinsic noise and the local sources (that it, what is theoretically considered "intrinsic") contribute 85% to intrinsic noise. At low induction, the relative contribution of the local sources is higher.

ANALYTICAL EXPRESSION FOR CORRELATION AND EXPLAINED FRACTION

The expressions for correlation $R(Y,\mu)$ (Fig. 4.4A) and extrinsic explained fraction $F_{extr,E}$ (Fig. 4.12D) can be calculated directly from eq. 4.9-4.15. For the correlation we obtain:

$$\text{Correlation: } R(Y,\mu) = \frac{a_{YG}\left(\langle\eta_\mu^2\rangle - \langle {\nu_{S0}}^2\rangle\right)}{\left(\phi\left(\eta_\mu^2 - \langle\nu_{S0}^2\rangle\right) + \langle\nu_{P0}^2\rangle + \langle\nu_{P0}^2\rangle\right)^{1/2}\left(\eta_\mu^2\right)^{1/2}} \tag{4.22}$$

Here, $\phi = a_{YG}^2 + a_{YP}^2 + a_{XX}^2$ is the slope of the line η_Y^2 vs. η_μ^2. For low growth noise, the correlation starts at zero and increases linearly with η_μ^2. At high growth noise, it saturates at $a_{YG}/\phi^{1/2} \approx 0.6$. This functional shape can be understood intuitively by considering the relative contribution of noise sources in the different growth noise regimes. The only source that couples Y and μ and thus induces correlations is the global noise N_G. At low noise, it vanishes because it does not have a non-zero noise floor and thus the fluctuations are uncorrelated. At high noise, the intensity-varying components of the noise sources are dominant and the saturation value of the correlation is defined by the relative strength of coupling to N_P and N_G.

For the explained extrinsic fraction we obtain:

$$\text{Explained extrinsic fraction: } F_{extr,E} = \frac{a_{CG}a_{YG}\left(\eta_\mu^2 - \langle {\nu_{S0}}^2\rangle\right)^2}{\eta_\mu^2\left(\gamma\left(\eta_\mu^2 - \langle\nu_{S0}^2\rangle\right) + \langle\nu_{P0}^2\rangle\right)} \tag{4.23}$$

Here, $\gamma = a_{YG}a_{CG} + a_{YP}a_{CP}$ is the slope of the line $Cov(Y,C)$ vs. η_μ^2. For low growth noise, $F_{extr,E}$ increases quadratically with η_μ^2 and at high growth noise it saturates at $a_{CG}a_{YG}/\gamma \approx 0.4$.

DERIVATION OF: TRANSIENTLY FASTER GROWING CELLS DISPLAY LOWER EXPRESSION NOISE

The expression noise of a subset of cells with a given growth rate can be calculated by using the probability distributions of Y and C conditioned on μ, instead of the global probability distributions. We here derive the analytical expression for extrinsic noise but intrinsic noise can be derived similarly. For a given growth rate μ, extrinsic noise of the subset of cells growing at this rate is

$$\eta^2_{extr}(\mu) = \frac{Cov(Y,C|\mu)}{\langle Y|\mu\rangle\langle C|\mu\rangle} \quad . \tag{4.24}$$

Using the expressions for growth rate and protein production fluctuations, as well as the equation for growth noise (eqs. 4.7, 4.8 and 4.9), we obtain the conditional expectation

$$\begin{aligned}\langle Y|\mu\rangle &= \langle Y\rangle + a_{YG}\frac{\eta^2_\mu - \langle \nu^2_{S0}\rangle}{\eta^2_\mu}\Delta\mu \\ &\equiv \langle Y\rangle + K_Y\cdot\Delta\mu \quad .\end{aligned} \tag{4.25}$$

Analogous for C. $\Delta\mu$ was defined as deviation of the current growth rate from its average, $\Delta\mu = \mu - \langle\mu\rangle$, and thus $\langle\Delta\mu\rangle = 0$. K_Y is a constant within each experiment but for different experiments depends positively on growth noise intensity. $\langle Y|\mu\rangle$ increases linearly with growth rate (Fig. 4.6 lines). Similarly, we obtain for the conditional covariance:

$$\begin{aligned}Cov(C,Y|\mu) &= a_{YG}a_{CG}\cdot Var(\nu_G|\mu) + a_{YP}a_{CP}\cdot\langle\nu^2_P\rangle + \langle\nu^2_{P0}\rangle = \ldots \\ &= \eta^2_{extr} - \eta^2_{extr,E} = \eta^2_{extr,U}\end{aligned} \tag{4.26}$$

Thus, the conditional covariance (also reflected in the size of red and green subset scatter plots in Fig. 4.4B) is constant and independent of the growth rate of the subset. Inserting the last two results into eq. 4.24 finally yields:

$$\eta^2_{extr}(\Delta\mu) = \frac{\eta^2_{extr,U}}{\big(\langle Y\rangle + K_Y\Delta\mu\big)\big(\langle C\rangle + K_C\Delta\mu\big)} \quad . \tag{4.27}$$

This means that the extrinsic noise of a subset of cells with specific growth rate depends inversely on the growth rate. The larger the growth noise, the larger K_Y (and K_C) and therefore the stronger the dependence of $\eta^2_{extr}(\Delta\mu)$ on $\Delta\mu$.

To obtain the functions plotted in Fig. 4.5, we used the previously fitted model parameters (same values for all experiments), as well as the average growth rate and growth noise for the specific experiment.

4.5.2. LINEARIZATION OF (CO)VARIANCE DECOMPOSITION

If fluctuations in production rate and growth rate depend linearly on each other, then the (co)variance decomposition of extrinsic and intrinsic noise (eq. 4.3, 4.4) can be simplified. Here, we derive the linearization of the decomposition of extrinsic noise (eq. 4.5), the derivation for intrinsic noise is very similar. Let

$$Y(t) = \alpha_Y + \beta_Y\mu(t) + r_Y(t) \quad . \tag{4.28}$$

Here, α_Y and β_Y are offset and slope of the linear relation between $Y(t)$ and $\mu(t)$, $r_Y(t)$ is a random noise term (deviation of data points from the line). Let $C(t)$ be defined in an analog way. For simpler notation we again omit the explicit notation of time dependence. We also subtract the mean values so that all averages are zero ($\langle\mu\rangle = \langle Y\rangle = \langle C\rangle = 0$), and thus α_Y=0. Then:

$$\langle Y|\mu\rangle = \beta_Y \mu \ , \text{ and} \tag{4.29}$$

$$R(Y,\mu) = \frac{Cov(Y,\mu)}{\sigma_Y \sigma_\mu} = \frac{\beta_Y \sigma_\mu}{\sigma_Y} \ , \text{ because } Cov(Y,\mu) = \beta_Y \sigma_\mu^2 \ . \tag{4.30}$$

σ denotes the standard deviation. An equivalent expression can be derived for C. Using eq. 4.29 and eq. 4.30, the explained noise is then

$$\eta^2_{extr,E} = Cov(\langle Y|\mu\rangle, \langle C|\mu\rangle) = \beta_Y \beta_C \sigma_\mu^2 = R(Y,\mu)R(C,\mu)\sigma_Y \sigma_C \ . \tag{4.31}$$

Then, the explained fraction of extrinsic noise can be written as a combination of correlations:

$$F_{extr,E} = \frac{\eta^2_{extr,E}}{\eta^2_{extr}} = \frac{R(Y,\mu)R(C,\mu)\sigma_Y \sigma_C}{Cov(Y,C)} = \frac{R(Y,\mu)R(C,\mu)}{R(Y,C)} \ . \tag{4.32}$$

We compared the general noise decomposition (eq. 4.3) with this linear approximation and found that the obtained values were very similar (Fig. 4.13A). For the model, which is by construction linear, the linear approximation derived here is the exact solution.

Note also that in the case of linear coupling between gene expression and growth the conditional expectation $\langle Y|\mu\rangle$ is a line, when plotted as function of μ. This is very close to the dependence obtained from the experimental data (black lines in Fig. 4.1D and 4.6). The slope of the line is $Cov(Y,\mu)/\sigma_\mu^2 = \beta_Y$.

4.5.3. ADDITIONAL FIGURES

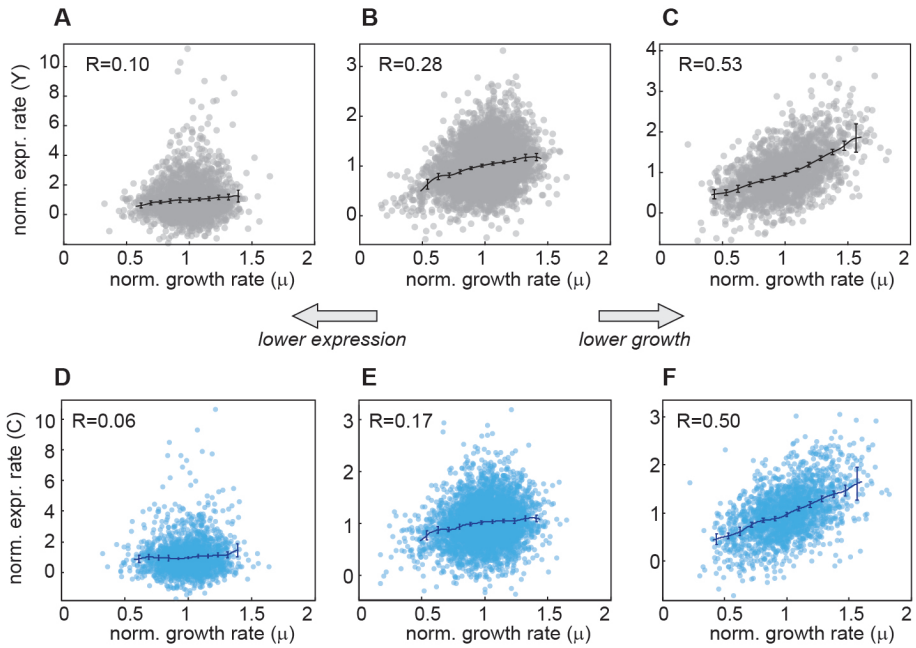


Figure 4.6: Correlation of YFP and CFP protein production rate with growth rate for different environments. (**A**) - (**C**) YFP production rate, (**D**) - (**F**) CFP production rate, obtained from the same experiments (and strain) as for YFP. Scatter plots and correlation coefficients (R) are similar for both reporters and change jointly with the growth conditions. Specifications of growth conditions: (**A**) ,(**D**) : μ=0.6dbl/hr, basal expression from P_{lac} . (**B**) ,(**E**) : μ=0.6dbl/hr, expression induced with 200μM Iptg. (**B**) is identical to Fig. 4.1D. (**C**) ,(**F**) : μ=0.2dbl/hr, expression induced with 200μM Iptg.

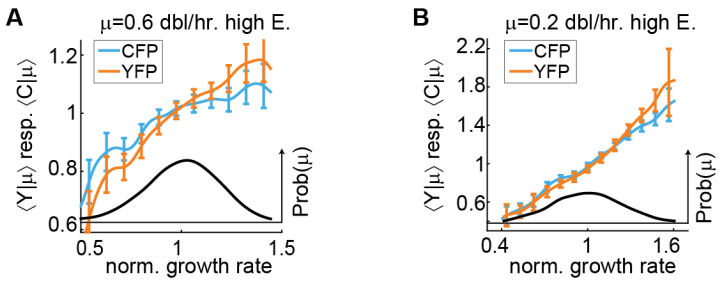


Figure 4.7: Average production rate in dependence of growth rate. The mean expression rates for given growth rate ($\langle Y|\mu\rangle$ and $\langle C|\mu\rangle$) are plotted for two different environments. (**A**) Dataset of Fig. 4.1D, which is the same as for Fig. 4.6B,E. (**B**) Dataset of Fig. 4.6C,F. Black curve: Growth rate distribution, thus indicating which expression rates of the yellow and blue curves are most abundant in a colony. Distribution not to scale between (**A**) and (**B**) . Error bars are SEM, obtained by bootstrapping. $\langle Y|\mu\rangle$ increases slightly stronger with μ than $\langle C|\mu\rangle$.

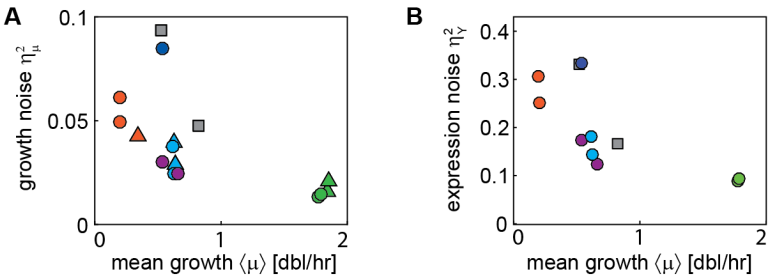


Figure 4.8: Growth noise (A) and gene expression noise (B) in dependence of average growth rate. (**A**) All datasets are displayed. (**B**) Experiments with high expression (induced P_{lac} promoters and constitutive promoters) are displayed. In both figures, noise typically decreases with increasing growth rate, but stressed cells deviated from the trend (squares, dark blue circle). Color code and symbols as in Fig. 4.2 (triangles: low expression).

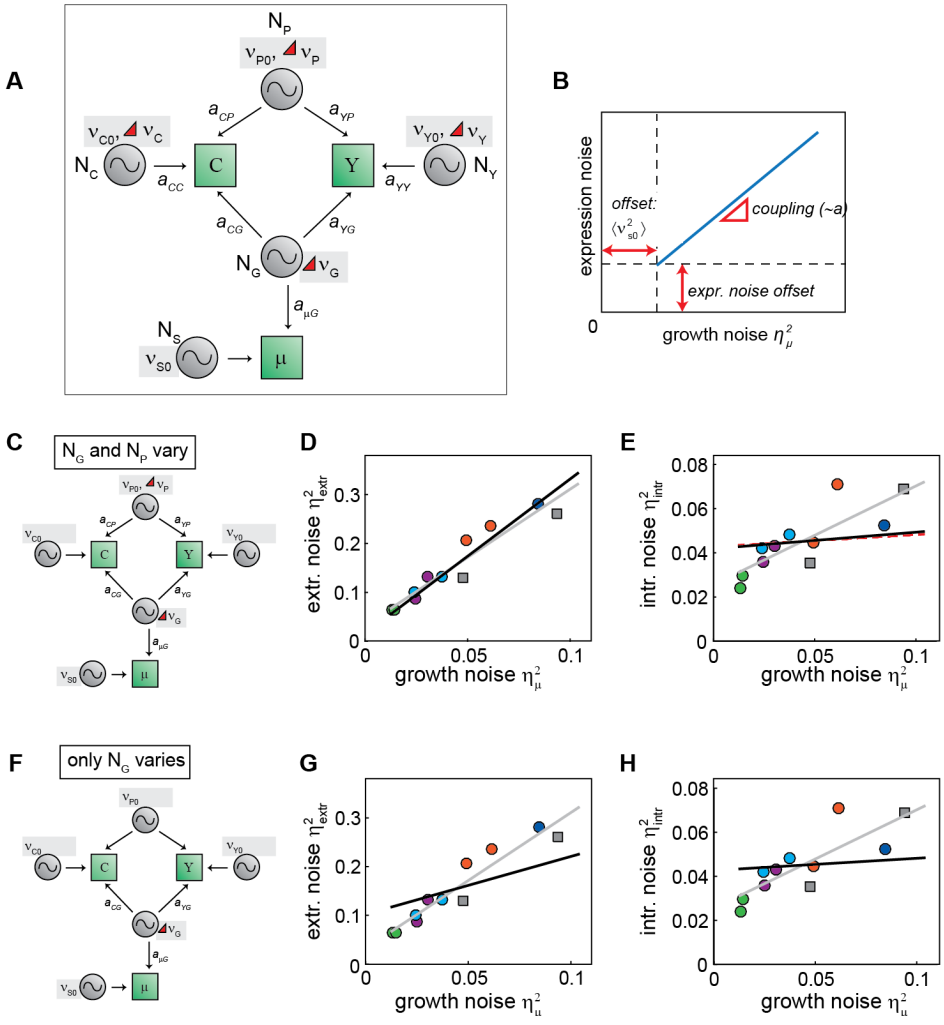


Figure 4.9: Model details and minimal model complexity. (**A**) Linear noise model with all relevant parameters. Red triangle indicates variation in intensity. Transmission efficiencies refer to the intensity-varying noise component. N_P is for example the sum of ν_{P0} and $a_{XP}\nu_P$ (a_{XP} is the noise transmission coefficient to observable Z) (**B**) Role of model parameters for the example line 'expression noise versus growth noise'. The slope is determined by the coupling coefficients, which couple to the intensity-varying common noise source(s). x-axis and y-axis offset are determined by the constant noise floors for growth ($\langle\nu^2_{S0}\rangle$) and expression (e.g. $\langle\nu^2_{P0}\rangle$ plus $\langle\nu^2_{Y0}\rangle$). (**C**) -(**H**) Simpler models with less varying noise sources fail to explain the data. Model parameters in (**C**) -(**E**) were fixed by fitting $Cov(Y,\mu)$, $Cov(C,\mu)$, η^2_Y and η^2_C against η^2_μ . In (**E**) -(**H**) only the covariances needed to be fitted to fix the model. Black lines: Predictions of noise scaling obtained from the simpler models, when parameters were fixed as described here. Gray lines: fits of full model from panel (**A**) and main text. Dashed line in (**E**) is equal to black line in (**H**) . Both simpler models fail to describe intrinsic noise, and the model of (**F**) also does not capture extrinsic noise.

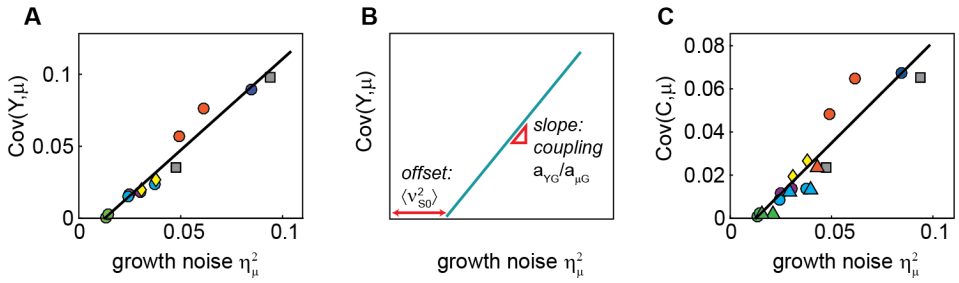


Figure 4.10: Linear dependence of covariances on growth noise (**A**) $Cov(Y,\mu)$ in dependence of η_μ^2 . (**B**) Relation of model parameters to the line in panel (**A**) . (**C**) $Cov(C,\mu)$ in dependence of η_μ^2 . Here, data for low expression is plotted as well (triangles). Color code and symbols as in Fig. 4.2.

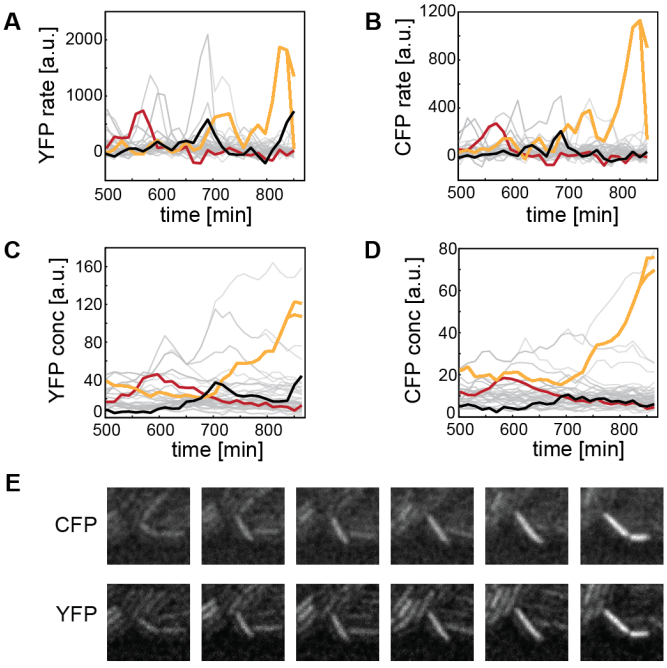


Figure 4.11: Simultaneous expression bursts at low expression. (**A**) , (**B**) At low average production, protein production rate can burst simultaneously for both reporters. Growth rate: 0.6dbl/hr. Highlighted lineages: Examples of simultaneous bursting. (**C**) , (**D**) The increase in production rate is followed by an increase in protein concentration. (**E**) Fluorescence images of the yellow lineage, covering a time range from 715min to 850min.

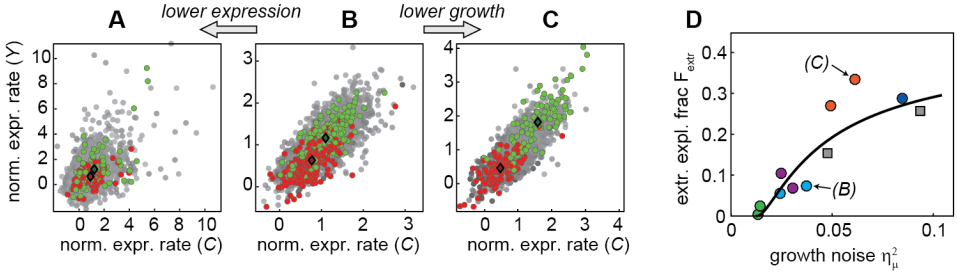


Figure 4.12: Coupling of growth fluctuations to extrinsic and intrinsic fluctuations in different environments. (**A**) -(**C**) Similar to Fig. 4.4B, but for three different environments. (**B**) is identical to Fig. 4.4B. Subsets with fastest (slowest) growth rates are colored in green (red), and their averages are marked as diamonds. Datasets and environments are the same as in Fig. 4.6. (**D**) Explained fraction of extrinsic noise in dependence of growth noise. Black curve is calculated from the line fits of the model (eq. 4.23). Numbers refer to the figures of corresponding scatter plots. Color code and symbols as in Fig. 4.2. Data points for growth on acetate (orange) deviate from the fit, which might be due to additional medium or mean growth rate dependent effects.

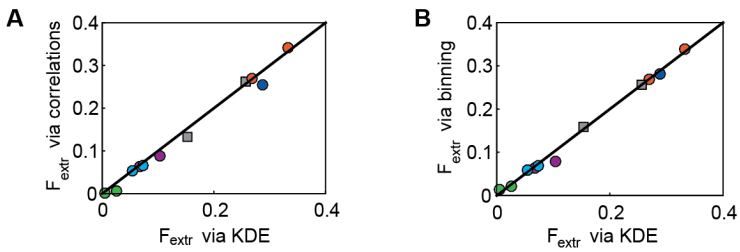


Figure 4.13: Determination of $F_{extr,E}$ with different methods. (**A**) The linearized $F_{extr,E}$, which is a function of the correlations between Y, C and μ (eq. 4.5), is plotted against the general $F_{extr,E}$ (obtained by using KDE). (**B**) $F_{extr,E}$ was determined by binning and by using KDE (described in Methods 4.4). Color code and symbols as in Fig. 4.2. Black lines are the identity function ($y = x$).

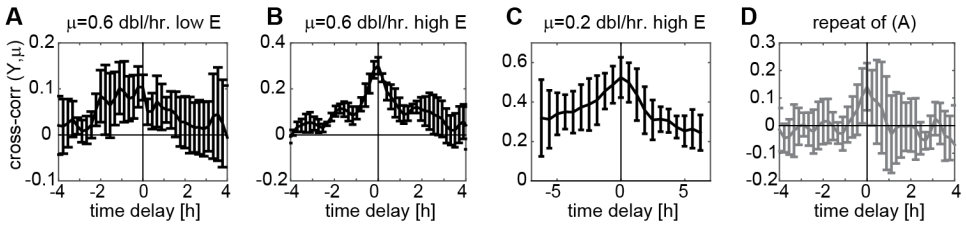


Figure 4.14: Cross-correlation between expression and growth rate. Displayed are cross-correlations between YFP production rate and growth rate for three different environments. (**A**) M9 + 0.2mM uracil + 0.1% maltose (low expression), same data as Fig. 4.6A,D and Fig. 4.12A. (**B**) M9 + 0.2mM uracil + 0.1% maltose + 200μM Iptg (high expression), same data as Fig. 4.6B,E and Fig. 4.12B. (**C**) M9 + 0.2mM uracil + 0.18% acetate + 200μM Iptg (high expression), same data as Fig. 4.6C,F and Fig. 4.12C. (**D**) repeat experiment of (**A**) . All cross-correlations peak at a delay of zero. The signal in (**A**) is difficult to interpret but a repeat experiment (**D**) supports the symmetry of the cross-correlation. Cross-correlations of CFP expression rate and growth rate are similar. Error bars are standard deviations, obtained by splitting the microcolony into 4 subgroups and calculating the cross-correlation for each subset (chapter 2.3.6).

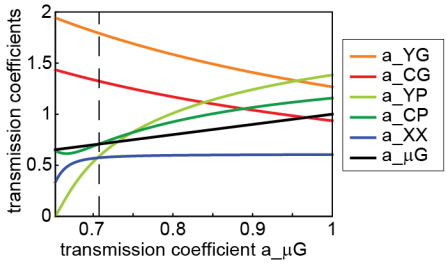


Figure 4.15: Varying the intensity of the growth specific noise source. Transmission coefficients towards growth obey the relation $a_{\mu S}^2 + a_{\mu G}^2 = 1$ (eq. 4.21). $a_{\mu G}$=1 corresponds to the model of the main text, that is, constant growth-specific noise. The dashed line corresponds to equal contribution of both noise sources towards growth fluctuations. a_{XX} refers to a_{CC} and a_{YY} (see model fitting). a_{CP} and a_{YP} probably diverge from each other because of the assumption that both reporters couple equally strong to intrinsic noise (a_{XX}).

4.5.4. ADDITIONAL TABLES

Medium	Expression	Data points	$\langle\mu\rangle$ (dbl/hr)	η^2_μ	η^2_C	η^2_Y	η^2_{extr}	η^2_{intr}	$R(Y,C)$	$R(C,\mu)$	$R(Y,\mu)$
strain MG22 (P_{lac} -*yfp*, P_{lac} -*cfp*)											
A	-	1262	0.34	0.21^2	0.84^2	1.05^2	0.60	0.30	0.68	0.14	0.11
A	+	1900	0.20	0.25^2	0.53^2	0.58^2	0.24	0.071	0.77	0.50	0.53
A	+	5154	0.20	0.22^2	0.47^2	0.53^2	0.21	0.045	0.83	0.46	0.48
M	-	1683	0.63	0.20^2	1.35^2	1.28^2	0.77	0.96	0.45	0.050	0.14
M	-	1947	0.64	0.17^2	1.11^2	1.26^2	0.61	0.81	0.43	0.063	0.10
M	+	3512	0.62	0.16^2	0.37^2	0.39^2	0.10	0.042	0.71	0.15	0.25
M	+	4225	0.62	0.19^2	0.42^2	0.43^2	0.13	0.049	0.73	0.17	0.28
M, h.i.	+	2831	0.54	0.29^2	0.52^2	0.63^2	0.28	0.053	0.86	0.45	0.49
L	+	3359	0.54	0.17^2	0.40^2	0.44^2	0.13	0.043	0.76	0.20	0.24
L	+	3763	0.66	0.16^2	0.35^2	0.35^2	0.088	0.036	0.71	0.21	0.30
RDM	-	929	1.85	0.13^2	0.79^2	0.93^2	0.27	0.48	0.37	0.019	0.065
RDM	-	1011	1.85	0.14^2	0.91^2	0.89^2	0.29	0.52	0.36	0.012	0.10
RDM	+	1117	1.77	0.12^2	0.31^2	0.29^2	0.064	0.024	0.72	0.023	0.009
RDM	+	806	1.79	0.12^2	0.32^2	0.29^2	0.064	0.030	0.69	0.055	0.068
strain M22 (P_λ -*yfp*, P_λ -*cfp*)											
M	+	1500	0.52	0.31^2	0.55^2	0.60^2	0.26	0.069	0.79	0.39	0.53
M	+	1567	0.82	0.22^2	0.41^2	0.41^2	0.13	0.035	0.79	0.26	0.40
strain ASC659 (P_{N25} -*gfp*)					η^2_G					$R(G,\mu)$	
L	+	5063	0.89	0.18^2	0.44^2	-	-	-	-	0.25	-
L	+	4627	0.88	0.20^2	0.47^2	-	-	-	-	0.29	-

Table 4.1: Summary of all data. Abbreviations used for growth media: A: M9 + 0.18% acetate, M: M9 + 0.1% maltose, L: M9 + 0.1% lactose. RDM: rich defined medium + 0.2% glycerol. h.i.=high toxic illumination. All M9 media were supplemented with 0.2mM uracil. Symbols used for expression level: "+": high expression, either through induction with 200µM Iptg (strain MG22) or constitutive expression (other strains). "-": basal expression, no induction. Data points: summed over all cells and all time points of each cell within the selected time window. The dark-gray shaded dataset (M+) was used for Fig. 4.1D and 4.4B. All three gray data sets were used in Fig. 4.6, 4.12, 4.14.

Transmission coefficients					
$a_{\mu G}$	a_{YG}	a_{CG}	a_{YP}	a_{CP}	$a_{YY} = a_{CC} = a_{ZZ}$
set to =1	1.27	0.94	1.38	1.16	0.60

Offset noise			
$\langle v_{S0}^2 \rangle$	$\langle v_{P0}^2 \rangle$	$\langle v_{Y0}^2 \rangle$	$\langle v_{C0}^2 \rangle$
0.0127	0.068	0.026	0.036

Table 4.2: Values of fitted model parameters. For description of model and fitting, see section 4.5.1.

5

GENERATION AND FILTERING OF GENE EXPRESSION NOISE BY THE BACTERIAL CELL CYCLE

Gene expression within cells is known to fluctuate stochastically in time. However, the origins of gene expression noise remain incompletely understood. The bacterial cell cycle has been suggested as one source. It involves chromosome replication, exponential volume growth, and various other changes in cellular composition. Elucidating how these factors give rise to expression variations is important to models of cellular homeostasis, fidelity of signal transmission, and cell-fate decisions.

Using single-cell time-lapse microscopy, we measure cellular growth, as well as fluctuations in the expression rate of a fluorescent protein and its concentration. We find that within the population, the mean expression rate doubles throughout the cell cycle with a characteristic cell-cycle phase dependent shape which is different for slow and fast growth rates. At low growth rate, we find the mean expression rate is initially flat, and then rises approximately linearly by a factor two until the end of the cell cycle. The mean concentration fluctuates at low amplitude with sinusoidal-like dependence on cell cycle phase. Traces of individual cells are consistent with a sudden two-fold increase in expression rate, together with other non-cell-cycle noise. A model is used to relate the findings and to explain how the cell-cycle induced variations depend on the chromosomal position.

We find that the bacterial cell cycle contribution to expression noise consists of two parts: a deterministic oscillation in synchrony with the cell cycle, and a stochastic component due to variable timing of gene replication. Together they cause half of the expression rate noise. Concentration fluctuations are partially suppressed by a noise canceling mechanism that involves the exponential growth of cellular volume.

The contents of this chapter have been published as Walker, N., Nghe, P., and Tans, S. J. *Generation and filtering of gene expression noise by the bacterial cell cycle.* BMC Biology **14**, 1 (2016). [132]

5.1. INTRODUCTION

Single-cell experiments have shown gene expression to fluctuate randomly under constant conditions [6–8, 16, 27, 113, 133], which can have key consequences for the fidelity of signal propagation [26], cell fate decisions [12, 15], and fitness [9, 11, 16, 30, 114, 134, 135]. Noise in gene expression is often quantified by the observed cell-to-cell variability in the production rate or concentration of a protein when observing many cells in an isogenic population [5, 7]. Fluctuations in gene expression can be caused by many local and global factors such as random binding events of RNA polymerase [120], fluctuating concentration of ribosomes or availability of amino acids [16, 19]. The cell cycle has been suggested as a general source of gene expression noise [5, 19]. That is, in a snapshot of a population two cells can differ in protein production rate or concentration because they are in different phases of their cell cycle. Alternatively, two cells at the same cell-cycle phase can differ because of cell-cycle independent effects. The key aim of this study is to quantify and disentangle these effects in *E. coli* , and to mechanistically understand cell-cycle contributions.

Eukaryotes exhibit distinct cell-cycle phases that display different levels of growth activity and of DNA replication, which in turn can result in varying expression levels as the cell-cycle progresses. Single-cell investigations of *Saccharomyces cerevisiae* have indeed shown quasi-periodic fluctuations of protein expression rates [119] and concentrations [136] in synchrony with the cell cycle. The prokaryotic cell cycle does not display such distinct replication and growth phases. *Escherichia coli* for instance, grows and replicates DNA continuously throughout its cell cycle, though for slow growth there are periods without replication activity [110, 137]. Expression activity can be dependent on the cell cycle nonetheless, for instance because the replication of a gene may double the transcriptional activity at a specific moment in time, as suggested by recent single-cell studies [5, 138, 139]. That doubling would then in turn affect enzyme concentration and could cause quasi-periodic fluctuations. But at the same time, cells may exploit specific regulatory mechanisms to filter such perturbations [140, 141]. Direct experimental investigations of the impact of the bacterial cell cycle on expression variability are lacking. Elucidating this question is important to understand the origins of gene expression noise, modeling of genetic circuits, and resulting impact on growth variability [56] as well as other forms of cellular heterogeneity [15].

To address these questions, we followed a single-cell approach. We imaged *E. coli* cells as they grew into microcolonies and measured gene expression as fluorescence signal of chromosomally encoded fluorescent proteins. As we show here, understanding the temporal dynamics requires detailed information on cellular volume increase in time, as protein concentrations are affected both by the time-dependent expression and by dilution. Hence we accurately determined protein expression and cell size at sub-cell-cycle resolution. We further developed a model to predict the cell-cycle dependence and amplitude of these quasi-periodic fluctuations in expression rate and concentration. The model predicted their dependence on chromosomal position, which we tested with genetic constructs.

5.2. RESULTS AND DISCUSSION

5.2.1. THE PROTEIN PRODUCTION RATE FLUCTUATES QUASI-PERIODICALLY.

To measure the effect of the cell cycle on protein expression we first determined protein production rate, as quantified by the time derivative of the total cellular fluorescence (Materials and Methods, section 5.4). Taking the data for all cells with completed cell-cycle (n=393) over all cell cycle phases, the protein expression rate displayed a total noise intensity (defined as standard deviation divided by the mean), of 0.48 [5]. When plotting the production rate versus cell cycle phase ϕ (where 0 is cell birth and 1 is cell division) and averaging over all cells (Fig. 5.1A), it displayed the following trend: it was approximately constant in the first half, after which it rose to about two-fold at the end of the cycle (Fig. 5.1B, 5.6). An initially constant rate and two-fold increase is consistent with the known chromosome replication pattern for the observed mean growth rate (0.6dbl/hr): a single chromosome copy in the first period of the cell cycle, after which replication occurs in the second period that produces two copies [142]. Each chromosome copy then yields a fixed expression rate. This is not unreasonable, as other components required for expression such as RNA polymerases and ribosomes also double throughout the cell cycle. At faster growth, replication occurs throughout the cell cycle for multiple nested chromosome copies [143]. Consistently, we found that the production rate was not flat initially but instead rose continuously throughout the cell cycle, when growing on a different medium that supported a higher mean growth rate of 1.8dbl/hr (see Fig. 5.7). The total increase remained two-fold, in agreement with an expected doubling of the number of gene copies. Overall, these data indicate that the mean protein expression rate is likely proportional to the gene copy number and hence doubles during chromosome replication. This variation is more continuous at high growth rate, because of the nested replication and overall higher gene copy numbers.

5.2.2. DETERMINISTIC CELL-CYCLE VARIATIONS CONTRIBUTE TO EXPRESSION NOISE.

To quantify the contribution of the mean cell-cycle fluctuations (Fig. 5.1B) to protein production noise, we split the single-cell production rate (which is distinct from the protein concentration) $p(\phi,\mathbf{x})$ into the population averaged rate $\overline{p_c}(\phi)$ and individual deviations $\delta p(\phi,\mathbf{x})$, which together capture all cell-to-cell variability (Fig. 5.1A,B):

$$p(\phi,\mathbf{x}) = \overline{p_c}(\phi) + \delta p(\phi,\mathbf{x}) \tag{5.1}$$

Here, ϕ denotes the cell-cycle phase and $\mathbf{x}$ all other causes of cell-to-cell variability. c refers to cell-cycle dependence, which here is redundant because it is implied by the ϕ-dependence but used for notation consistency. $\overline{p_c}(\phi)$ can be estimated by the curve in figure 5.1B, and subtracted from individual traces to obtain an estimate for $\delta p(\phi,\mathbf{x})$. The noise intensity caused by the deterministic cell-cycle fluctuation $\overline{p_c}(\phi)$ is 0.26, which was obtained by considering the phase ϕ as a random variable and then calculating the variance of the trace. Noise of the individual expression traces $\delta p(\phi,\mathbf{x})$, averaged over all cells and ϕ, was 0.42 (see Fig. 5.8A). These values are consistent with a scenario in which population mean trace $\overline{p_c}(\phi)$ and deviation traces $\delta p(\phi,\mathbf{x})$ are independent and thus

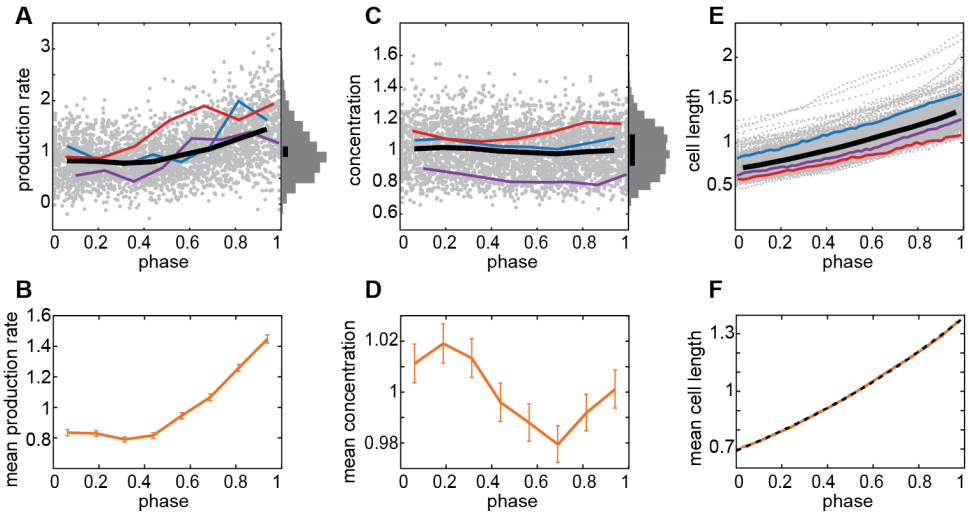


Figure 5.1: Dependence of protein production rate (A,B), protein concentration (C,D) and cell length (E,F) on cell-cycle phase. Values are normalized by the average. (**A**) , (**C**) , (**E**) Data for 393 cells (gray) with three example traces and the binned colony average (thick black line). A total frequency histogram is plotted and the black bar indicates a width of 0.2 times the population mean. (**B**) , (**D**) , (**F**) Phase-dependence of the binned data. In (**F**) an exponential function (black dashed line) is fitted to the averaged cell length. Error bars are obtained by bootstrapping. For cell length, error bars are plotted but are smaller than the line thickness. Growth was on minimal medium supporting a growth rate of 0.6dbl/hr.

their variances (squared noise) can be added up: $0.48^2 \approx 0.26^2 + 0.42^2$. This population-average cell-cycle contribution towards production rate noise does not include cell-cycle stochasticity of individual cells and we will consider that below.

5.2.3. CONCENTRATION FLUCTUATIONS ARE BUFFERED BY DILUTION.

Fluctuating production rates can cause noise in the protein concentration. To determine the latter, we quantified the mean fluorescence per unit area of the cell. The noise intensity of 0.15 (0.10 for fast growth), which was obtained by taking the data of all cells and at all cell-cycle phases, was consistent with previous reports [7]. After ordering by cell cycle phase and averaging (Fig. 5.1C), the concentration also showed systematic variations (Fig. 5.1D, 5.6): it increased slightly right after cell birth, then decreased and finally rose again. The amplitude of these variations was 4% of the mean. These low values (see also Fig. 5.8B) and the initial increase seemed inconsistent with the large amplitude of the production rate fluctuations and its initially constant value (Fig. 5.1B) [139].

To get a more intuitive understanding of these differences, we formulated a minimal cell-cycle model based on the measured cell-cycle dependency of the production rate (Fig. 5.1B, 5.2A). The concentration cannot be determined by simply integrating the production rate, as this would ignore dilution due to volume growth. To quantify the volume growth, we determined for each cell its length and its dependence on the cell-

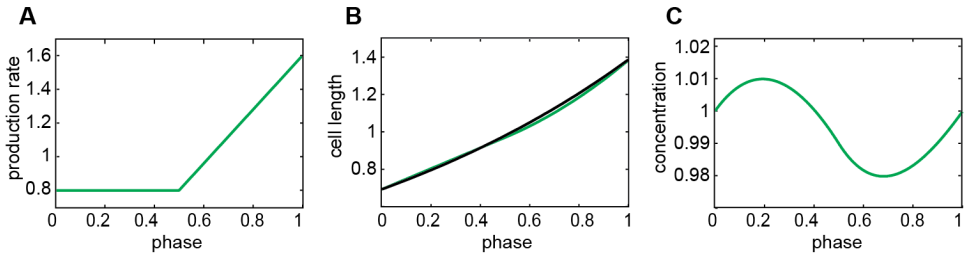


Figure 5.2: Model for cell-cycle dependence of protein concentration. (**A**) Average protein production rate normalized by the mean. (**B**) Exponential length increase (black). Population average protein production rate integrated in time, or the population average total fluorescence (green). (**C**) Determined cellular protein concentration, given by the green signal divided by the black line in panel (**B**) .

cycle phase (Fig. 5.1E, Materials and Methods) [56]. The population mean cell length $\overline{L}(\phi)$ was well described by an exponential function (see Fig. 5.1F) [144–146], and not by bi-linear or linear functions (Fig. 5.9), as suggested previously [147–150]. Therefore, an exponential function for cell size was used as input for the minimal model (Fig. 5.2B). With a mean protein production $\overline{p}(\phi)$ at phase ϕ (Fig. 5.2A), the concentration $\overline{E}(\phi)$ can then be written as:

$$\overline{E}(\phi) = \frac{\left(\overline{F_0} + \int_0^{\phi} \overline{p}(\phi')d\phi'\right)}{\overline{L}(\phi)} \quad , \tag{5.2}$$

where $\overline{F_0}$ is the total amount of protein at cell birth.

By design, $\overline{E}(\phi)$ (Fig. 5.2C) reproduced the measured data (Fig. 5.1D), and provided an intuitive explanation for the observed functional form. The production rate is low early in the cell-cycle but the dilution rate is even lower, resulting in an increasing concentration. For intermediate phases the dilution rate exceeds the production rate, and near the end the dilution rate is smaller again. This also explains why concentration fluctuations are small: the functional form of the total fluorescence (as a function of the cell cycle phase) is almost identical to that of the volume (Fig. 5.2B).

5.2.4. STOCHASTIC REPLICATION TIMING CONTRIBUTES TO EXPRESSION NOISE.

The single cell data also suggested that stochasticity in replication timing is a source of protein production noise, which is supported by previous studies [110, 151] (Fig. 5.1A, thin lines). In other words, $\delta p(\phi, \mathbf{x})$ would be the sum of fluctuations caused by cell cycle stochasticity $\delta p_c(\phi, \nu)$ and of fluctuations $\delta p_{nc}(\mathbf{x})$ unrelated to the cell cycle (Fig. 5.3A). Here, ν is the cell-cycle phase at which the gene of interest is replicated and ν varies from cell to cell. Thus, the sum of $\delta p_c(\phi, \nu)$ and the population-average $\overline{p_c}(\phi)$ yield all the fluctuations $p_c(\phi, \nu)$ caused by the cell cycle. To determine the stochastic contribution of the cell cycle to the expression noise, one needs to quantify $\delta p_c(\phi, \nu)$. However, it is not trivial to distinguish $\delta p_c(\phi, \nu)$ from the other stochastic, non-cell-cycle variations in the experimental single-cell traces.

To overcome this problem, we started with $p_c(\phi,\nu)$ and followed a variance decomposition approach using the law of total variance [123, 124]. The variance of the full cell-cycle fluctuations can be decomposed as follows:

$$Var\left(p_c(\phi,\nu)\right) = \left\langle Var\left(p_c(\phi,\nu)|\phi\right)\right\rangle + Var\left(\left\langle p_c(\phi,\nu)|\phi\right\rangle\right) \tag{5.3}$$

Here, angular brackets denote averaging, and the notations $Var(...|\phi)$ and $\langle ...|\phi\rangle$ indicate respectively the variance and the average for a given phase ϕ (conditioned on ϕ). In the second term, the brackets thus indicate an averaging over the stochastic variable ν, which yields $\overline{p_c}(\phi)$. Next, the variance is taken. This variance was in fact calculated previously, and found to be $(0.26)^2$ (see Fig. 5.1B). Thus, the second term indicates the

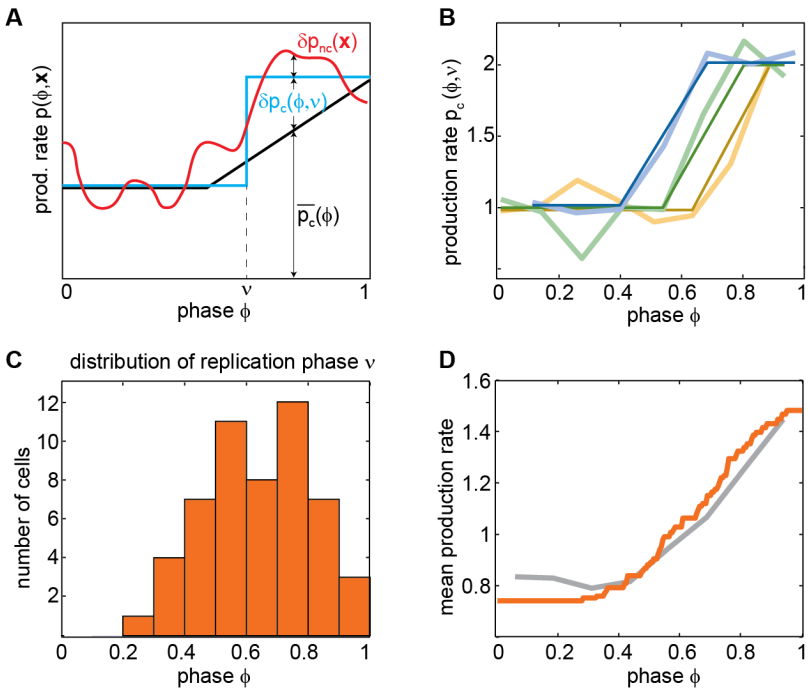


Figure 5.3: Production rates of single cells. (**A**) Description of variables used for noise decomposition. The protein production rate $p(\phi,\mathbf{x})$ (red line) is the sum of three contributions: 1) the population-average cell-cycle fluctuations $\overline{p_c}(\phi)$ (black line), 2) the contribution due to stochastic replication timing (difference between blue and black line, $\delta p_c(\phi,\nu)$), and 3) stochasticity resulting from other, unknown, noise sources (difference between red and blue line, $\delta p_{nc}(\mathbf{x})$). The sum of $\delta p_c(\phi,\nu)$ and $\delta p_{nc}(\mathbf{x})$ represents all of the stochastic contributions $\delta p(\phi,\mathbf{x})$. The phase at which replication occurs is denoted by ν. (**B**) Experimental traces of three different cells (thick lines) and fitted step functions (thin lines). See appendix (section 5.5) for definition of step function. Initial value was set to 1 and data is slightly vertically shifted for clarity. (**C**) Histogram of ν. Data is from 53 cells in which a step-function could be discerned from the rest of the noise (14% of the traces). (**D**) Comparison of experimental average production rate curve (gray line) and mean of ideal step functions (orange line).

deterministic contribution to the cell cycle induced noise.

In the first term, the variance of $p_c(\phi, v)$ is determined conditionally on ϕ, and then averaged. This term thus denotes the stochastic contribution to the cell cycle induced noise. The data does not directly provide an estimate of this variance, because the cell-cycle induced noise and noise from other sources are confounded in the measured single-cell traces of the production rate (Fig. 5.1A). Indeed, in these traces, other noise sources such as metabolism [56] and fluctuating transcription factors [7] are substantial and can mask the quick two-fold increase expected from gene replication events. However, in a subset of traces the two-fold increase was clear (Fig. 5.3B,C, Materials and Methods). Fitting each of these traces with a step-function (Fig. 5.10) provided a distribution of the step-moment, v. We obtained a wide distribution for v with a mean 0.64 and a standard deviation of 0.17 (Fig. 5.3C). To check whether this distribution was consistent with the full dataset, we compared the average of the fitted step-functions to the average of all measured traces ($\overline{p_c}(\phi)$, Fig. 5.1B), and found that they were similar (Fig. 5.3D). These findings suggested that gene duplication events with stochastic timing in individual cells underlie the smooth shape of the population average production rate (Fig. 5.1B).

The distribution of v (Fig. 5.3C) now allowed us to estimate the first term in equation 5.3, by first determining the variance of the step-functions at fixed phase, and then averaging over all phases (Fig. 5.11A). We obtained a value of $(0.23)^2$ for this stochastic contribution of the cell-cycle to expression noise, which is comparable in magnitude to the deterministic contribution denoted by the second term ($(0.26)^2$, Table 5.1). Thus, variability in initiation timing contributes substantially to the cell cycle induced noise. The deterministic and stochastic contributions together ($p_c(\phi, v)$) thus caused a variance of $(0.23)^2 + (0.26)^2 = (0.35)^2$, which is about half (52%) of the protein production variance (Fig. 5.5B, Table 5.1).

To estimate how the protein concentration noise is affected by the cell-cycle, we computed the concentration traces resulting from the step-like production rate functions (Fig. 5.11A). For each $p_c(\phi, v)$ of the set (Fig. 5.3C) the corresponding concentration curve was computed, using that proteins are diluted due to volume growth (Fig. 5.11B). We found that the cell-cycle fluctuations contributed less than 1.5% to the variance in protein concentration (Fig. 5.11B and Fig. 5.5B). Note that one can distinguish contributions from the population average trend (Fig. 5.1D) and the stochastic deviations around it due to variability in replication timing (less than 1% contribution each, Table 5.1).

5.2.5. LOCATION ON THE CHROMOSOME AFFECTS EXPRESSION FLUCTUATIONS.

Chromosome replication is initiated at the origin of replication (*oriC*) from which two replication forks then progress simultaneously and bi-directionally along the two strands of DNA [152]. This raises two expectations: First, genes located at opposite sides but at the same distance from *oriC* should be duplicated at the same time and thus show the same cell-cycle dependence of protein production and concentration. Second, if one gene is located upstream of the other, the increase in protein production should occur earlier. To test the first prediction, we investigated a *cfp* gene positioned symmetrical to the *yfp* gene studied so far, at the opposite strand at the same distance from *oriC* (Materials and Methods, Fig. 5.4A inset). We indeed found that both reporters dis-

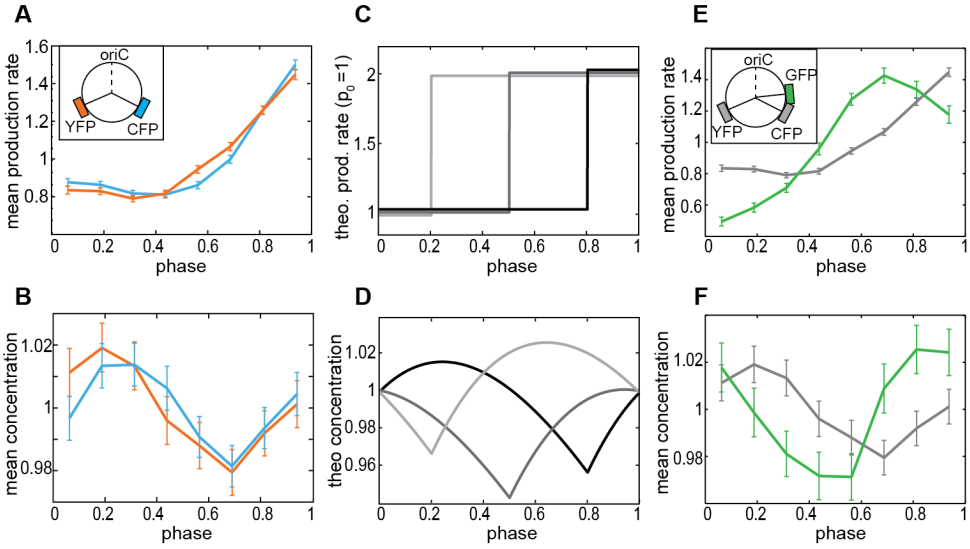


Figure 5.4: Influence of chromosomal position. (**A**) , (**B**) Production rate and concentration for genes at equidistant and symmetric positions with respect to the origin of replication. (Inset) Locations of fluorescent genes and origin of replication *oriC* on the DNA. (**C**) , (**D**) Different replication times ((**C**) , light to dark gray) and their predicted influence on the concentration (**D**) . Production rates are slightly vertically shifted for clarity. (**E**) , (**F**) Production rate and concentration of GFP (strain ASC636, green line) compared to YFP (strain MG22, gray line). GFP data is from 296 cells with complete cell cycle that have on average 7 data points/cell cycle. (Inset) Location of *gfp* compared to other fluorescent genes. Error bars are obtained by bootstrapping.

played a similar dependence of production rate and concentration on cell-cycle phase (Fig. 5.4A,B, 5.6).

To change the position we studied a *gfp* gene under P_{lac} control closer to *oriC* than *yfp* or *cfp* (Materials and Methods, Fig. 5.4E). As expected from the earlier replication, the GFP production rate indeed increased earlier than the previous YFP signal (Fig. 5.4E). It started comparatively low, then increased more than two-fold and subsequently decreased again to end at twice the initial rate (Fig. 5.4E). The cause of the high fold-change and decrease is unknown, but changes in chromosome structure or transient improvement in competition for RNA polymerases for this promoter (two binding sites at the two replicated genes) could play a role. As predicted by the model (Fig. 5.4C,D), the dip in GFP concentration occurred earlier and the initial increase disappeared (Fig. 5.4F). The magnitude of fluctuations remained at around 4%. Overall, these data show that gene position on the chromosome affects cell cycle related noise.

5.3. CONCLUSIONS

In summary, we found that the cell cycle can be a major causal factor of observed noise in the rate of gene expression (52%), with the rest coming from other sources such as

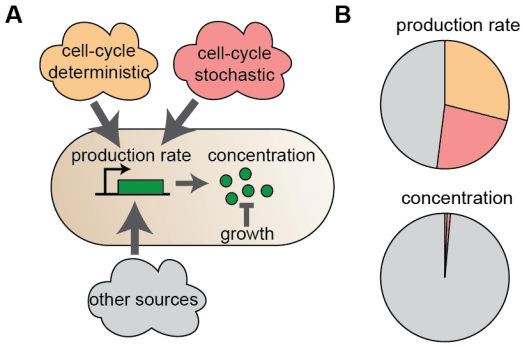


Figure 5.5: Summary of observed contributions to gene expression noise. (**A**) The cell cycle causes fluctuations in the protein production rate, through deterministic and stochastic contributions. Other non-cell-cycle related sources contribute as well. The fluctuations in protein concentration are determined by protein production noise and dilution due to growth. (**B**) Contributions of the different noise sources as described in panel (**A**) , as fractions of the total observed variance in gene expression (Table 5.1).

metabolism [56, 153, 154], transcription factors [26], or expression machinery [120] (see Fig. 5.5A). Within the cell cycle contribution, the data suggests two components: a deterministic mean contribution determined by cell phase (29%), and a stochastic contribution caused by variability in the timing of replication (23%) (Fig. 5.5B). The initially flat production rate suggested gene copy number is the main factor in cell-cycle induced expression rate variations, though alternative explanations cannot be ruled out.

The analysis indicated a noise-canceling mechanism: even sudden two-fold production rate increases caused by replication of the gene are effectively compensated for by a concurrent acceleration of dilution due to exponential growth [140, 141] (Fig. 5.5B). The observed minor effect of the cell cycle on the protein concentration is thus due to a passive homeostasis mechanism that exploits the balance between synthesis and dilution. When proteins are actively degraded, this noise canceling mechanism would be less efficient. We note that a similar, but likely active, balancing between synthesis and dilution was observed in mammalian cells where transcription rate is adjusted to cell size [155, 156]. The homeostatic mechanism we observed does not necessarily act on noise from other sources, such as fluctuations in RNA polymerase availability [120] or transcription factors [8], if they are not synchronized with exponential volume growth. Indeed, concentrations do display significant noise intensities (0.15 for slow growth, 0.10 for fast growth). We note that other canceling mechanisms can act on non-cell-cycle expression noise. For instance, metabolic noise that causes expression noise is partially compensated for by increased growth [56].

Our findings provide insight into how elementary processes such as gene replication events and volume growth can cause and filter noise in bacterial cells. Elucidating the sources of gene expression noise is important to obtaining a bottom-up understanding of cellular heterogeneity, cellular homeostasis and cell-cycle regulation, and to providing input for mathematical models of gene expression networks. Our results indicate

that variance decomposition can be a useful tool in disentangling different noise sources within cells.

5.4. MATERIALS AND METHODS

Experimental and analysis methods were described in detail in chapter 2. Here, we specifically explain additional or deviating methods.

5.4.1. EXPERIMENTS AND IMAGE ANALYSIS

E. coli strain MG22 [7] was used for all experiments unless noted otherwise. This strain is a derivative of MG1655 that contains *yfp* and *cfp* under control of a *lac* promoter, which were inserted into the chromosome at the *intC* and *galK* locus. These two loci are equidistant from the origin of replication, on opposite halves of the circular chromosome. Additionally, we used strain ASC636 in which gene *lacA* of the *lac* operon was replaced by *gfp* (constructed by A. Böhm).

For microscopy experiments we used either M9 minimal medium (main text figures) or rich defined medium (MOPS EZ rich defined medium from Teknova, Fig. 5.7). M9 was supplemented with 0.2mM uracil and 0.1% maltose was added as carbon source, yielding a growth rate of 0.6dbl/hr. To the rich medium we added 0.2% glycerol as carbon source yielding a growth rate of 1.8dbl/hr. In all experiments we also added 200 μM IPTG to fully induce the *lac* promoters.

Cells were grown and prepared for microscopy as described in chapter 2. Movies of growing microcolonies were acquired on the Nikon TE2000-E inverted microscope (chapter 2.2). Imaging frequency was on average 55 phase contrast and 8 fluorescence images per cell cycle and fluorescence illumination time was kept short (YFP: 25ms, CFP: 30ms, GFP: 50ms). Each experiment was performed at least twice (Fig. 5.6).

Images were analyzed offline as described (chapter 2). Cells were segmented and tracked and cell length was determined (cell width is constant: Fig. 5.12, [32], contrary to results in [157]). Fluorescence images were corrected and concentration and production rate was calculated for each cell. To obtain autofluorescence intensity, we measured a non-fluorescent strain (MG1655) with the same illumination settings as standard experiments. Measured signal intensity was 2.5% (CFP) resp. 0.4% (YFP) of the actual concentration signal from fluorescent proteins. Fluorescence signals were not corrected for autofluorescence because autofluorescence was small compared to the signal and did only introduce a constant offset with no effect on our results.

5.4.2. DATA PROCESSING

We analyzed cell cycles within a time window of the experiment that showed a constant population-mean growth rate and mean protein concentration. Population mean growth and concentration were considered constant when they fluctuated less than 5% around the long-term average (for GFP in strain ASC636 10% was used as cutoff criterion). Cells that stopped growing or filamented were removed (less than 15 cells per dataset). The fraction of analyzed cells relative to all cells observed with complete cell cycle was over 86% for MG22 datasets and over 50% for ASC636 datasets. The main conclusions were robust to taking the complete data set of growing non-filamentous cells

with complete cell cycles. Datasets contained between 215 selected cells (large cells in rich medium) to 435 selected cells (minimal medium). If one dataset was used for multiple plots (e.g. Fig. 5.1 & 5.4A,B) the same cells were analyzed.

Traces of production rate (Fig. 5.3) for individual cells were considered "step-traces" when they deviated from a fitted step trace (see also appendix) less than a fixed threshold. To be considered a step-trace, the mean squared deviation of a data point on the trace from the fitted value had to be below 2% of the squared trace-average. Figures and percentages in the main text are determined from one microcolony per strain and growth condition. Results for the repeat experiment are shown in Fig. 5.6.

To determine the average dependence of a signal (e.g. concentration) on cell-cycle phase, we binned the signal according to phase and averaged it within each bin (see Fig. 5.1B,D,F). Error bars are standard errors of the mean from a resampled distribution of the signal, obtained by bootstrapping from the experimental data for each bin.

The contribution of a specific noise source X (for example deterministic cell-cycle variations) to total protein production noise was calculated by using the additivity of variances for independent variables. The production rate p was for example written as sum $p = X + Y$ with Y being the unknown, not-measured, fluctuations (such as $\delta p(\phi, \mathbf{x})$, see also main text). Then, $Var(p) = Var(X) + Var(Y)$, and the fraction of the variance in p which is caused by X is $Var(X)/Var(p)$. We normalized variables such as p by their mean so that the squared noise is identical to the variance. For example, the contribution of $\overline{p_c}(\phi)$ to $p(\phi, \mathbf{x})$ is $0.26^2/0.48^2$. For protein concentration the calculation is identical.

5.5. APPENDIX

5.5.1. ADDITIONAL METHODS

FITTING BACTERIAL GROWTH WITH AN EXPONENTIAL, LINEAR AND BILINEAR FUNCTION

Since the precise law of bacterial growth is important for cell-cycle fluctuations of protein concentration, we assessed fits of different growth functions to the binned cellular length data. We examined exponential [144–146], linear [149, 150] and bilinear [147, 148] fit functions for growth. Goodness of fit was evaluated via the mean square error MSE which is the average squared deviation of a data point from the fitted curve:

$$MSE = \frac{1}{n} \sum \epsilon_i^2 \,. \tag{5.4}$$

Here, ϵ_i is the distance of a data point from the fitted value, n is the number of data points and the sum runs over all data points. Average length of each cell is normalized to one. More elaborate measures which introduce a penalty for increasing number of fitting parameters (four vs. two free parameters for bilinear vs. exponential fit) gave essentially the same result.

The best fit to the population average function of cellular growth was an exponential function (Fig. 5.9, $MSE = 7.4 \cdot 10^{-6}$). A bilinear fit was slightly worse ($MSE = 31.1 \cdot 10^{-6}$), while cellular growth clearly does not follow a linear growth law ($MSE = 316.8 \cdot 10^{-6}$). For several single-cell individual traces, a bilinear fit was better than an exponential fit.

Since our analysis holds for exponential and bilinear growth functions and an exponential growth law may be more physiologically relevant we chose an exponential function.

FIT OF STEP FUNCTIONS TO PRODUCTION RATE TRACES

We determined production rate as the slope of a linear fit to three subsequent total-fluorescence data points. This discrete derivative smoothens a theoretical perfect stepwise increase of production rate and makes the increase appear more gradual. In order to fit step functions to production rate traces (Fig. 5.3 in the main text), we needed to know what a perfect and immediate step function would look like after our data processing. We used this result to fit the single cell traces in the main text.

Let $d_1, d_2, \ldots, d_n$ be the phase points for which fluorescence data exists (n fluorescent data points for one cell cycle, $0 \leq d_i \leq 1$) (Fig. 5.10A,B). All data points have an equal phase difference of $\Delta d = d_{(i+1)} - d_i$. The sudden increase of production rate is supposed to occur at phase x which is in between d_i and $d_{(i+1)}$ but may also be on either boundary of this interval. Δx is the phase between the step time and the next data point $d_{(i+1)}$. The initial production rate is p_0 and doubles at x to be $2p_0$.

For calculation of production rate at phase point d_k the total fluorescence data $F_{(k-1)}$, F_k and $F_{(k+1)}$ at phase $d_{(k-1)}$, d_k and $d_{(k+1)}$ has to be fitted by a linear function under a least square criterion. For $k < i$, resp. $k > (i+1)$ the fitted production rate is simply p_0, resp. $2p_0$. However, for the two data points close to the transition in production rate, i.e. at phase d_i and $d_{(i+1)}$, a smoothing effect will occur. For calculation of production rate at these points we need the total fluorescence at the following phase points:

phase	total fluorescence	
$d_{(i-1)}$	$F_{(i-1)}$	
d_i	$F_i = F_{(i-1)} + p_0\Delta d$	(5.5)
$d_{(i+1)}$	$F_{(i+1)} = F_{(i-1)} + p_0\Delta d + p_0\Delta d + p_0\Delta x$	
$d_{(i+2)}$	$F_{(i+2)} = F_{(i-1)} + 2p_0\Delta d + p_0\Delta x + 2p_0\Delta d$	

Fitting a linear function to $F_{(i-1)}$, F_i, $F_{(i+1)}$ to obtain the production rate $p(d_i)$, resp. to F_i, $F_{(i+1)}$, $F_{(i+2)}$ to obtain $p(d_{(i+1)})$, we get (see Fig. 5.10C):

$$\begin{aligned} p(d_i) &= p_0 \cdot (1 + 0.5\Delta x/\Delta d) \\ p(d_{(i+1)}) &= p_0 \cdot (1.5 + 0.5\Delta x/\Delta d) \end{aligned} \tag{5.6}$$

Thus, after data processing, we obtain one (Fig. 5.10C, blue and green line) or two (red line) data points with intermediate production rate. Our time resolution is set by the acquisition frequency of fluorescent images and we therefore restricted the potential phases for step-events to these time points (i.e. $\Delta x = 0$). Therefore, we fitted single cell traces with a smoothed step function that contains one intermediate data point at $1.5p_0$ (Fig. 5.10C, blue and green line).

We excluded fitted steps where doubling occurred within the first 25 minutes after cell birth. DNA duplication takes at least 33 minutes in fast growth but around 70 minutes at our slow growth rate [158]. The observed genes are located at least halfway downstream from the origin of replication, so duplication should occur at least 35 minutes

after cell birth. Nested replication forks and hence pre-birth start of replication are not to be expected to play a relevant role at slow growth rates. Hence, we can expect that excluding the first 25 minutes for gene duplication does not introduce a bias.

5.5.2. ADDITIONAL FIGURES AND TABLES

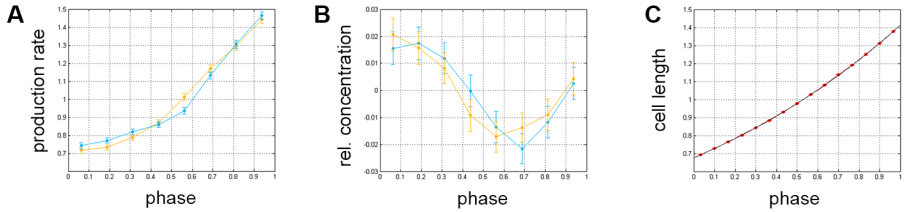


Figure 5.6: Population average cell-cycle dependencies from a replicate experiment. Figures are to be compared to Fig. 5.1 and Fig. 5.4A,B. yellow: YFP (Fig. 5.1), cyan: CFP (Fig. 5.4)

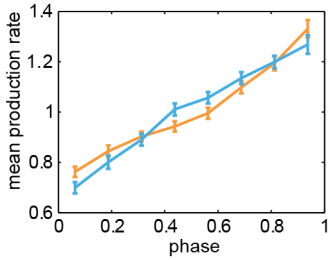


Figure 5.7: Dependence of average YFP (orange line) and CFP (blue line) production rate on cell cycle phase in rich medium. Data is obtained from 215 cells with complete cell cycle, strain is MG22. Error bars are obtained by bootstrapping.

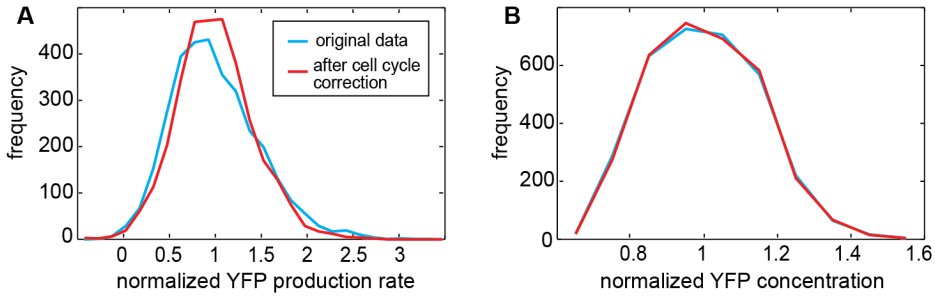


Figure 5.8: Distribution of normalized YFP production rate (A) and concentration (B) . Histograms of data before (blue) and after (red) subtraction of cell-cycle trends are shown. Variation in production rates is reduced while concentration distribution is almost unchanged.

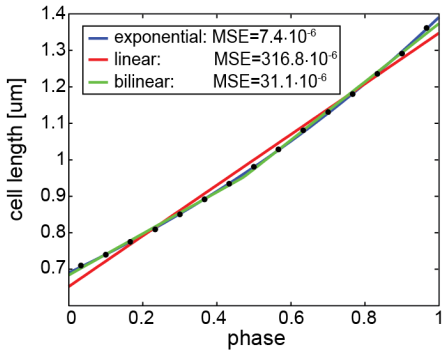


Figure 5.9: Functional shape of cellular growth. Different growth functions are fitted to the binned phase dependence of cellular length (black dots): Tested fit functions are: exponential growth (blue line), linear growth (red line) and bilinear growth (green line). The exponential fit is best and data points deviate the most from a linear fit.

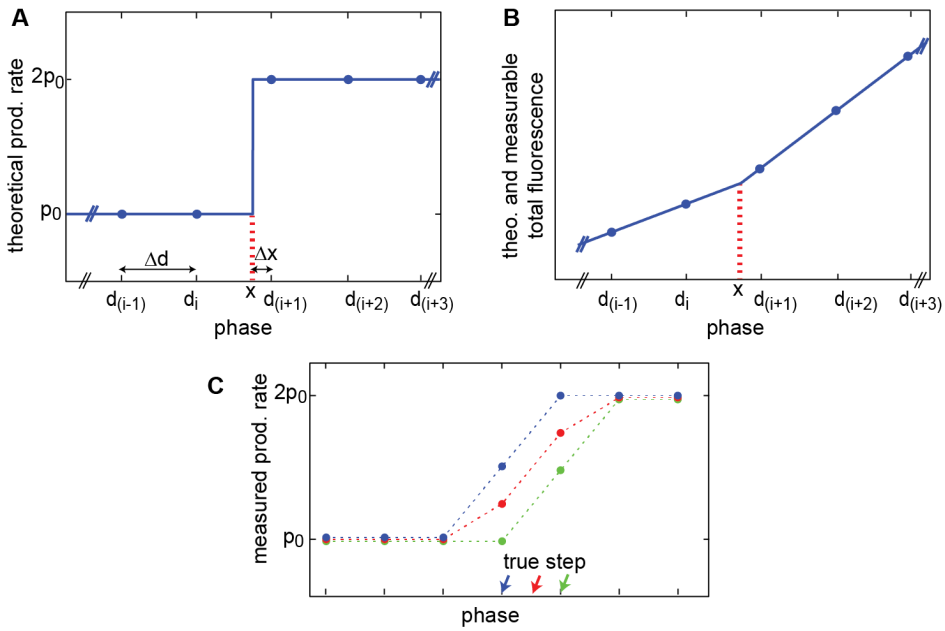


Figure 5.10: Influence of determination method of protein production rate on a step-like increase. (**A**) The theoretical production rate doubles instantaneously (blue line). Dots denote the time points at which production rate will be measured in an experiment, the vertical axis is however not directly measurable due to finite time delays between acquired images. (**B**) Total fluorescence (line) is the integrated production rate. For production rate calculation, total fluorescence is measured at the indicated phase points (dots). (**C**) Inferred production rate from total fluorescence data for an idealized step-like input rate. Production rate is smoothened and contains one or two data points at intermediate rates. The applied fit function in the main text corresponds to the blue and green lines, i.e. when the step occurred at one of the measured time points.

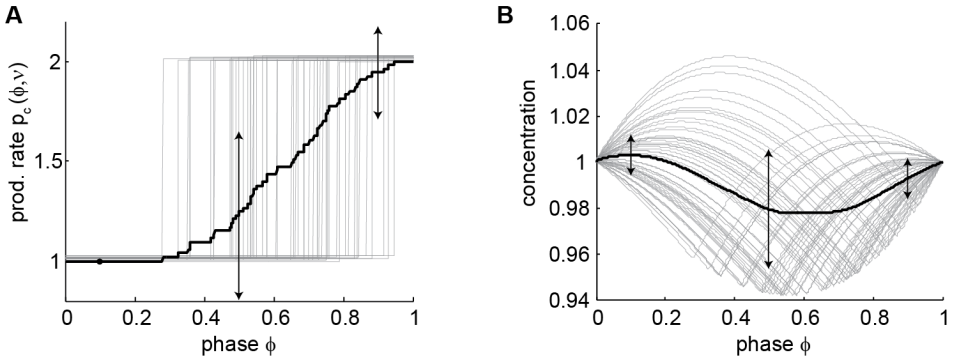


Figure 5.11: Determination of protein noise caused by stochastic replication timing. (**A**) Production rate. Fitted step-like production rate functions, for 53 individual cell cycles (thin gray lines), and the resulting averaged production rate (black line, same as orange line in Fig. 5.3D). Arrows indicate three examples of the conditional variance $Var\left(p_c(\phi,\nu)|\phi\right)$ at three phases (size corresponds to standard deviation of the single-cell traces at that phase, first arrow has length zero). (**B**) Concentration. We estimated the concentration traces (thin gray lines) using the production rate traces from (**A**) and dilution due to volume growth. The thick black line is the average of all single-cell traces. The stochastic and deterministic cell-cycle contributions to the concentration fluctuations were then determined in the same way as for production rate, using variance decomposition. Specifically, the conditional variance (the variance of the protein concentration at a particular phase) is displayed for three example phases (arrows). Resulting noise contributions are found in Table 5.1. Initial values here are normalized to 1.

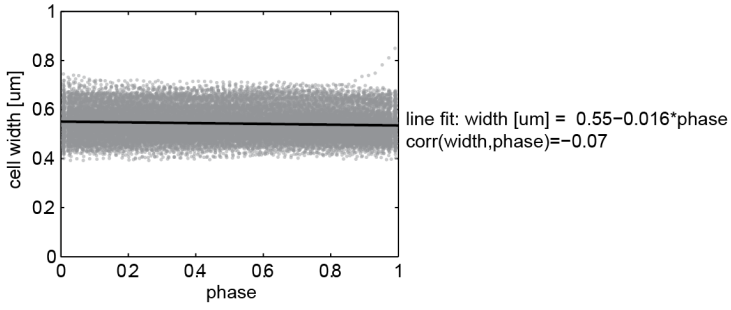


Figure 5.12: Cell width in dependence of cell-cycle phase. The width is found to be independent of cell-cycle phase.

	a) total variance	b) variance of deterministic cell-cycle fluctuations $Var\left(\overline{p_c}(\phi)\right)$	c) variance caused by stochastic replication timing $Var\left(\delta p_c(\phi,\nu)\right)$	total variance due to cell-cycle fluctuations: b)+c)	variance caused by non-cell-cycle effects: a)-b)-c)
production rate	0.48^2	0.26^2	0.23^2	0.35^2	0.33^2
concentration	0.145^2	0.011^2	0.013^2	0.017^2	0.144^2

Table 5.1: Contributions of different components to protein production and concentration noise. Values are given as variances (squared noise) because then individual contributions can be added up. Dataset of main text (e.g. Fig. 5.1), strain MG22 grown on M9+0.1% maltose.

6

STOCHASTICITY OF METABOLISM AND GROWTH AT THE SINGLE-CELL LEVEL

Elucidating the role of molecular stochasticity [159] in cellular growth is central to understanding phenotypic heterogeneity [160] and the stability of cellular proliferation [161]. The inherent stochasticity of metabolic reaction events [153] should have negligible effect, because of averaging over the many reaction events contributing to growth. Indeed, metabolism and growth are often considered to be constant for fixed conditions [162, 163]. Stochastic fluctuations in the expression level [7, 28, 159, 164] of metabolic enzymes could produce variations in the reactions they catalyse. However, whether such molecular fluctuations can affect growth is unclear, given the various stabilizing regulatory mechanisms [34, 165, 166], the slow adjustment of key cellular components such as ribosomes [116, 167] and the secretion [168] and buffering [169, 170] of excess metabolites. Here we use time-lapse microscopy to measure fluctuations in the instantaneous growth rate of single cells of Escherichia coli, and quantify time-resolved cross-correlations with the expression of lac genes and enzymes in central metabolism. We show that expression fluctuations of catabolically active enzymes can propagate and cause growth fluctuations, with transmission depending on the limitation of the enzyme to growth. Conversely, growth fluctuations propagate back to perturb expression. Accordingly, enzymes were found to transmit noise to other unrelated genes via growth. Homeostasis is promoted by a noise-canceling mechanism that exploits fluctuations in the dilution of proteins by cell-volume expansion. The results indicate that molecular noise is propagated not only by regulatory proteins [5, 26] but also by metabolic reactions. They also suggest that cellular metabolism is inherently stochastic, and a generic source of phenotypic heterogeneity.

The contents of this chapter have been published as Kiviet*, D. J., Nghe*, P., Walker, N., Boulineau, S., Sunderlikova, V., and Tans, S. J. *Stochasticity of metabolism and growth at the single-cell level.* Nature **514**, 376 (2014). [56]

6.1. RESULTS AND DISCUSSION

To investigate the dynamics of cellular growth, we followed individual *E. coli* cells growing on different nutrients. Among them was the synthetic sugar lactulose [171], which is imported and catabolized by the LacY and LacZ enzymes like its analogue lactose, but unlike lactose does not induce *lac* operon expression (Fig. 6.1A). Mixtures of lactulose and the gratuitous inducer isopropyl-β-D-thiogalactoside (IPTG) thus allowed us to vary the mean lac expression level independently and hence to explore different regimes of noise transmission. We determined the instantaneous growth rate $\mu(t)$ of individual cells within microcolonies at sub-cell-cycle resolution for various growth conditions, using time-lapse microscopy [5] at high acquisition rates and automated image analysis (appendix and chapter 2). We found that $\mu(t)$ varied considerably in time, both within one cell-cycle and between different cell-cycles (Fig. 6.1B,C and Fig. 6.4), with noise inten-

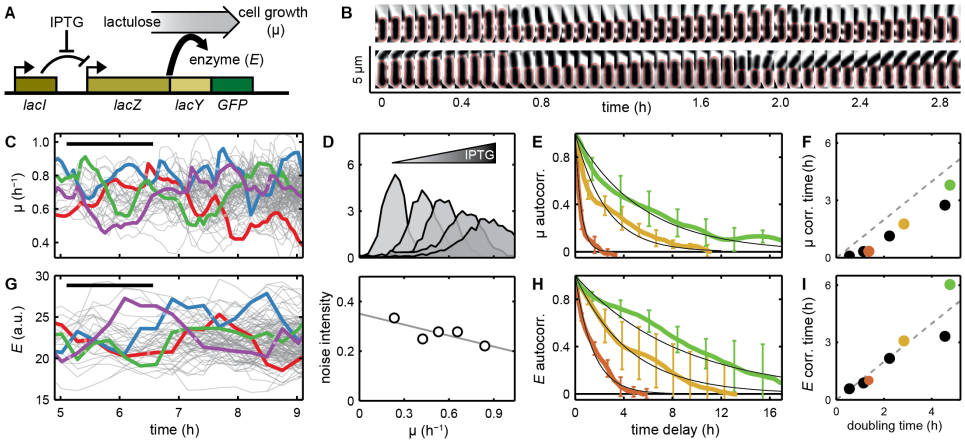


Figure 6.1: Growth rate variability in single *E. coli* cells. (**A**) Schematic diagram of the studied system. Lactulose is metabolized by the *lac* enzymes, but does not induce *lac* expression. Mean *lac* expression can hence be varied independently by the inducer IPTG. GFP is fused transcriptionally in the *lac* operon (Table 6.2). (**B**) Aligned phase-contrast images for two lineages. Microcolonies were grown on polyacryl pads (0.1% lactulose and 200 μM IPTG) for eight to nine generations. Up to 48 images were taken per hour. Red line: cell boundary from image analysis. (**C**) Instantaneous growth rate $\mu(t)$ against time, determined by fitting exponentials to the cellular length. Four lineages are colored for clarity. Black bar, mean division time; light points, division events. (**D**) Top: histograms of μ values for different IPTG levels. Bottom: noise intensity (standard deviation over the mean). (**E**) Autocorrelation function of $\mu(t)$ for low (4 μM, green), intermediate (6 μM, ochre) and high (200 μM, brown) IPTG levels. For clarity, error bars denoting the standard deviation are indicated only for a fraction of the points. Black lines: exponential fits that provide the correlation time. Correlation functions were determined along the branched lineages (see chapter 2.3.6, Fig. 2.8). (**F**) Graph of $\mu(t)$ correlation time versus mean doubling time. Colors are as in (**E**) ; black points are for growth on defined rich, lactose, succinate and acetate (in order of increasing doubling time). (**G**) - (**I**) As (**C**) , (**E**) and (**F**) , but for the fluorescence intensity reporting for $E(t)$ within single cells. Protein concentrations were determined by the mean fluorescence per unit area (see chapter 2.3.3, Fig. 2.5).

sities (standard deviation over the mean) ranging between 0.2 and 0.4 (Fig. 6.1D). Consistently, the growth rates of sister cells were significantly correlated (Fig. 6.5). We found that the typical timescales of the fluctuations were somewhat smaller than the mean cellular doubling time, as quantified by the autocorrelation functions $R_{\mu\mu}(\tau)$ (Fig. 6.1E,F). Such a scaling with doubling time is typical for protein concentration fluctuations [23]. Thus, the data indicated randomly fluctuating growth limitations, and suggested they could be caused by concentration fluctuations of cellular components.

To study the relation between growth and *lac* enzymes, we quantified the fluctuations in the *lac* production rate $p(t)$ and concentration $E(t)$ using green fluorescent protein (GFP) labeling (Fig. 6.1A,G-I, Fig. 6.4 and Fig. 2.5). We computed the cross-correlation functions $R_{p\mu}(\tau)$ and $R_{E\mu}(\tau)$, which indicate whether expression fluctuations correlate with μ-fluctuations occurring time τ later, and thus inform on the direction of transmission [27, 28]. Both $R_{p\mu}(\tau)$ and $R_{E\mu}(\tau)$ showed positive correlations regardless of the IPTG concentration (Fig. 6.2A,E-G). Their shapes and symmetries did depend on IPTG, however. At low and intermediate IPTG, $R_{E\mu}(\tau)$ was nearly symmetric around $\tau = 0$ while $R_{p\mu}(\tau)$ was asymmetric with larger weight at $\tau > 0$ (Fig. 6.2E,F Fig. 6.7). This would indicate that p fluctuations on average correlated more strongly with μ fluctuations that occur later. Such a delay in μ is consistent with the idea that *lac* expression fluctuations produce variations in lactulose catabolism, which in turn propagate through the metabolic network and perturb growth.

High IPTG $R_{E\mu}(\tau)$ displayed a positive peak at $\tau < 0$ (Fig. 6.2G and Fig. 6.7). Thus, E fluctuations correlated more strongly with μ fluctuations occurring earlier, which suggested backward transmission from growth to expression. Such a growth-to-expression coupling could be caused by specific regulatory interactions [53, 116, 172], or more generally by growth fluctuations that cause variations in general components that are required for transcription and translation. Overall, the data suggested that noise not only propagated forward, from expression to growth, but also backward, from growth to expression.

To determine whether back-and-forth transmission could explain the correlations, we developed a stochastic model. A black-box approach was followed, in which noise propagation is represented by phenomenological transmission coefficients that do not specify molecular details (Fig. 6.2B). Despite the circulating noise, the system could be decomposed into distinct noise transmission modes; here termed the *lac* catabolism, common noise and dilution modes (Fig. 6.2D). The cross-correlation curves for all induction levels (Fig. 6.2E-G) were fitted jointly, using the transmission strength from the common noise source to p as a single free parameter (Fig. 6.2H-J).

The effects of induction could be explained by altered intensities of the modes. At low and intermediate IPTG, the *lac* catabolism mode was dominant, with lac noise causing up to 30% of the growth noise (Table 6.1). At higher IPTG this mode weakened because of decreased transmission from E to μ. This decrease is plausible, as catalyzed reactions are less dependent on catalyst when the latter is abundant, consistent with the observed relation between the mean $\overline{E}$ and $\overline{\mu}$ (Fig. 6.2C). On the other hand, the rather constant $R_{p\mu}(0)$ (Fig. 6.2E-G) indicated that the common-noise mode had an almost fixed intensity for all IPTG concentrations. To probe the generality of this mode further, we made a number of genetic modifications. We found that it remained active when we knocked-

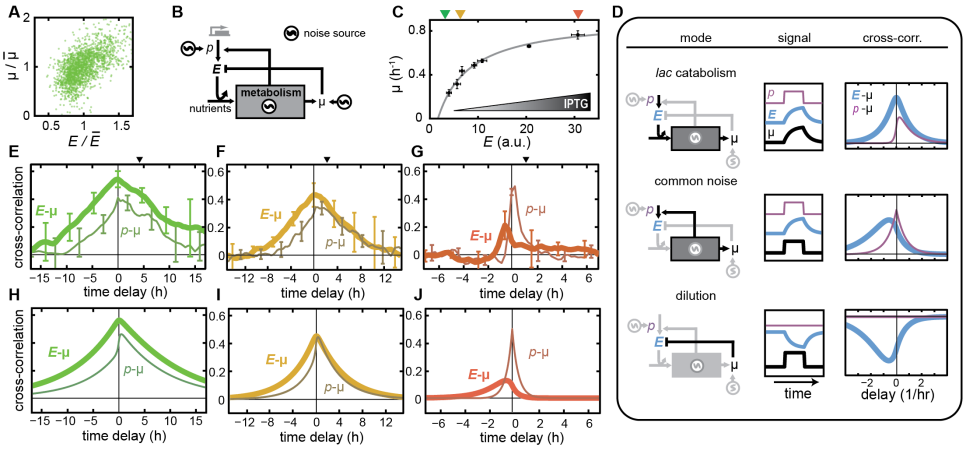


Figure 6.2: Cross-correlation functions and mathematical model. (**A**) Instantaneous growth rate against *lac* enzyme concentration from one microcolony, corresponding to the cross-correlation value $R_{E\mu}(0)$ in (**E**) . (**B**) Model of the coupling between expression and growth noise. Two noise sources are specific to p and μ, one is common to p and μ. Correlations arise when noise emitted from one source is received by two observables (p, E or μ). Analytical solutions revealed all contributing pathways, and showed they were finite despite the looped network structure (appendix). (**C**) The mean growth rate versus the mean expression level, as measured for different levels of IPTG induction. Line: fit to a Monod growth model. (**D**) Three classes of noise transmission modes. As an example, a noise source (left) emits a block wave, giving rise to signals μ, p and E (middle) and their cross-correlations (right). Other pathways contribute as well. For instance, common noise can also drive the catabolism mode. (**E**) -(**G**) Cross-correlation functions $R_{p\mu}(t)$ for the enzyme production rate $p(t)$ and growth rate $\mu(t)$ (thin line), as well as $R_{E\mu}(t)$ for the enzyme concentration $E(t)$ and $\mu(t)$ (thick line). Growth is on lactulose (0.1%) with IPTG: 4 μM (**E**) , 6 μM (**F**) , 200 μM (**G**) . Top triangles indicate mean division time. Error bars denoting the standard deviation are indicated for some data points only. The main features were robust to changing the growth determination method and taking the cell width into account (Fig. 6.6E-H). Growth and expression differences typically did not correlate with location within the microcolony (Fig. 6.6I). Protein production rates were determined by the time-derivative of the total fluorescence per cell (Fig. 2.5, Fig. 6.4A). (**H**) -(**J**) Fits to the experimental data (**E**) -(**F**) .

out the *lac* repressor, changed the GFP position within the operon, altered the type of fluorescent protein or used an exogenous constitutive promoter (Fig. 6.6A-D). These data suggest that common noise transmits to expression in general, which does not exclude additional coupling by specific regulatory interactions.

Next, we tested key findings. First, if the asymmetry in $R_{p\mu}(\tau)$ (Fig. 6.2E,F) is indeed caused by *lac* catabolism, this asymmetry should be suppressed when carbon enters central metabolism via another pathway. Growth on acetate was similarly slow as on lactulose and low induction, but $R_{p\mu}(\tau)$ was now indeed nearly symmetric (Fig. 6.3A,B and Fig. 6.7). At the same time, $R_{E\mu}(\tau)$ became more asymmetric as predicted for a dominant common noise mode transmission (Fig. 6.3A,B and Fig. 6.7). When growing on other natural substrates including lactose, the $R_{E\mu}$ peak-width scaled roughly with doubling time

consistent with dilution setting the transmission delay timescales (Fig. 6.3B and Fig. 6.8). To test further whether *lac* fluctuations could be causal in the growth noise, we exposed the cells to IPTG pulses in a microfluidic device. The resulting pulses in *lac* expression were indeed followed by a pulse in growth (Fig. 6.9A). Next, we aimed to mimic common noise fluctuations by growing cells on glucose minimal medium and pulsing with amino acids. These pulses indeed produced transient increases in μ and p (Fig. 6.9B), consistent with common noise propagating to enzyme expression and to growth.

Second, the network structure implied a homeostatic control mechanism: upward fluctuations in common noise increase E when transmitted via p, but also decrease E when transmitted via μ (Fig. 6.2B). These opposing effects offer a direct prediction: if the positive pathway dominates, $R_{E\mu}(\tau)$ should be positive, as is the case so far. If the negative pathway would dominate, however, $R_{E\mu}(\tau)$ should become negative (Fig. 6.3C). One cannot manipulate how volume changes affect dilution. To tilt the balance, we thus looked for constructs with a weaker coupling to common noise in the positive pathway, as measured by $R_{p\mu}(0)$. A constitutively expressed mCherry with a twofold lower $R_{p^*\mu}(0)$ indeed displayed negative $R_{E^*\mu}(\tau)$ (Fig. 6.3D and Fig. 6.7). Thus, two parallel antagonistic pathways that together form a so-called incoherent feedforward network motif [173] can partly cancel noise. This canceling also explains why $R_{E\mu}(0)$ is low even though $R_{p\mu}(0)$ is high at high induction where common noise dominates (Fig. 6.2G). Interestingly, while up-fluctuations in μ are associated with up-fluctuations in E (Fig. 6.2G), increases in mean $\overline{\mu}$ lead to decreases in $\overline{E}$ (Fig. 6.8E) [53, 116]. These opposing dependen-

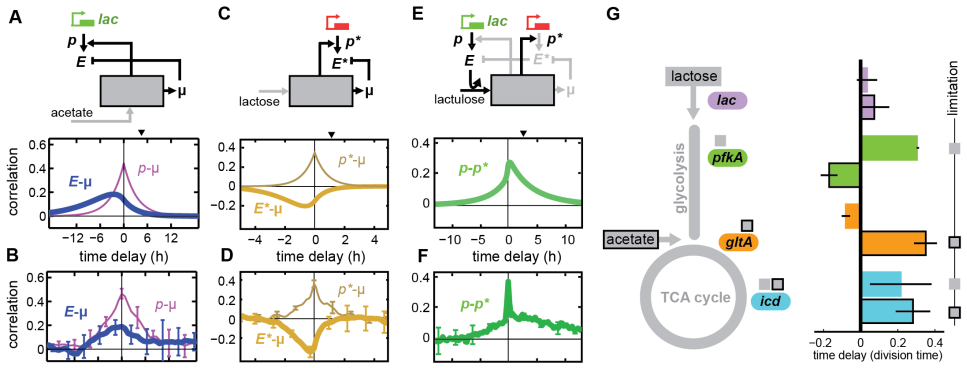


Figure 6.3: Model predictions and experimental tests. Top: re-wired noise transmission networks with predicted dominant pathways (black). Colored genes indicate labeling with GFP and mCherry. Middle: predicted cross-correlation with mean doubling time (triangle). Bottom: measured cross-correlation. Error bars denote the standard deviation. (**A**) , (**B**) For growth on acetate the *lac* enzymes are catabolically inactive. (**C**) , (**D**) Gene with a weaker coupling from common noise to expression (compared with the *lac* operon), leading to dominant dilution. (**E**) , (**F**) Transmission from the *lac* genes to another gene via growth. When the *lac* genes do not transmit because cells grow on acetate, the correlation is symmetric (Fig. 6.10C,D). (**G**) Time delays for *lac* , *pfkA*, *gltA* and *icd* in lactose (not boxed) and acetate media (boxed), as derived from the correlation functions $R_{p\mu}(\tau)$ (Fig. 6.10E). Small square boxes indicate which gene is considered limiting in steady-state in a particular medium (see main text).

cies suggest that different mechanisms underlie these two types of expression variation.

Third, if *lac* enzymes transmit to growth and growth transmits to expression in general, then *lac* enzymes ought to transmit also to other genes. Hence we quantified $p^*(t)$ of mCherry controlled by promoters with no known functional interactions with the *lac* system. For lactulose and low induction, mCherry fluctuations indeed occurred after *lac* fluctuations on average (Fig. 6.3F and Fig. 6.10A,B) in accordance with predictions (Fig. 6.3E). In contrast, this delay was absent for acetate, which is consistent because *lac* then does not transmit to growth (Fig. 6.10C,D). Noise in *lac* expression can thus couple to other genes without specific regulatory interactions.

For the *lac* genes, the *lac* catabolism mode transmitted to growth only when the mean *lac* expression was kept artificially low and limited the mean growth rate. Hence, we wondered whether limiting enzymes in central metabolism could similarly perturb growth. For growth on lactose, glycolysis is considered limited by *pfkA*, and the tricarboxylic acid cycle by *icd* but not by *gltA*; while in acetate, *gltA* is limiting, *icd* may be limiting but *pfkA* is not [174–176]. We indeed observed positive time delays in $R_{p\mu}$ for *pfkA* and *icd* in lactose, and for *gltA* and *icd* in acetate, but not in the other cases (Fig. 6.3G and Fig. 6.10E). This pattern of correlation delays is consistent with the mechanism found for *lac* , in which growth limitation in steady-state resulted in noise transmission to growth. Notably, the differences in noise transmission behavior were observed for enzymes catalyzing nearby reactions in the pathway. For instance, *icd* acts almost directly after *gltA*, but *icd* displayed delayed correlation in lactose while *gltA* did not. This excludes the possibility that the delayed correlations are caused by synchronous fluctuations of *pfkA*, *gltA*, *icd* and other central metabolic genes. Together, the results indicate that expression-to-growth noise propagation occurs more generally for limiting genes.

6.2. CONCLUSIONS

Our study shows that fluctuations in gene expression can affect the growth stability of a cell, and, in turn, growth noise affects gene expression. This entanglement between growth and expression noise reflects the inherent auto-catalytic nature of self-replicating systems: metabolic enzymes help synthesize the building blocks for their own synthesis. The results raise the question how different fluctuating metabolic activities within the cell are coordinated, and which regulatory mechanisms are implicated in maintaining growth homeostasis. Metabolic stochasticity could allow clonal cells in a population to adopt a wide spectrum of metabolic states, and hence enable bet-hedging strategies to exploit new conditions optimally. Metabolic stochasticity could represent a generic source of cellular heterogeneity [15], but also prevent optimal growth [177] and limit efficient biosynthesis. Novel approaches are required to incorporate noise transmission within the current theoretical framework of metabolism.

6.3. APPENDIX

6.3.1. EXPERIMENTAL MATERIAL AND METHODS

TIME LAPSE MICROSCOPY, IMAGE ANALYSIS AND CROSS-CORRELATIONS

The experiments and analysis methods were extensively described in chapter 2. Therefore, we here only give a brief summary and focus on details that differed from the commonly used methods.

We used derivatives of *E. coli* strain MG1655 (see Table 6.2) for the experiments. To measure expression of the *lac* operon, *lacA* was replaced with *GFPmut2* [84]. *mCherry* and *GFPmut2* controlled by the constitutive promoter P_{N25} [178] was inserted into the chromosome at different locations using the Datsenko & Wanner protocol [58]. *LacZ-GFPmut2* fusion was performed using the Hamilton & al. protocol [179].

For the microscopy experiment cells were grown in minimal M9 medium supplied with one of various carbon sources or in EZ defined rich medium. First, cells were inoculated into TY medium at 37 °C in the morning and after several hours highly diluted into the medium used for microscopy. On the next morning the still exponentially growing cells were diluted to OD≈0.005. An acryl gel pad which was presoaked in medium was placed into a glass chamber and 1µL of cells were applied onto it. The glass cavity was closed with a coverslip and clamped tight. Despite thorough washing of glassware and using distilled water, our gels contained organic contamination as observed by slow but significant cell growth on minimal M9 media without sugar. Such organic contaminations have also been observed for growth in batch cultures [180–182]. To make sure that cells would not use these contaminants as carbon source in experiments at low growth rate, we first grew cells with a knocked-out *lac* operon on the gel (strain NCM520), as to consume the organic contaminants before the actual experiment.

Growth of cells into microcolonies was observed with an inverted microscope (Nikon TE2000) at 100x magnification and 1.5x intermediate magnification. Image acquisition was automatized and phase contrast and fluorescence time-lapse movies were acquired. We measured maturation times of GFPmut2 and mCherry to be approximately 5 and 15 minutes, respectively (Fig. 3.4), similar to other reported values [102, 110]. The movies were analyzed offline with a custom software based on Schnitzcells [39]. Cells were segmented and tracked, resulting in a full lineage tree of several hundred cells. Growth rate as derivative of cell length was determined at sub-cell-cycle resolution. Fluorescence images were corrected and apparent protein concentration and production rate were determined as described in chapter 2. All obtained parameters were corrected for quasi-periodic cell-cycle fluctuations (Fig. 6.11). The temporal fluctuations of two signals relative to each other were determined by cross-correlations. Here, the redundancy of data in the tree-like lineage structure was taken into account (Fig. 2.8).

PULSE EXPERIMENTS IN MICROFLUIDIC DEVICE

The microfluidics devices were similar to those described in ref. [31]. Briefly, concentrated cells are loaded into a PDMS structure bound to glass, after which cells grow in a narrow channel of limited length. Due to growth, cells are pushed out of this narrow channel into the wider flow channel that supplies nutrients and washes away the excess cells. Media is pumped through the device at a rate of 0.5mL/hr using syringe pumps

(New Era). Media is switched using a valve (Hamilton), with a characteristic half-time of less than 1 minute (determined using a fluorescent dye). Cells were imaged with an inverted Olympus X81 microscope with ZDC hardware focus, equipped with an oil objective (Olympus UPLFLN100x), cooled CCD camera (Olympus XM10), mercury lamp with liquid light guide (X-Cite 120), GFP filter set (Chroma N49002), automated stage (Märzhäuser Scan IM 120x80) and an incubation chamber (Life Imaging Services) allowing precise 37°C temperature control.

6.3.2. MODELING MATERIAL AND METHODS

ANALYTICAL EXPRESSION OF THE CORRELATIONS

We note E the enzyme concentration, p its production rate and μ the rate of increase of volume. Given that enzymes are long-lived compared to the cell cycle time:

$$\dot{E} = p - \mu \cdot E \tag{6.1}$$

The rates p and μ can vary due to fluctuations in the concentration or the activity of global factors (such as ribosomes, ATP, amino-acids or common regulatory elements) or due to fluctuations of specific factors, such as *lacI* repressor binding to the *lac* operon or enzymes catalyzing cell wall production. Hence we consider three independent noise sources: N_G is the effect of common components, N_E and N_μ are components specific respectively to E (more specifically: protein production p) and μ. These N_X are modeled as independent colored noises:

$$\dot{N}_X = \theta_X - \beta_X \cdot N_X \quad . \tag{6.2}$$

Here, β_X is the decay rate and θ_X the amplitude of a white noise source. We computed analytical solutions for the time-correlations under a linear response approximation, which is suited to probe quantitatively the short term response of the network when fluctuations are of sufficiently limited amplitude. We defined the perturbed variables $\delta X(t) = X(t) - X_0$ (where X_0 is the mean of X) and logarithmic gains T_{XY} representing how a variable X responds to the fluctuations of a source Y. Given the network of noise coupling interactions described in the main text (Fig. 6.2B), we write the following additional relations between variables p, E, and μ, and the noise sources N_G , N_E , and N_μ :

$$\frac{\delta\mu}{\mu_0} = T_{\mu E} \cdot \frac{\delta E}{E_0} + T_{\mu G} \cdot N_G + N_\mu \tag{6.3}$$

$$\frac{\delta p}{E_0 \cdot \mu_0} = T_{EE} \cdot \frac{\delta E}{E_0} + T_{EG} \cdot N_G + N_\mu \tag{6.4}$$

A first order development of equation 6.1 gives:

$$\frac{\dot{\delta E}}{E_0 \cdot \mu_0} + \frac{\delta E}{E_0} = \frac{\delta p}{E_0 \cdot \mu_0} + T_{E\mu} \cdot \frac{\delta\mu}{\mu_0} \tag{6.5}$$

where the transmission of dilution fluctuations is denominated as $T_{E\mu} = -1$. Fourier transforms of equation 6.3-6.5 ($\widetilde{X}$ indicates the Fourier transform of X) result in a linear

system for E and μ fluctuations in the frequency domain:

$$\frac{\widetilde{\delta E}}{E_0} = \frac{\mu_0}{\mu_E + i\omega}\left[\widetilde{N}_E + \left(T_{EG} + T_{E\mu}T_{\mu G}\right)\cdot\widetilde{N}_G + T_{E\mu}\widetilde{N}_\mu\right] \tag{6.6}$$

$$\frac{\widetilde{\delta\mu}}{\mu_0} = \frac{\mu_0}{\mu_E + i\omega}T_{\mu E}\cdot\left[\widetilde{N}_E + \left(T_{EG} + T_{E\mu}T_{\mu G}\right)\cdot\widetilde{N}_G + T_{E\mu}\widetilde{N}_\mu\right] + T_{\mu G}\widetilde{N}_G + \widetilde{N}_\mu \tag{6.7}$$

where the rate $\mu_E = \mu_0(1 - T_{E\mu}T_{\mu E} - T_{EE})$ sets the timescale of E fluctuations.

Let $R(\tau) = \left\langle \int_{-\infty}^{+\infty} \frac{E(t)}{E_0}\frac{\mu(t+\tau)}{\mu_0}d\tau \right\rangle$ be the cross-covariance between E and μ normalized by their mean, R_{EE} and $R_{\mu\mu}$ their respective auto-covariance. In the frequency domain we have:

$$\widetilde{R}_{\mu E} = \left\langle \frac{\widetilde{\delta\mu}}{\mu_0}\cdot\frac{\widetilde{\delta E}^*}{E_0}\right\rangle, \qquad \widetilde{R}_{EE} = \left\langle \frac{\widetilde{\delta E}}{E_0}\cdot\frac{\widetilde{\delta E}^*}{E_0}\right\rangle, \qquad \widetilde{R}_{\mu\mu} = \left\langle \frac{\widetilde{\delta\mu}}{\mu_0}\cdot\frac{\widetilde{\delta\mu}^*}{\mu_0}\right\rangle \tag{6.8}$$

Using the independence of noise sources (eq. 6.2) together with eq. 6.6- 6.8, the cross-covariance in the time domain is:

$$\begin{aligned} R_{\mu E}(\tau) = {} & 1\cdot T_{\mu E}\cdot S_E(\tau) + \left(T_{EG} + T_{E\mu}T_{\mu G}\right)\cdot T_{\mu E}\left(T_{EG} + T_{E\mu}T_{\mu G}\right)\cdot S_G(\tau) \\ & + T_{E\mu}\cdot T_{\mu E}T_{E\mu}\cdot S_\mu(\tau) + \left(T_{EG} + T_{E\mu}T_{\mu G}\right)\cdot T_{\mu G}\cdot A_G(\tau) + T_{E\mu}\cdot 1\cdot A_\mu(\tau) \end{aligned} \tag{6.9}$$

Each term of the sum on the right-hand side of equation 6.9 is the product of three factors separated by dots, respectively representing: fluctuations transferred to E, fluctuations transferred to μ, and a τ-dependent function quantifying delayed correlation arising from a particular noise source. We find these functions have one of two forms: S_X (symmetric) or A_X (asymmetric, negatively delayed), which are inverse Fourier transforms of the frequency dependent components of the cross-correlation (obtained following the method described in detail in [27]):

$$S_X(\tau) = \theta_X^2\frac{\mu_0^2}{2(\beta_X^2 - \mu_E^2)}\left(\frac{e^{-sign(\tau)\cdot\mu_E\cdot\tau}}{\mu_E} - \frac{e^{-sign(\tau)\cdot\beta_X\cdot\tau}}{\beta_X}\right) \tag{6.10}$$

$$A_X(\tau) = \begin{cases} \theta_X^2\left(\frac{\mu_0}{(\beta_X^2-\mu_E^2)}e^{\mu_E\cdot\tau} - \frac{\mu_0}{2\beta_X(\beta_X-\mu_E)}e^{\beta_X\cdot\tau}\right) & \text{if } \tau < 0 \\ \theta_X^2\cdot\frac{\mu_0}{2\beta_X(\beta_X-\mu_E)}e^{-\beta_X\cdot\tau} & \text{if } \tau \geq 0. \end{cases} \tag{6.11}$$

The first three terms of eq. 6.9 correspond to the symmetric modes generated by direct transmission fluctuations from E to μ, originating from N_E , N_G and N_μ (Fig. 6.2D). These modes thus depend on *lac* catabolism (they disappear when $T_{\mu E} = 0$), and are hence named *'lac catabolism'* modes (see Fig. 6.2D). The fourth term represents the asymmetric contribution arising from the joint dependence of protein synthesis and cell volume increase on noise in common factors, and was hence named the *'common noise'* mode (see Fig. 6.2D). The fifth term is the asymmetric contribution originating from the effect of volume growth on enzyme concentration by dilution ($T_{E\mu} = -1$), and was thus termed the *'dilution'* mode (see Fig. 6.2D). Similarly, we obtain analytical expressions for the auto-covariance for $\tau \geq 0$:

$$R_{EE}(\tau) = S_E(\tau) + \left(T_{EG} + T_{E\mu}T_{\mu G}\right)^2\cdot S_G(\tau) + T_{E\mu}^2\cdot S_\mu(\tau) \tag{6.12}$$

$$R_{\mu\mu}(\tau) = T_{\mu E}^2 \cdot A_E(\tau) + 2\frac{\mu_E}{\mu_0} T_{\mu E}(T_{EG} + T_{E\mu}T_{\mu G}) \cdot S_G(\tau) + 2\frac{\mu_E}{\mu_0} T_{\mu E} T_{E\mu} \cdot S_\mu(\tau) + T_{\mu G} \cdot B_G(\tau) + B_\mu(\tau) \tag{6.13}$$

where $B_X(\tau) = \frac{\theta_X^2}{2\beta_X} e^{-\beta_X \cdot \tau}$.

The cross-covariance function between the production rate and growth rate is obtained in the same way:

$$R_{p\mu}(\tau) = T_{EG} \cdot T_{\mu G} \cdot B_G(\tau) + T_{EE} \cdot T_{\mu E} \cdot \left[S_E(\tau) + \left(T_{EG} + T_{E\mu}T_{\mu G}\right)^2 S_G(\tau) + T_{E\mu}^2 S_\mu(\tau) \right] + T_{EE}(T_{EG} + T_{E\mu}T_{\mu G}) \cdot T_{\mu G} \cdot A_G(-\tau) + T_{EG} \cdot T_{\mu E}(T_{EG} + T_{E\mu}T_{\mu G}) \cdot A_G(\tau) + T_{EE}T_{E\mu} \cdot 1 \cdot A_\mu(-\tau) + 1 \cdot T_{\mu E} \cdot A_E(\tau). \tag{6.14}$$

In the case of a second constitutively expressed gene F, we computed similarly the different types of correlation functions, assuming the response of the production rate p^* to be:

$$\frac{\delta p^*}{F_0 \cdot \mu_0} = T_{FE} \cdot \frac{\delta E}{E_0} + T_{FG} \cdot N_G + N_F. \tag{6.15}$$

6

FITTING PARAMETERS FOR GROWTH EXPERIMENTS ON LACTULOSE

The parameters of the model (Table 6.3) are the average growth rate μ_0, the amplitudes of the noise sources $\{\theta_E, \theta_\mu, \theta_G\}$, their relaxation constants $\{\beta_E, \beta_\mu, \beta_G\}$ and the transfer coefficients $\{T_{\mu G}, T_{EE}, T_{EG}\}$. Additional coefficients $T_{E\mu}$ and $T_{\mu G}$ were used in the model for notation consistency but have imposed values: $T_{E\mu} = -1$ models the physical dilution and $T_{\mu G}$ was arbitrarily set to 1 as it only normalizes T_{EG} and N_G.

Most parameters were determined directly from the measurements as follows. The population average growth rate μ_0 was determined as the average of the measured instantaneous growth rates. Timescales $1/\beta_\mu$ and $1/\beta_G$ (growth specific and common noise respectively) were taken equal to the measured autocorrelation time of μ. The timescale $1/\beta_E$ (*lac* specific noise) was varied to fit $R_{p\mu}(0)$ and was in all cases found to be consistent with previously reported timescales [5] for gene specific noise (≈ 9 min in doubling units). The amplitudes of the noise sources $\{\theta_E, \theta_\mu, \theta_G\}$ were determined from the experimentally measured $\{R_{EE}(0), R_{\mu\mu}(0), R_{E\mu}(0)\}$ by solving the linear system obtained from equations 6.9, 6.12, 6.13 in real time space at $t = 0$. The feedback of E on itself T_{EE} determines the characteristic time of E fluctuations (see μ_E in equations 6.6 and 6.7) and was taken to match the measured autocorrelation time of E.

The two remaining parameters T_{EG} (response of *lac* expression to common noise N_G) and $T_{\mu E}$ (response of μ to E fluctuations) were the only parameters that determined the values of the cross-correlation functions $R_{E\mu}$ and $R_{p\mu}$ for time delays different than zero. $T_{\mu E}$ was determined independently from the cross-correlations as the slope of the $E - \mu$ clouds of Fig. 6.12. Consequently, the only fitting parameter for the shape of the functions $R_{E\mu}(\tau)$ and $R_{p\mu}(\tau)$ was T_{EG}, for which we could use the single value of 1.3 for all experiments.

ROBUSTNESS OF THE FIT

We quantified the quality of the fit of the cross-correlations by computing the weighted sum of squared errors $\theta^2 = \frac{1}{N}\sum_t \frac{(y(t)-f(t))^2}{\sigma(t)^2}$, where $y(t)$ is the measured correlation at time delay t, $f(t)$ the predicted correlation, $\sigma(t)$ the experimental variance and N the number of time points. $\theta^2 \approx 1$ can be interpreted as the amplitude of the residuals being comparable to the experimental variance, while $\theta^2 > 1$ indicates that the fit is poor. We computed θ^2 for $R_{E\mu}$ and $R_{p\mu}$ for different T_{EG}. The range $T_{EG} = 1.3 - 1.5$ yielded best quality fits $\theta^2 < 0.6$ for all IPTG induction levels on lactulose. For simplicity, we used the unique value $T_{EG} = 1.3$ which is optimum at low induction (Fig. 6.13A). The symmetries of the cross-correlations were robust to variations of the parameters T_{EG} and $T_{\mu E}$ that determine the shape of the cross-correlation functions (Fig. 6.13B,C).

PREDICTIONS FOR REWIRED NETWORKS

For the other conditions of Fig. 6.3, average growth rate μ_0 was measured experimentally. Following the trends observed on lactulose experiments, we took $\beta_\mu = \beta_M = 2\mu_0$ and $1/\beta_E = 9$ min. For $T_{\mu E}$ and T_{EE} , we took the values found for lactulose experiments at the same induction level, otherwise zero when the enzyme was in excess or inactive. We considered $T_{EE}/T_{\mu G}$ to be promoter specific and kept the value of 1.3 for the *lac* operon. We found $T_{CG}/T_{\mu G} = 0.7$ for the exogenous constitutive promoter independently of the genomic location of its chromosomal insertion. θ_E, θ_μ and θ_G were fitted to reproduce the experimentally measured values $R_{EE}(0)$, $R_{\mu\mu}(0)$ and $R_{p\mu}(0)$. In all cases (except panel G, for which there is no prediction), the cross-correlation between growth rate and gene concentration was fully predicted. The correlation between the production rates p of *lac* enzyme and p^* of mCherry under a constitutive promoter (Fig. 6.3E, Fig. 6.10C,D) were fully predicted: we took the parameters already obtained from the experiments on the same growth medium, assumed that noise transmission from the *lac* catabolism affected similarly the production rate of *lac* and the production rate of the constitutive gene ($T_{CE} = T_{EE}$), and arbitrarily took the same noise level η_F for the constitutive promoter as measured on lactose.

EFFECT OF ACTIVE DEGRADATION

The model can be extended to take into account active degradation of *lac* enzymes at rate γ_0, by replacing the expression of μ_E by $\gamma_0 + \mu_0(1 - T_{E\mu}T_{\mu E} - T_{EE})$. Interestingly, active degradation would result in less self-sustained fluctuations of E, and the model predicts a reduction of 90% in E noise and a corresponding decrease of 19% in μ noise (Fig. 6.14). This indicates a potential mechanism for noise reduction at the expense of higher protein production costs.

6.3.3. ADDITIONAL FIGURES

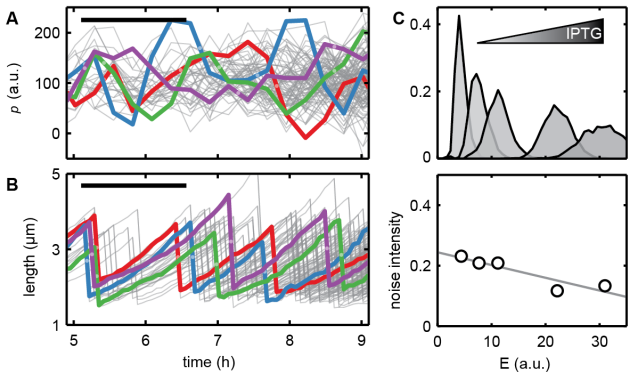


Figure 6.4: Cell length, enzyme concentration and production rate. (**A**) Enzyme production rate against time $p(t)$ for all lineages within a microcolony, from 5 h into the experiment and onwards. Four lineages are colored for clarity. Black bar, mean division time; light points, division events. (**B**) Cell length against time $L(t)$ as in (**A**) . (**C**) Histograms of observed E values for different IPTG induction levels. Bottom panel indicates the noise intensity, defined as the standard deviation over the mean. For information on cell segmentation, determination of cell length and enzyme concentration (part of the published figure) we refer the reader to chapter 2, Fig. 2.2-2.5.

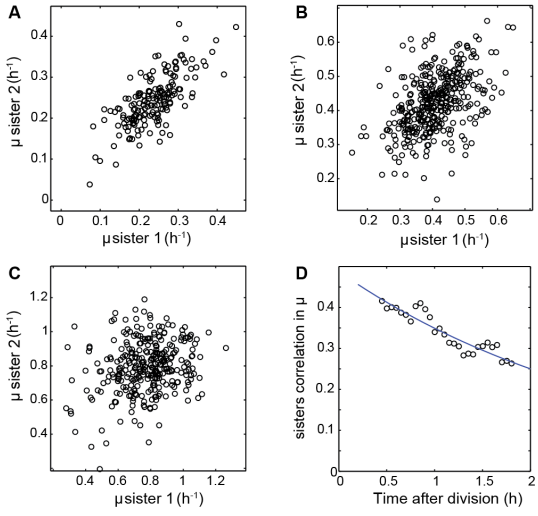


Figure 6.5: Correlations between the growth rate of sister cells during growth on lactulose for increasing levels of IPTG induction. (**A**) At 4μM IPTG, $R = 0.72$, $n = 171$, $P < 10^{-27}$ (t-test). (**B**) At 6μM IPTG, $R = 0.42$, $n = 382$, $P < 10^{-16}$. (**C**) At 200μM IPTG, $R = 0.32$, $n = 314$, $P < 10^{-8}$. (**D**) Evolution in time of the correlation coefficient between growth rate of sisters, for 6μM IPTG. A decreasing exponential was fitted with a decay time of 2.86 h.

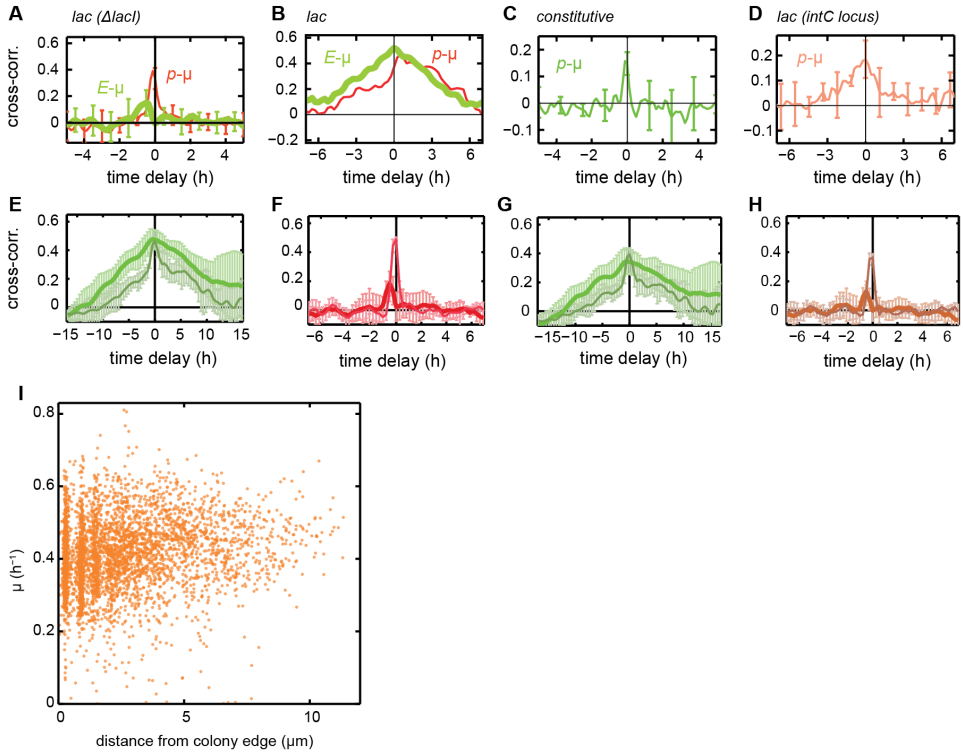


Figure 6.6: Cross-correlations of control experiments and using different methods of growth-rate determination. (**A**) Expression of *lac* in a *lacI* repressor knockout strain on lactose minimal medium (to be compared with Fig. 6.2G). (**B**) Expression of *lac* measured with a GFP fusion to LacZ shows the same result as co-transcriptional expression of GFP on 0.1%lactulose and 6 μM of IPTG (to be compared with Fig. 6.2F). (**C**) Exogenous constitutive promoter (P_{N25}) driving the production of GFP, inserted in the *cheZ* locus, on minimal medium with lactose. (**D**) The *lac* promoter driving the production of yellow fluorescent protein (YFP), inserted in the *intC* locus, on minimal medium with maltose. (**E**) Cross-correlations for lactulose growth at low IPTG (4 μM), with growth rate determined as follows: $S(t)$ is the surface area of the cell silhouette versus time (Fig. 2.4A). The growth rate is the time derivative of $S(t)$. (**F**) The same, for lactulose growth at high IPTG (200 μM). (**G**) Cross-correlations for lactulose growth at low IPTG (4 μM), with growth rate determined as follows: $S(t)$ is the surface area of the cell silhouette versus time, $L(t)$ is the length of the cell silhouette versus time (Fig. 2.4A,B). The growth rate is the derivative of $L(t)\cdot(S(t)/L(t))^2$. Note that $S(t)/L(t)$ is taken as a measure for the width of the cell, and the width squared times the length as a measure for the cell volume. (**H**) The same, for lactulose growth at high IPTG (200 μM). These cross-correlations display the same shape and symmetry as in Fig. 6.2E, G, where the growth rate is determined as the derivative of the length of the cell silhouette (Fig. 2.4C). Hence the central features are robust to different methods of growth rate determination. (**I**) Scatter plot of instantaneous growth rate and cell position within the microcolony. The cell position was calculated as the minimal distance of the center of a cell to the edge of the microcolony. Data obtained during growth on lactulose at intermediate IPTG induction (6 μM).

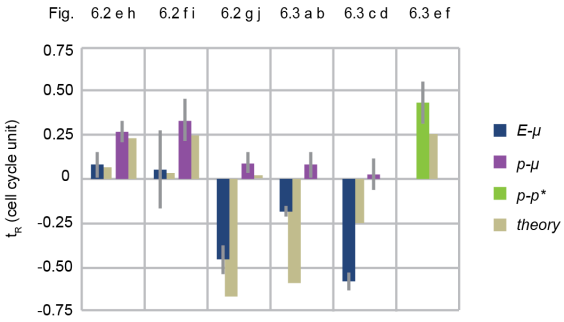


Figure 6.7: Quantification of symmetry of cross-correlation functions. For each cross-correlation (corresponding figure indicated at top), we computed the weighted average of the time delay $\tau_R = \sum_{t=I}(R_t \cdot t)/\sum_{t=I} R_t$, with R_t the correlation intensity at time delay t, considering significantly cross-correlations (t-test, $P < 0.05$, $n = 4$) within the interval $I = [-2,2]$ cell cycles. A positive (respectively negative) τ_R indicates that the cross-correlation R has more weight at positive (respectively negative) times. Error bars denote the standard deviation of the symmetry values determined for four subbranches. Note that the $E-\mu$ cross-correlations of Fig. 6.3C,D are negative, and hence we display $-\tau_R$.

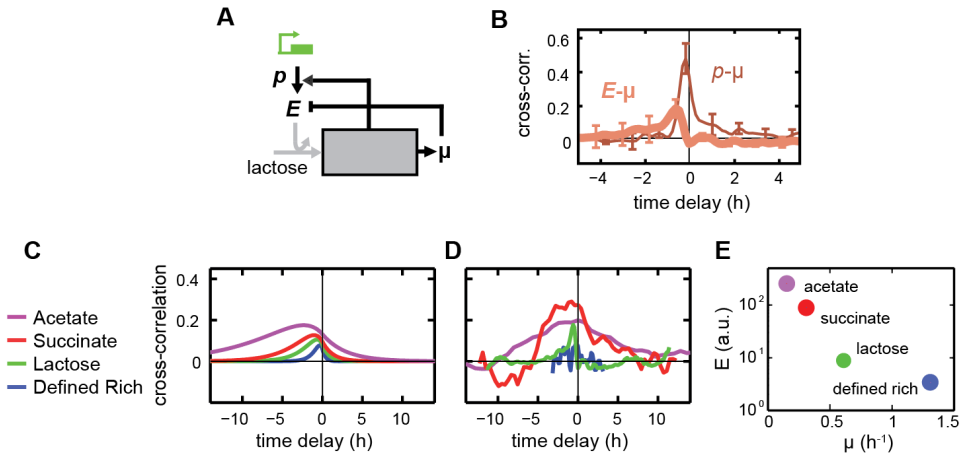


Figure 6.8: Cross-correlations for growth on different carbon sources. (**A**) Schematic diagram of noise transmission during growth on lactose, which is predicted to be similar to the case of growth on lactulose at high IPTG induction (see Fig. 6.2G, J). (**B**) Corresponding measured cross-correlations. (**C**) Theoretical cross-correlations obtained by using the parameters during growth on lactulose and changing exclusively the population average growth rate to the experimentally measured value. This prediction displays a positive asymmetric peak towards negative time and a width scaling with the average growth rate. (**D**) Corresponding measured cross-correlations. (**E**) Population average *lac* enzyme concentration versus the population average growth rate on minimal medium supplemented with varying carbon sources.

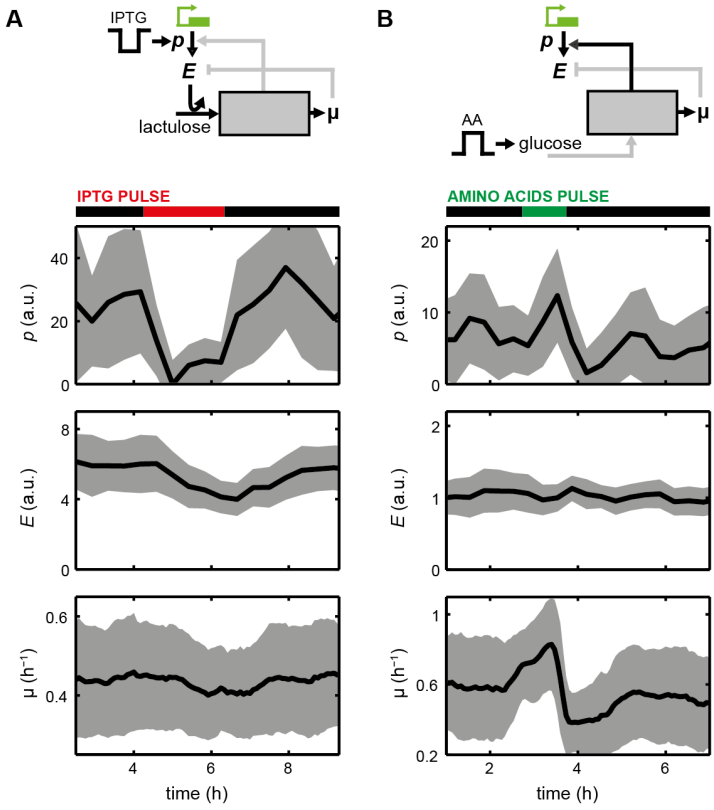


Figure 6.9: External media perturbations in microfluidic device. (**A**) Growth of AB460 in microfluidic device (see appendix) on M9 medium with 0.1% lactulose, 0.01% Tween-20 and 16 μM IPTG. A 2-h pulse to medium with 3 μM IPTG is indicated in red. Black line is the mean, and gray area is the standard deviation, of approximately 60 cells. Indicated are the *lac* production rate (p), lac concentration (E) and cell growth rate (μ). The duration and intensity of the pulse was chosen to reflect the naturally occurring fluctuations in *lac* expression. Upon the pulse, the production rate transiently decreased, followed by a gradual transient decrease in *lac* concentration, and a transient decrease in growth rate. These data are consistent with the catabolism transmission mode (top). (**B**) Growth of ASC631 in microfluidic device on M9 medium with 0.1% glucose, 0.01% Tween-20 and 1 μM IPTG. To mimic fluctuations in common components, a 1-h pulse of amino acids (Teknova M2104) added to the medium is indicated in green. Both growth and production rate increase immediately upon addition of amino acids, reflecting the common noise transmission mode (top). The enzyme concentration remained relatively stable, showing that for these perturbations the production increase and dilution increase canceled each other. These data are consistent with the common noise mode (top).

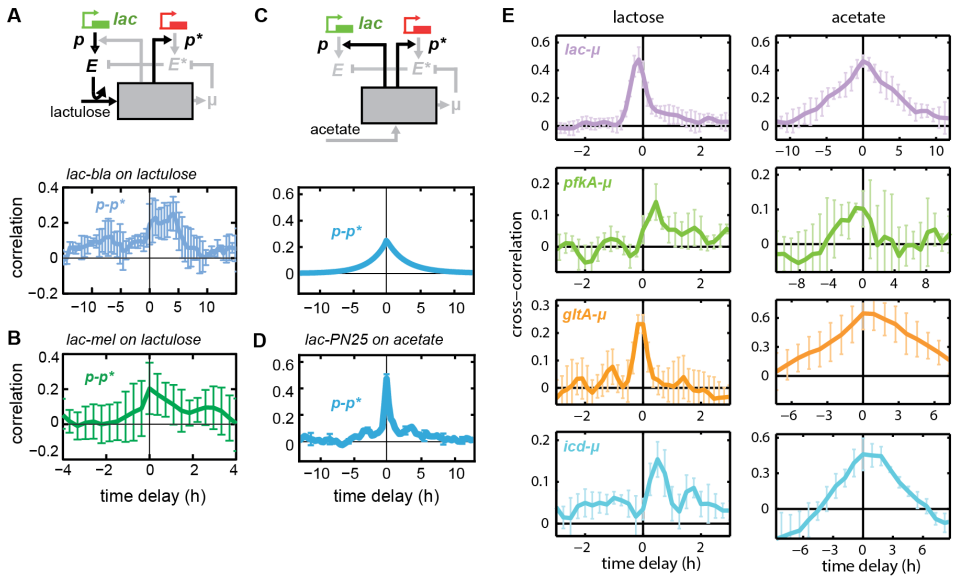


Figure 6.10: Cross-correlations of additional constructs. (**A**) Transmission from *lac* to another gene via growth (on 0.1% lactulose and 6 μM IPTG) shown by the asymmetric cross-correlations between *lac* production rate and mCherry production driven by the constitutive *bla* promoter. (**B**) The same for *lac* production rate and mCherry driven by the *mel* promoter induced by 0.2% melibiose ($\Delta melA$ strain). (**C**) Symmetric cross-correlation between *lac* production rate p and other gene production rate p^* predicted for growth on acetate (see (**D**)). (**D**) Absence of transmission shown by the cross-correlation between *lac* production rate p and the mCherry production rate p^* driven by the constitutive P_{N25} promoter, on minimal medium with 0.18% acetate, consistent with predictions (see (**C**)). (**E**) Cross-correlations ($R_{p\mu}$) for *lac* , *pfkA*, *gltA* and *icd* in lactose (left) and acetate media (right).

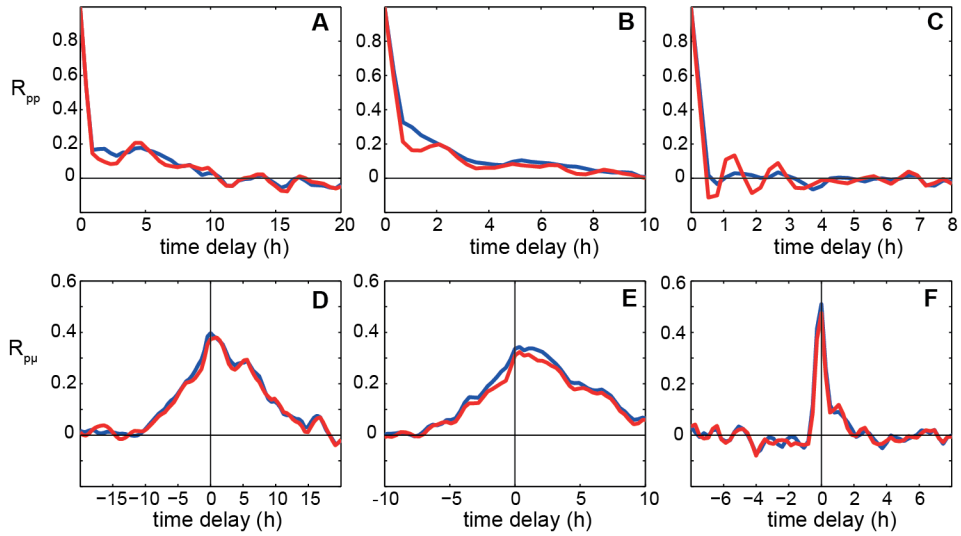


Figure 6.11: Effect of the cell cycle correction (red without and blue with correction) on the autocorrelation of the production rate of *lac p* (A-C) and the cross-correlation between p and μ (D-F) during growth on lactulose for the respective induction levels IPTG = 4 / 6 / 200 µM. The increase of the gene copy number across the cell cycle causes periodic variations in the production rate, which appear as oscillations in R_{pp} with the periodicity of the average cell cycle duration. Subtracting the averaged trend over all cell cycles from the signals allow to remove most of this signal. Cross-correlations $p-\mu$ are negligibly affected by this correction.

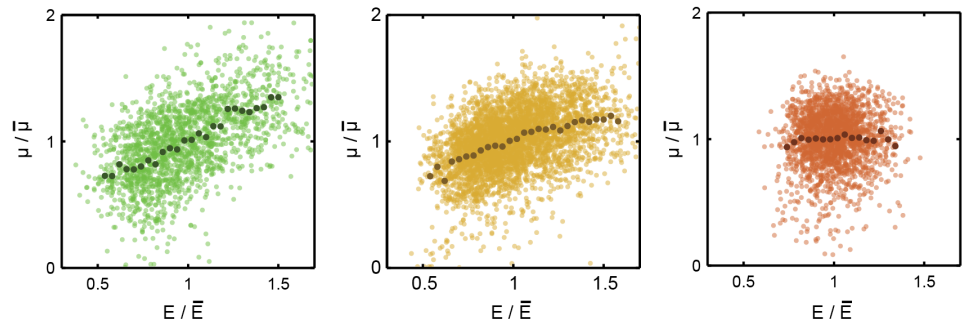


Figure 6.12: Scatter plots of the growth rate and the enzyme concentration normalized by their average. Darker dots indicate the μ average over bins of E. These plots have been obtained for increasing IPTG induction levels on minimal medium and lactulose 0.1%: from left to right, IPTG = 4 µM, 6 µM and 200 µM.

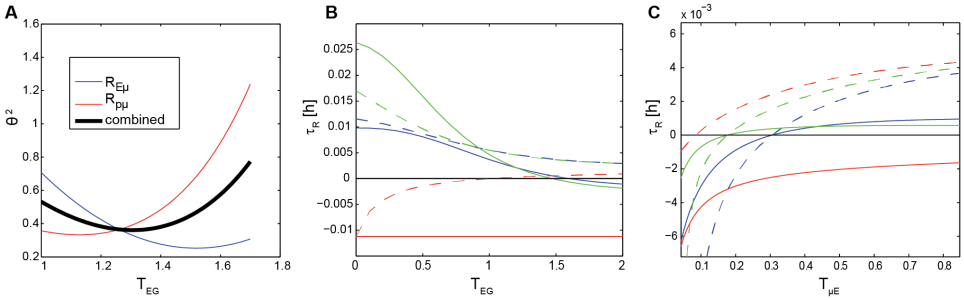


Figure 6.13: Robustness of the fitting. (**A**) θ^2 for different values of the fitting parameter T_{EG} at low induction, for the fit of $R_{E\mu}$ (blue), $R_{p\mu}$ (red) and their average (black). The fitting is optimal for $T_{EG} = 1.3$ and within the experimental error ($\theta^2 < 1$) (**B**-**C**) Symmetry of the cross-correlations was quantified by the value τ_R as described in the caption of Fig. 6.7, for $R_{E\mu}$ (full lines) and $R_{p\mu}$ (dotted lines) for the three induction levels (low in blue, intermediate in green and high in red) when varying the parameters T_{EG} (**B**) and $T_{\mu E}$ (**C**) . Symmetries are conserved for domains of the curves which are on the same side of the axis $\tau_R = 0$. In particular, the symmetry of all cross-correlations is stable for $T_{EG} = 1 - 1.5$ at all induction levels (model value 1.3), for $T_{\mu E} > 0.31$ at low induction (model value 0.7), $T_{\mu E} > 0.18$ at intermediate induction (model value 0.5) and $T_{\mu E} < 0.08$ at high induction (model value 0).

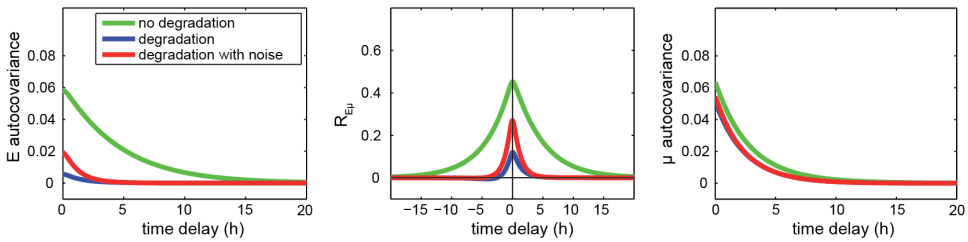


Figure 6.14: Model prediction for the effect of active degradation of *lac* enzymes with a timescale of 60 min during growth on lactulose with 6 µM IPTG with an average doubling time of 140 min. We compare the cases when there is no active degradation and active degradation taking or not into account the possibility of additional noise specifically due to degradation (we assumed that degradation noise was of same amplitude than production noise).

6.3.4. ADDITIONAL TABLES

Lactulose experiments	Noise observed in	Transmitted from E
low induction	p	12%
(Iptg=4 µM)	E	34%
	μ	31%
intermediate induction	p	9.80%
(Iptg=6 µM)	E	19%
	μ	20%
constitutive gene	p^*	13%
(Iptg=6 µM)	E^*	<1%

Table 6.1: Contribution of noise transmitted from *lac* concentration *E* to different variables in various culture media. The contribution of noise transmitted from E was computed by comparing the coefficient of variation of a given variable with or without transmission from E, using the values fitted with the model. Note that a decomposition of noise as a sum of coefficient of variations is not possible here, given the feedback of E on itself, which leads to self-sustained fluctuations which impact the noise intensity in a non-additive way.

Strain	Genotype	Origin
AB460	Δ*lacA::gfp-cat*	Constructed by A. Böhm
ASC631	Δ*lacA::gfp-cat*, Δ*php::-mCherry-kan*^{R}	This study
ASC636	Δ*lacA::gfp-cat*, Δ*cheZ::-mCherry-kan*^{R}	This study
ASC638	Δ*cheZ::-gfp-kan*^{R}	This study
ASC639	Δ*lacA::gfp-cat*, Δ*lacI::kan*^{R}	This study
ASC662	*lacZ-gfp*	This study
ASC640	Δ*lacA::gfp-cat*, Δ*php::Bla-mCherry-kan*^{R}	This study
ASC644	Δ*lacA::gfp-cat*, Δ*melA::mCherry-kan*^{R}	This study
ASC666	*L31-mCherry-kan*^{R}, *gltA::gfpA206K-cat*	This study
ASC677	*L31-mCherry-kan*^{R}, *pfkA::gfpA206K-cat*	This study
ASC678	*L31-mCherry-kan*^{R}, *icd::gfpA206K-cat*	This study
MG22	Δ*intC* ::$P_{L-lacO1}$-*yfp*	Elowitz lab
NCM520	Δ*lacAYZ*	Coli Genetic Stock Center

Table 6.2: List of strains used in this study. Construction of strains is described in chapter 2.2.1.

	lactulose + 4 μM IPTG	lactulose + 6 μM IPTG	lactulose + 200 μM IPTG	acetate	lactose, P_{N25} promoter	lactulose + 6 μM IPTG (2colors)	acetate (2 colors)
	Fig. 6.2H	Fig. 6.2I	Fig. 6.2J	Fig. 6.3A	Fig. 6.3C	Fig. 6.3E	Fig. 6.10C
μ_0 (h^{-1})	**0.23**	**0.42**	**0.84**	**0.22**	**0.91**	**0.35**	**0.29**
$\beta_\mu = \beta_M$ (h^{-1})	0.33	0.58	3.23	0.45	1.82	0.71	0.58
β_E (h^{-1})	5.63	6.06	7.36	7	7	7	7
η_E	1.03	0.78	0.48	0.60	-	0.78	0.60
η_μ	0.16	0.12	0.03	0.07	0	0.12	0.07
η_G	0.22	0.19	0.22	0.24	0.17	0.19	0.24
T_{EE}	1	0.68	-0.23	0	-	0.68	0
$T_{\mu E}$	0.7	0.5	0	0	-	0.5	0
T_{EG}	1.3	1.3	1.3	1.3	-	1.3	1.3
η_F	-	-	-	-	0.3	0.3	0.3
T_{FE}	-	-	-	-	0	0.5	0
T_{FG}	-	-	-	-	0.7	0.7	0.7

Table 6.3: Parameter values used to fit the experimental data. Values are for growth on lactulose at different IPTG concentrations (first 3 columns corresponding to model fits in Fig. 6.2H-J) and for the predictions of the rewiring experiments (Fig. 6.3A,C,E and Fig. 6.10C). Population average growth rate is measured and indicated in bold. Grey boxes correspond to parameters inferred from other experiments with similar conditions, as explained above. The noise sources amplitudes are expressed as the standard deviation of the random variable N_X which corresponds to $\eta_X = \theta_X / \sqrt{2\beta_X}$.

Daan Kiviet and Sander J. Tans conceived and designed the experimental approach. The experiments were performed by Daan Kiviet, Philippe Nghe, the author (especially central metabolic genes, noise transmission to constitutive genes) and Sarah Boulineau. Vanda Sunderlikova designed and constructed the strains. Philippe Nghe developed the theoretical model. Daan Kiviet, Philippe Nghe and Sander J. Tans wrote the published manuscript.

7

INFLUENCE OF RIBOSOME EXPRESSION DYNAMICS ON CELLULAR GROWTH

In this chapter we study the influence of fluctuations in ribosome content on cellular growth. Ribosomes are the machines that produce proteins and are highly abundant in cells. On the population level, their concentration appears to be precisely tuned to the cellular growth rate. Their role within the cell is central because cells cannot grow faster then ribosomes replicate themselves (and other components). We aimed to investigate whether transient fluctuations away from the average concentration render ribosomes growth-limiting and whether ribosomes are tuned to an optimal level. Therefore, we first characterize different fluorescent reporters to monitor ribosome production. Then, we determine temporal cross-correlations and the results suggest that ribosomes are not dynamically limiting growth rate. Neither do ribosomes appear to fluctuate around a clear optimal concentration. The results of this study are of preliminary nature and we conclude the chapter with suggestions for future research directions.

7.1. INTRODUCTION

Ribosomes are a central component of every living cell: They translate the genetic information encoded in mRNA into proteins, which then carry out virtually every function in a cell [183]. A ribosome is a complex machine consisting of ribosomal RNA and many proteins and is present in thousands of copies in a bacterial cell [1]. The number of required ribosomes increases strongly with growth rate [1, 53, 55] but producing these large numbers is costly. Therefore, on the population level ribosome synthesis is strictly regulated and precisely adjusted to the need imposed by the environment [184, 185]. But just as all other cellular components ribosomes are likely subject to noise and their concentration or efficiency may dynamically fluctuate in single cells. Considering their central role in cellular growth, any fluctuation away from the mean concentration could have an effect on growth rate. Indeed, it is commonly believed that fluctuations in ribosome content directly affect gene expression ("extrinsic noise") and thus may cause fluctuations in biomass generation (growth) [5, 123].

In this chapter we aim to address the question whether fluctuations in ribosome content dynamically limit growth of single cells. In other words, are ribosomes a bottleneck and do cells that temporarily have a higher ribosome concentration also grow faster? Or is the opposite true and these cells only feel the extra production burden while they cannot use the extra translational capacity, resulting in transiently slower growth? Related to this question is a third scenario: The mean ribosome concentration could be optimal and yield maximal growth while both up- and down-fluctuations could reduce growth rate. Moreover, it is an open question whether cells dynamically regulate ribosome content and thus counteract fluctuations to achieve growth homeostasis. To address these questions, we constructed a fluorescent reporter for ribosomes and measured growth and expression fluctuations over time. We then used amongst others cross-correlation analysis (chapter 2.3.6) to test whether ribosome fluctuations might transmit to growth rate. Limitations would manifest themselves as delay between ribosome production and growth rate fluctuations. This project is work in progress and we here report preliminary results and give an outlook on potential future directions.

Designing relevant experiments and interpreting data requires detailed understanding of ribosome production and its regulation. Therefore, we first give a brief introduction to ribosome composition and functionality and discuss regulation of ribosome synthesis. We also review how ribosome concentration and growth rate are coupled on the bulk level, then introduce our reporter strains and subsequently describe our results.

7.1.1. RIBOSOME BIOGENESIS AND FUNCTIONALITY

E. coli ribosomes are composed of a large 50S and a smaller 30S subunit [183] (Fig. 7.1). Together, they consist of three stable ribosomal RNAs (rRNA) and 54 ribosomal proteins (r-proteins) [183]. Around 65% of a ribosome is rRNA and around 35% are r-proteins [186]. rRNA is transcribed from seven operons which are all very similar, highly accessed and located close to the origin of DNA replication [184, 187]. Still, these operons are not transcribed at maximum capacity and only five of the seven operons are needed (then transcribed at maximum capacity) to produce the required rRNA and to yield near maximal growth rate [187]. This means that expression from rRNA promoters is gene-

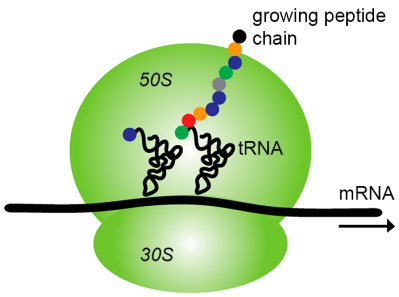


Figure 7.1: Illustration of a translating ribosome. The ribosome (green) consists of a small (30S) and a large subunit (50S), both of which contain rRNA and r-proteins. During translation, it moves along an mRNA and converts each nucleotide triplet into an amino acid, which is added to the growing polypeptide chain (protein). tRNAs provide the "nucleotides-amino-acid" pairs. Image was taken from [199] and labels were adjusted.

dosage compensated [188]. Genes encoding r-proteins are arranged in 19 operons of variable size, ranging from one to eleven genes per operon [189, 190]. To form a mature ribosome, the rRNAs need to be modified and fold and all r-proteins need to bind in the correct order [184]. Ribosome assembly takes around 1-4 minutes [191, 192].

For translation, the ribosome binds mRNA at the ribosome binding site (RBS) and starts decoding at the start codon, which is a specific nucleotide triplet [184]. Each codon is recognized by a transfer RNA (tRNA) which is charged with the encoded amino acid (Fig. 7.1). The tRNA binds the mRNA, adds the amino acid to the growing peptide chain and then unbinds [184]. This process is repeated until the stop codon is reached. rRNA performs the core catalytic functions during translation while r-proteins have rather assisting functionality [193, 194].

Ribosomes are not distributed uniformly throughout the cytoplasm but avoid the nucleoid region and are mostly located at the cell poles [195, 196]. In growing cells, 80% of all ribosomes are actively translating, independent of the growth rate [1, 184]. An exception are very slow growth conditions where the inactive fraction might be increased [197]. Ribosomes that are not active might still be in the final stage of de novo assembly [1] or might be in the process of being reactivated (repaired) [198]. The elongation rate during translation is independent of growth rate [53, 55, 197]. Together with the constant fraction of active ribosomes, this implies that the total rate of protein synthesis can only be changed by changing the number of ribosomes in a cell.

RELEVANCE FOR THE PROJECT

Since ribosomes consist of many components, it is not trivial to decide which part to label with a fluorescent reporter. One needs to choose whether expression of r-proteins or rRNA should be tracked and of which representative precisely within these classes. Ribosomal proteins can be expressed as translational fusions, but location within the operon and on the ribosome should be considered to retain functionality. rRNA expression can be monitored by introducing a promoter reporter. The fast assembly of the ribosomes suggests that expression fluctuations of one component will be a good proxy for fluctuations in the number of assembled, functional ribosomes.

7.1.2. REGULATION OF RIBOSOME SYNTHESIS

TRANSCRIPTION AND TRANSLATION

Producing large amounts of ribosomes is costly, as is translation of unnecessary proteins by overabundant ribosomes. Therefore, ribosome synthesis is strictly controlled. We here give an overview of regulation, for more detailed reviews on ribosome regulation see for example [184, 185], and on ppGpp (see below): [200, 201].

Expression of rRNA is regulated on the transcriptional level and r-protein synthesis is controlled during transcription and translation [202]. An important regulator of rRNA transcription is a small molecule called ppGpp (Fig. 7.2A) [201, 203]. It binds to RNA Polymerase (RNAP) and inhibits transcription from ribosomal RNA promoters [204]. The effect of ppGpp is potentiated by the protein DksA which binds to RNAP and modifies it [205, 206]. Additionally, rRNA expression is regulated by the transcription factor Fis [207] and the concentration of initiating nucleoside triphosphate (iNTP) [208]. iNTP is especially important during outgrowth from stationary phase and Fis levels change during transitions to and from stationary phase, but ppGpp dominates regulation during nutrient shifts and exponential growth [209].

Expression of r-proteins is balanced with rRNA expression by the following translational feedback mechanism (Fig. 7.2A): When r-proteins exist in excess and free rRNA is absent, many of these proteins bind with lower affinity to their own mRNAs [197]. This prevents the mRNA from being translated and additionally enhances its degradation [190, 210]. For a long time it was thought that free rRNA concentration is the master regulator which controls expression of r-proteins [197, 210, 211]. Recently, it was however shown that additionally transcription of r-proteins is directly regulated by ppGpp and DksA, similar to regulation of rRNA transcription [202]. It was suggested that these two mechanisms may complement each other: one could adjust the r-protein level upon severe up- or downshifts while the other mechanism might be responsible for fine-tuning within a steady state [202].

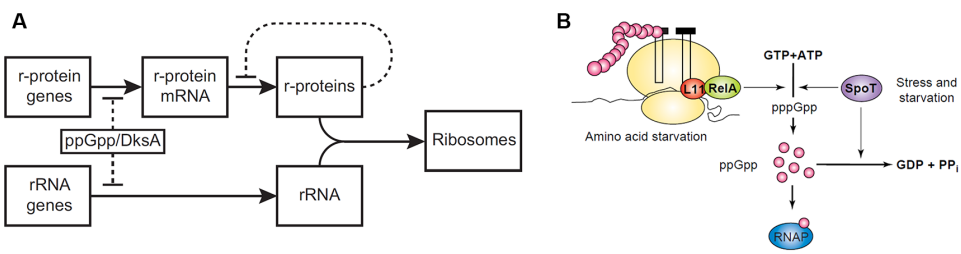


Figure 7.2: Regulation of ribosome synthesis. (**A**) Transcription of rRNA and of many r-proteins is regulated by ppGpp and DksA. Additionally, r-proteins are regulated on the translational level because excess r-proteins bind to their own mRNA and inhibit translation. (**B**) ppGpp is synthesized by RelA and SpoT and different stresses trigger ppGpp production by these enzymes. SpoT also hydrolyzes ppGpp. Image (**A**) was taken from [202] and (**B**) from [200].

THE SIGNALING MOLECULE PPGPP

The alarmone (p)ppGpp is synthesized by RelA and both synthesized and hydrolyzed by SpoT [212, 213] (Fig. 7.2B). RelA senses amino-acid starvation and is bound to the ribosome when non-active [214, 215]. When an uncharged tRNA enters the ribosome, RelA starts synthesizing ppGpp and thus triggers the "stringent response" [214]. Increased ppGpp levels then downregulate ribosome production (see above) and upregulate for example expression of amino acid synthesis enzymes [211]. The synthase activity of SpoT remained unknown for a long time but recently it was found that synthesis is triggered by multiple stresses. Among them are carbon starvation [216], fatty acid limitation [217], phosphate starvation [218], but not amino acid starvation (for a review see [201]).

SpoT always synthesizes a basal level of ppGpp, also in non-stressed conditions, and the basal ppGpp concentration depends inversely on the growth rate [1, 185, 213, 219]. As we will see in more detail in the next section, ribosome concentration increases with increasing growth rate ("growth rate control") and much research was done to find the mechanism of that control [219–221]. Now, it is thought that the basal ppGpp level sets the ribosome concentration [221]: The better the growth medium, the lower the ppGpp level and therefore the higher the expression from ribosomal promoters. It was recently also shown that regulation of ribosome synthesis by ppGpp is sufficient to robustly optimize growth rate because it balances supply and demand of amino acids [222]. For completeness, it should be noted that ppGpp regulates much more than ribosome synthesis: It upregulates expression of the general stress response sigma factor (rpoS), expression of genes necessary for survival in stress conditions, expression of biosynthetic enzymes, etc. [200, 201]. ppGpp is a key regulator during entry into stationary phase [201] and can convey persistence during antibiotics treatment [10, 223].

RELEVANCE FOR THE PROJECT

It should be checked that translational fusions do not interfere with the translational feedback mechanism (e.g. do not affect degradation). Further, fluctuations in r-proteins could lag behind rRNA fluctuations because of that regulation mechanism.

The role of ppGpp was so far only investigated in steady state and under severe stresses. It would be interesting to test whether ppGpp is also important during dynamic fluctuations and for example mediates a balance between ribosome content and growth. It is unclear whether SpoT or RelA would be the more relevant ppGpp source during dynamic adjustments and whether the relevance depends on the growth medium (for example with/without supplied amino acids). Further, externally manipulating the ppGpp level could potentially be used to alter ribosome concentration.

7.1.3. CORRELATION OF RIBOSOME CONCENTRATION AND GROWTH RATE AT THE BULK LEVEL

To interpret how fluctuations of ribosomes are related to growth rate fluctuations, it is useful to know how their averages (in different environments) are coupled. Therefore, we here summarize literature on bulk measurements: When growing cells in different media, the concentration of ribosomes increases strongly and linearly with the growth rate (Fig. 7.3A) [1, 53, 55]. As mentioned above, ppGpp is thought to be the regulator of

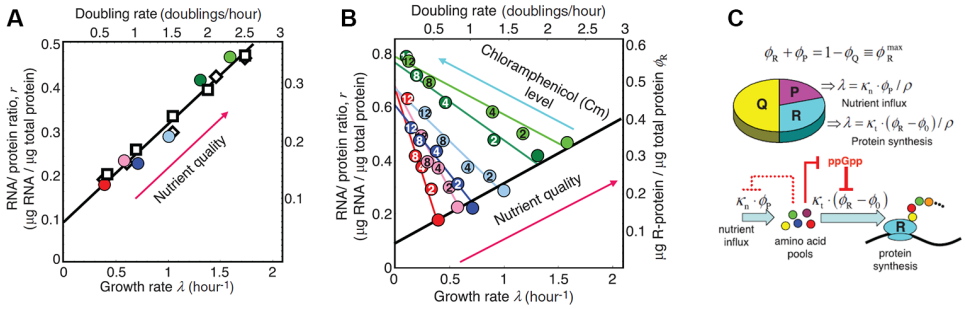


Figure 7.3: Scaling of ribosome concentration with growth rate. (**A**) Ribosome concentration (measured as RNA/protein ratio) depends linearly on growth rate. Different symbols denote different strains, the color corresponds to different growth media (displayed for one strain). (**B**) Ribosome concentrations for various growth media (color) and chloramphenicol concentrations [Cm] (numbers, in µM). At higher [Cm] the translational efficiency decreases, therefore the ribosome concentration is raised and growth rate is lower. (**C**) The linear correlations can be explained with a resource allocation model. Upper panel: The proteome is divided into a ribosomal fraction ϕ_R and two non-ribosomal fractions (ϕ_P and ϕ_Q). Lower panel: Optimal fractions are determined by balancing fluxes and depend on the growth rate. Figures were taken from [53].

this growth rate control [221] by coupling ribosome synthesis to supported growth rate. Transcription rate of rRNA promoters thus increases strongly with growth rate. For r-proteins, however, differential transcription (relative to total mRNA synthesis) is rather independent of growth rate and the linear increase is controlled by increasing translation [210]. The total number of r-proteins per cell ranges from around 7,000 in slow growth conditions to 70,000 at maximum growth, which corresponds to a fraction of 9% resp. 21% of the proteome [1]. In most experiments, the fraction of r-proteins was determined by indirect methods (e.g. inferring from rRNA concentration), but recently the Weissman group directly measured synthesis rates in different conditions and results are consistent [224]. The fraction of synthesized stable rRNA and tRNA relative to total RNA is high and ranges from 40% during slow growth to 85% during fast growth [1].

Few years ago, the Hwa group revisited this growth rate scaling in the light of optimal resource allocation [53]. We here give a brief summary (see Fig. 7.3): The ribosome concentration scales with growth rate and scaling is independent of the bacterial strain and precise medium composition (Fig. 7.3A). This linear correlation suggests that ribosomes are growth limiting [53], or, in other words, that a certain growth rate could not be achieved if ribosome concentration was below the line in Fig. 7.3A. We note that the medium-independence was challenged in [225] which argued for a growth *medium* dependence (minimal medium vs. added amino acids). The slope in Fig. 7.3A is determined by the translation rate per ribosome and the offset by the fraction of inactive ribosomes. Decreasing the translation efficiency by e.g. adding antibiotics or genetic manipulation is compensated with an increased ribosomal fraction (Fig. 7.3B). This is at the cost of synthesis of other proteins and results in a lower growth rate. The Hwa group showed that ribosomal fraction of the proteome can be predicted with a resource allocation model (Fig. 7.3C). A key ingredient is the balancing of nutrient influx on the

one side and amino acid requirements for protein synthesis on the other side. Given a constant ribosome efficiency [1], the protein synthesis flux can then only be adjusted by changing the ribosomal fraction (ϕ_R). For example, in rich media less resources need to be devoted to synthesis of precursors. Thus, more ribosomes can be produced to process these precursors, that is, the ribosomal fraction increases.

WHAT LIMITS THE GROWTH RATE OF E. COLI?

The maximum growth rate of *E. coli* is bounded due to the finite ribosome elongation rate. Cells cannot grow faster than ribosomes can synthesize a copy of themselves. In the theoretical extreme case where this is their only task, they would need around 6 minutes for this copy to be produced [53]. In reality, many more proteins are required in a cell and their copies need to be produced as well, which increases the workload of each ribosome. Taking that "replication burden" into account, the maximum growth rate was estimated to be a doubling time of around 20 minutes [53].

Apart from that extreme limit, ribosomes are often thought to be growth limiting in a given environment [53]. That is, cells cannot grow faster without increasing their ribosomal fraction. However, at the same time there exists an optimal ribosome fraction for each growth medium and cells cannot grow faster by simply increasing that fraction [53]. The reason is that the increase would be at the cost of other enzyme concentrations, which would then limit metabolite fluxes. Whether similar arguments hold for dynamic fluctuations away from the mean value is unknown and the topic of this chapter.

Interestingly, other work suggests that not ribosome capacity but supply of amino acids is rate-limiting in minimal medium [226]. Therefore, the actually growth-limiting process might be different in minimal medium and in complex medium with supplied amino acids (see also Outlook).

7.1.4. GENETIC CONSTRUCTS USED

To monitor fluctuations in ribosome content we constructed translationally fused reporters for two different r-proteins and a promoter reporter for rRNA expression (see Table 7.1). We chose the large subunit proteins L19 and L31 which were successfully labeled

Strain	Genotype	Remarks
ASC656	*L31-mCherry-kan*^{R} (no linker)	
ASC657	*L19-mCherry-kan*^{R} (no linker)	
ASC680	*L31-mCherry-kan*^{R} (no linker), Δ*cheZ::P*$_{rrn}$*-GFP -cat*^{R}	
ASC779	Δ*cheZ::P*$_{rrn}$*-GFP*	not used
ASC810	*L31-mCerulean-kan*^{R}	

Table 7.1: Ribosome constructs. The study was started with ASC656, ASC657 and ASC680 and we later switched to ASC810. Strain ASC779 was tested but showed unexplainable low expression and slow growth and was therefore discarded. GFP is monomeric in all constructs (A206K mutation), as are mCerulean and mCherry. Strain ASC810 contains a standard GC-linker between L31 & mCerulean, which is the linker that was used for most other constructs in this thesis as well.

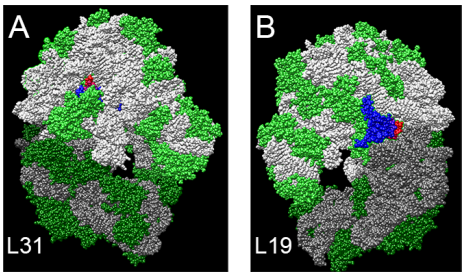


Figure 7.4: Labeled ribosomal proteins. Ribosomal proteins L31 (**A**) and L19 (**B**) , which were fused to a fluorescent reporter in different constructs, are shown in blue-red. The reporter gene was fused to the C-terminal (red). Color code: green: r-proteins, gray: rRNA, light color: 50S subunit, dark color: 30S subunit. The tunnel through which mRNA proceeds is visible as black hole. Structure PDB files are 1vs5 and 1vs6 and obtained from http://www.ebi.ac.uk/. UCSF Chimera (http://www.cgl.ucsf.edu/chimera/index.html) was used for visualization.

before [215] (Fig. 7.4). Their C-terminal, to which the fluorescent gene was fused, is at the ribosome surface and accessible. L31 (*rpmE*) is non-essential, only loosely attached to the ribosome and expressed from a one-gene operon. L19 (*rplS*) is essential, located at the interface between the two ribosomal subunits and expressed as last gene from a four-gene operon (www.ecocyc.org). The proteins were originally fused to mCherry (without linker). Due to some issues (see next section) we later constructed a L31-mCerulean fusion that also contained a linker between the proteins. The promoter reporter for rRNA expression (P_{rrn}) [102] was chromosomally inserted at the *cheZ* location and controlled expression of GFPmut2. Note that likely only full-length rRNA or complete ribosomes can feedback on rRNA expression and convey dosage compensation [188]. Therefore, we do not expect the reporter to affect expression from other rRNA promoters.

7.2. GENERAL CHARACTERIZATION OF R-PROTEIN-MCHERRY AND RRNA REPORTERS

7.2.1. GROWTH RATE AND RIBOSOME LOCALIZATION PATTERN

Before investigating questions on dynamic fluctuations, we characterized general properties of the obtained constructs to assess their quality. Labeling a protein could compromise its function and correct insertion into the ribosome. The high copy number of additionally produced fluorescent proteins, both as fusion and as promoter reporter, may introduce a growth burden. Especially mCherry expression (from a plasmid) has been reported to create a growth defect [112]. Therefore, we first checked key characteristics of the constructs: First, is the growth rate similar to wild type (WT) and second, can labeled r-proteins be inserted into the ribosome? The latter can be tested by observing the localization pattern: Ribosomes are enriched at the cell poles due to nucleoid exclusion and the fluorescence signal should show the same non-uniform distribution. We also compared properties between the L31-mCherry and L19-mCherry strain. The L31-mCerulean reporter, which was constructed much later, will be characterized in section 7.4. However, in a few cases (for example Table 7.2) a direct comparison was convenient and then results are already discussed here.

Strain	μ(dbl/hr) (microscope)	μ(dbl/hr) (plate reader)
ASC555 (WT)	1.0	not measured
ASC656 (L31-mCherry)	0.97	0.95
ASC657 (L19-mCherry)	0.95	0.87
ASC680 (L31-mCherry, P_{rrn}-GFP)	0.75	0.77
ASC810 (L31-mCerulean)	1.0	not measured

Table 7.2: Growth rates of ribosome constructs. Growth rates were measured on M9 + 0.2 mM uracil + 0.1% lactose. Microscopy growth rates are averages of multiple colonies. Plate reader growth rates were averaged over 3-5 wells (measured on the same 96 well plate). The slow growth of ASC657 in the plate reader is atypical because usually cells grow at least as fast in bulk compared to a surface (microscope). It could be due to stress or transient entry into stationary phase during preparation steps.

We found that coexpression of mCherry led to only a minor decrease in growth rate of <5%, despite the high copy number of mCherry proteins (Table 7.2). Since experiments in Table 7.2 were performed on different days, which can cause growth variability, it is not clear whether the decrease is statistically significant. The later used strain containing the L31-mCerulean fusion did not have any detectable growth defect compared to WT. In general, these high growth rates were a positive surprise because thousands of additional fluorescent proteins were produced in the reporter strains. However, the double-label strain with the additional P_{rrn}-GFP did grow slower. Depending on the growth medium, the growth decrease was around 15% (rich medium) to 25% (M9 with lactose, see Table 7.2). Tests with the single label strain P_{rrn}-GFP (ASC779) produced a growth rate similarly slow as with the double-label strain. This suggested that the specific expression from an rRNA promoter could cause the growth burden and not the additionally produced proteins per se (which should be much higher in the double label strain). For some constructs, we also measured bulk growth rate (Table 7.2) and these growth rates were very similar compared to microscopy data.

Next, we checked whether r-protein fusions were correctly inserted into the ribosome by examining the spatial distribution of the fluorescence signal. The reporter for rRNA naturally only reports on expression and is therefore uniformly distributed in the cell. We found that for all used r-protein fusions the fluorescence intensity was distributed non-uniformly: it was low at the nucleus location and high at the cell poles and the location between the nuclei (e.g. fast growing cells presumably contained multiple nuclei), see Fig. 7.5. This localization pattern is typical for ribosomes and was shown independent of fluorescent-protein techniques by for example cryofixation [227]. Therefore, the observation indicated that the fusion was properly incorporated into the ribosome. We observed very similar localization patterns for L31-mCherry, L19-mCherry and L31-mCerulean. Note that the segregation strength between ribosome and nucleus was stronger at fast growth compared to slow growth (Fig. 7.5), for which we do not know the reason. This observation seems however consistent with ref. [195] (their Fig. S11).

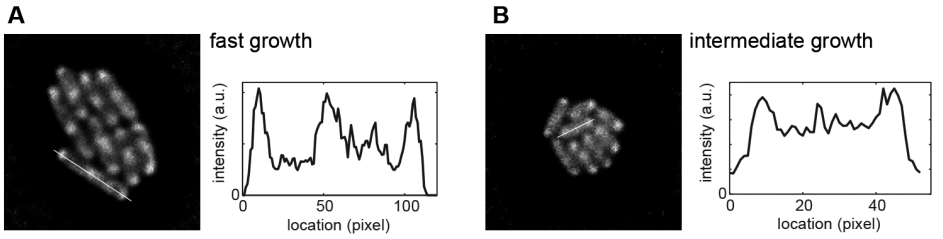


Figure 7.5: Localization pattern of labeled r-proteins. L31-mCherry localization pattern is displayed for growth on (**A**) rich medium (EZ RDM + 0.2% glycerol, μ≈2dbl/hr) and (**B**) minimal medium (M9 + 0.2mM uracil + 0.1% lactose, μ≈1dbl/hr). The intensity histograms are taken along the displayed lines. The reason for the stronger ribosome-nucleoid separation during fast growth is unknown.

7.2.2. MORPHOLOGY OF MICROCOLONIES

Despite this promising first characterization we unfortunately encountered several issues with the r-protein-mCherry fusions. First, for an unknown reason, the morphology of microcolonies was affected and cells suddenly seemed to repel each other (Fig. 7.6A). This effect was dependent on the growth medium and was often observed for growth on lactose (Fig. 7.6A) but never on acetate (Fig. 7.6B) or rich medium. We excluded changes in the glass and acryl gel surface as potential cause. Strains without labeled r-proteins almost never formed colonies with sparsely distributed cells. Growth rate was not affected by this colony morphology but cross-correlations were usually irreproducible and differed from results for dense colonies. We suspect that the apparent cell areas in the phase contrast images may play a role in the changed cross-correlations because isolated cells appear thicker. This could introduce artificial differences in measured fluorescence concentration for which cellular area is used. We can only speculate on the reason for this phenotype but it might be a stress reaction, cells could secrete some messenger molecules or the production of pili could be altered. The stress could be caused by misfolded mCherry proteins.

Since the rRNA reporter is in a background strain with r-protein-mCherry fusion, these experiments also produced the same open colony shape. The effect was even more severe for the double-label strain. In large contrast to that, the later used L31-mCerulean construct did not suffer from unexpected colony morphology and grew normally into a dense colony. Thus, the altered colony shape was likely due to either the choice of fluorescent protein or the absence of a linker between r-protein and reporter.

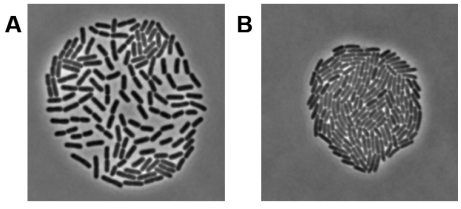


Figure 7.6: Morphology of microcolonies. (**A**) L31-mCherry cells were grown on M9 + 0.2mM uracil + 0.1% lactose. Cells seem to repel each other. (**B**) Same strain grown on M9 + 0.2mM uracil + 30mM acetate. Colonies were always dense.

7.2.3. SCALING OF FLUORESCENCE CONCENTRATION WITH GROWTH RATE

ONLY THE RRNA REPORTER CONCENTRATION SCALES WITH GROWTH RATE

We next tested whether the characteristic linear increase of ribosome concentration with growth rate (Fig. 7.3A) could be reproduced with our constructs. Therefore, we measured the fluorescence intensity for three different growth conditions, supporting growth rates from around 0.3dbl/hr to 2dbl/hr. We found that the concentration of L31-mCherry and L19-mCherry decreased slightly with increasing growth rates and both reporters produced very similar concentration values (Fig. 7.7A). The similar concentration supported the equivalence of the two different labels. However, the growth scaling strongly differed from the linear increase established in literature and was hence unexpected [1, 53]. Contrary to that, the concentration of GFP expressed from the rRNA promoter increased linearly with growth, in agreement with results reported literature [1, 53] (Fig. 7.7B). Therefore, P_{rrn}-GFP seemed to be a better reporter for ribosome expression. The properties of L31-mCerulean will be discussed in detail in section 7.4. For the remainder of this chapter we will refer to the linear increase of concentration with growth rate as "growth (rate) scaling".

WHY COULD GROWTH SCALING BE ABSENT?

There are two possible interpretations why r-protein concentration did not show growth rate scaling: First, the effect is real and r-protein concentration does not scale with growth rate. Second, the effect is an artifact and was in one way or the other introduced by mCherry.

Why could the lack of scaling be real and have been omitted in literature? It would be possible that at slow growth r-proteins are overproduced. Then, not all proteins could be incorporated into ribosomes but many would be freely diffusing. Thus, the r-protein concentration would be higher than the concentration of assembled ribosomes. Most studies on growth rate scaling did not measure r-protein concentration directly but mea-

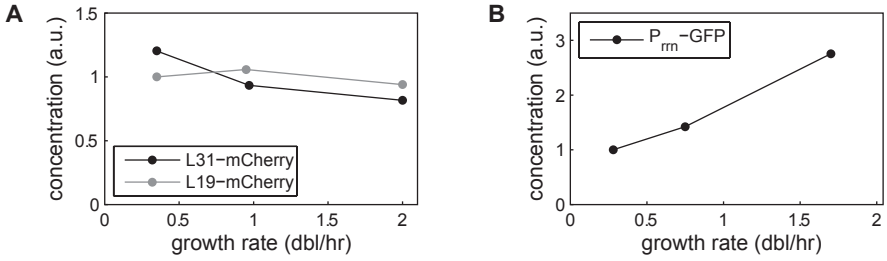


Figure 7.7: Dependence of measured ribosome concentration on growth rate. (**A**) Concentration of r-protein-mCherry fusions plotted against growth rate (ASC656 and ASC657). The concentration does not scale as expected. (**B**) Concentration of GFP expressed from a P_{rrn} promoter plotted against growth rate (ASC680). The linear correlation is as expected [53]. Each data point was obtained as average of ≥45 cells, which were pregrown in the respective medium and then imaged under the microscope. Growth rates were taken from time-lapse experiments with these strains in the respective media.

sured concentration of stable RNA and used it as proxy for ribosomes [53, 228]. Other studies which focused on stoichiometry of the ribosome typically used sedimentation techniques to obtain (complete) ribosomes, but could not detect freely diffusing proteins [229–231]. Also in newer super-resolution studies freely diffusing proteins would likely be too fast to detect [196].

However, r-protein synthesis rate was measured directly by Cole & Nomura [210] and recently by the Weissman group [224]. Both studies did find a strong increase of r-protein synthesis with growth rate, that is, growth rate scaling (see e.g. Table mmc1 in [224], which includes the r-proteins L31 and L19). Additionally, the scenario of overproducing r-proteins at slow growth would be inefficient for a cell, would deplete resources, and would contradict the translational feedback mechanism (section 7.1.2). Therefore, we suggest that the lack of scaling is most likely an artifact caused by mCherry-labeling.

One possibility is that mCherry interferes with the regulation of r-protein expression. It might affect the translational feedback mechanism and alter mRNA lifetimes: mCherry is a very large protein with 256 amino acids (AAs) compared to L31 (70 AAs) and L19 (115 AAs). Therefore, the secondary structure of mCherry-mRNA might sterically block r-proteins from binding their own mRNA and thus disable the translational feedback control. Growth rate control is mostly exerted on the translational level and a purely transcriptional reporter would decrease in concentration with increasing growth rate, similar to our observation (see Fig. 7.8 which is reproduced from [210]). Additionally, at least some mRNAs of r-proteins are degraded in the 3'-to-5' direction, which is unusual for mRNA degradation [232]. Thus, the mCherry mRNA, which is fused to the 3' end (the later C-terminal), could affect degradation rate. Taken together, we think that a likely scenario is abolishment of translational feedback by the mCherry fusion and the creation of an effective transcriptional reporter. Still, it remains unclear why two different fusions showed exactly the same altered growth rate behavior.

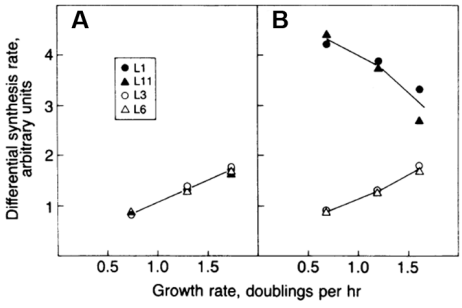


Figure 7.8: Growth rate dependence of transcription and translation of r-proteins. This figure shows work of Cole & Nomura [210]. Relative synthesis rate of four different r-proteins (see legend) was measured for different growth rates. L1 and L11 are in the same operon, which is translationally regulated by L1-binding. (**A**) Intact translational feedback mechanism. All proteins show growth rate scaling. (**B**) Translational feedback of L1 and L11 is disrupted by a mutation in L1. Then, relative synthesis rate decreases with increasing growth rate (black symbols), similar to our observation (Fig. 7.7A).

EFFECT OF MATURATION TIME

Additionally, the maturation time of mCherry, which seems to be slow at least in some conditions (chapter 3), may be play a role in the lack of growth rate scaling. If fluorophore maturation is slow, then the observed fluorescence intensity will be lower than

the actual number of proteins present because newly produced proteins are not visible yet. This effect is more severe at fast growth rates because proteins are then diluted faster. To quantify the impact of slow maturation times and to test if it can explain the lack in growth rate scaling, we modeled protein maturation as first order rate equation, following the derivations in [88, 109]. The total protein concentration E is the sum of mature, fluorescent proteins M, which can be measured, and immature, non-fluorescent proteins I, which cannot be measured: $E = M + I$. The time dynamics of I and M are:

$$\begin{aligned} \dot{I} &= \alpha - \mu_e I - \lambda_{mat} I \ , \\ \dot{M} &= \lambda_{mat} I - \mu_e M \ . \end{aligned} \tag{7.1}$$

Here, α is the translation rate, $\mu_e = \mu \cdot \ln 2$ the growth rate, computed with base e, and $\lambda_{mat} = 1/\tau_{mat}$ the maturation rate (base e). In words, immature proteins are produced by translation and disappear (are effectively degraded) due to maturation and growth. The mature proteins are produced by maturation and disappear due to growth.

In steady state, the time derivatives in eq. 7.1 are zero and the second equation determines the relative fraction of mature and immature proteins:

$$I = \frac{\mu_e}{\lambda_{mat}} \cdot M \tag{7.2}$$

The first equation determines the absolute protein concentration, which is of less relevance here. The total protein concentration as function of measured concentration is then

$$\begin{aligned} E &= \left(1 + \tfrac{\mu_e}{\lambda_{mat}}\right) \cdot M \\ &= \left(1 + \mu \cdot \tau_{mat} \cdot \ln 2\right) \cdot M \\ &\equiv s(\mu, \tau_{mat}) \cdot M \ , \end{aligned} \tag{7.3}$$

where $s(\mu, \tau_{mat})$ is defined as scale factor between measured and real protein concentration.

The measured concentration M of e.g. L31-mCherry was approximately constant for a wide range of growth rates (Fig. 7.7A). For a maturation time of 15min (see chapter 3.6) and constant measured concentration, the real concentration E would be around 30% higher at the fastest measured growth rate compared to the slowest growth rate (Table 7.3). This increase is by far not enough to explain the factor 3 difference in ribosome concentration for these growth rates (Fig. 7.7B, and Scott et al. [53]). Also a maturation time of 40 min would create less than a factor 2 difference for E during fast and slow growth. Only an unrealistically slow maturation time of 180 min could account for the absence of growth scaling (Table 7.3). Thus, we conclude that maturation times caused the measured ribosome concentration to increase somewhat less steeply with growth rate. However, this effect is likely not the main cause for the absence of growth scaling. Note that the estimates presented here assumed that maturation times are independent of the growth medium. If maturation was slower at slower growth rate (see section 7.3.1), its influence on growth scaling would be even weaker.

growth rate (dbl/hr)	τ_{mat} 5 min	15 min	40 min	180 min
0.35	1.02	1.06	1.16	1.73
1	1.06 (1.04)	1.17 (1.1)	1.46 (1.3)	3.08 (1.8)
2	1.11 (1.09)	1.35 (1.3)	1.92 (1.8)	5.16 (3.0)

Table 7.3: Scale factor between real and measured protein concentration. The scale factor $s(\mu,\tau_{mat})$ of eq. 7.3 is shown for the three growth rates of Fig. 7.7 and four different examples for maturation times. Small numbers in brackets are the rounded multiplicative factor between real protein concentration in the current and the slowest growth medium, when the measured protein concentration is the same in both media. τ_{mat} = 180min was chosen such that this factor is 3 between fastest and slowest growth, which would be required to explain the absence of growth scaling solely by maturation time (Fig. 7.7).

7.3. FLUCTUATIONS, LIMITATION AND OPTIMALITY

7.3.1. CROSS-CORRELATIONS AT DIFFERENT GROWTH RATES

From the previous section we concluded that both rRNA and r-protein reporters had advantages and disadvantages. For example, P_{rrn}-GFP caused a reduction in growth rate but displayed growth rate scaling as expected, while the opposite was true for r-protein fusions. Since no label was clearly better we continued using both labels when now turning towards the actual scientific question: Are fluctuations in ribosome concentration dynamically limiting growth rate? If so, do cells with transiently less ribosomes grow slower and with more ribosomes grow faster or does ribosome concentration fluctuate around an optimum and all deviating cells grow slower?

We first investigated whether ribosomes are a bottleneck that limits growth by testing whether fluctuations in ribosome content are transmitted to growth rate. To that extent, we calculated the cross-correlation functions $R_{p\mu}(\tau)$ between ribosome production rate p and growth rate μ, as well as $R_{E\mu}(\tau)$ between ribosome concentration E and growth rate. If, for example, ribosomes were growth limiting, then an up-fluctuation in ribosome production rate should be followed by an increase in growth rate, thus $R_{p\mu}(\tau)$ would be asymmetric and shifted towards $\tau > 0$. Since ribosome concentration varies strongly with the growth conditions, we tested for a wide range of growth rates (0.3-1.7dbl/hr) whether limitations and noise transmission are present.

RESULTS

Fig. 7.9 displays the cross-correlations $R_{p\mu}(\tau)$ and $R_{E\mu}(\tau)$ for the r-protein fusions L31-mCherry and L19-mCherry, obtained for three different growth rates. At slow growth the protein production rate and growth rate were weakly positively correlated (Fig. 7.9A). Surprisingly, the production rate fluctuations followed behind the growth fluctuations, with a delay of $\approx 0.7T_d$ (T_d=1/μ being the mean interdivision time). Such a delay was previously neither observed for constitutively expressed nor regulated genes (for an exception, see below). $R_{E\mu}(\tau)$, however, displayed a shape that we frequently observed for constitutive promoters ("dilution mode", chapter 6): Protein concentration lagged behind growth fluctuations and their correlation was strongly negative (≈-0.4). Cross-

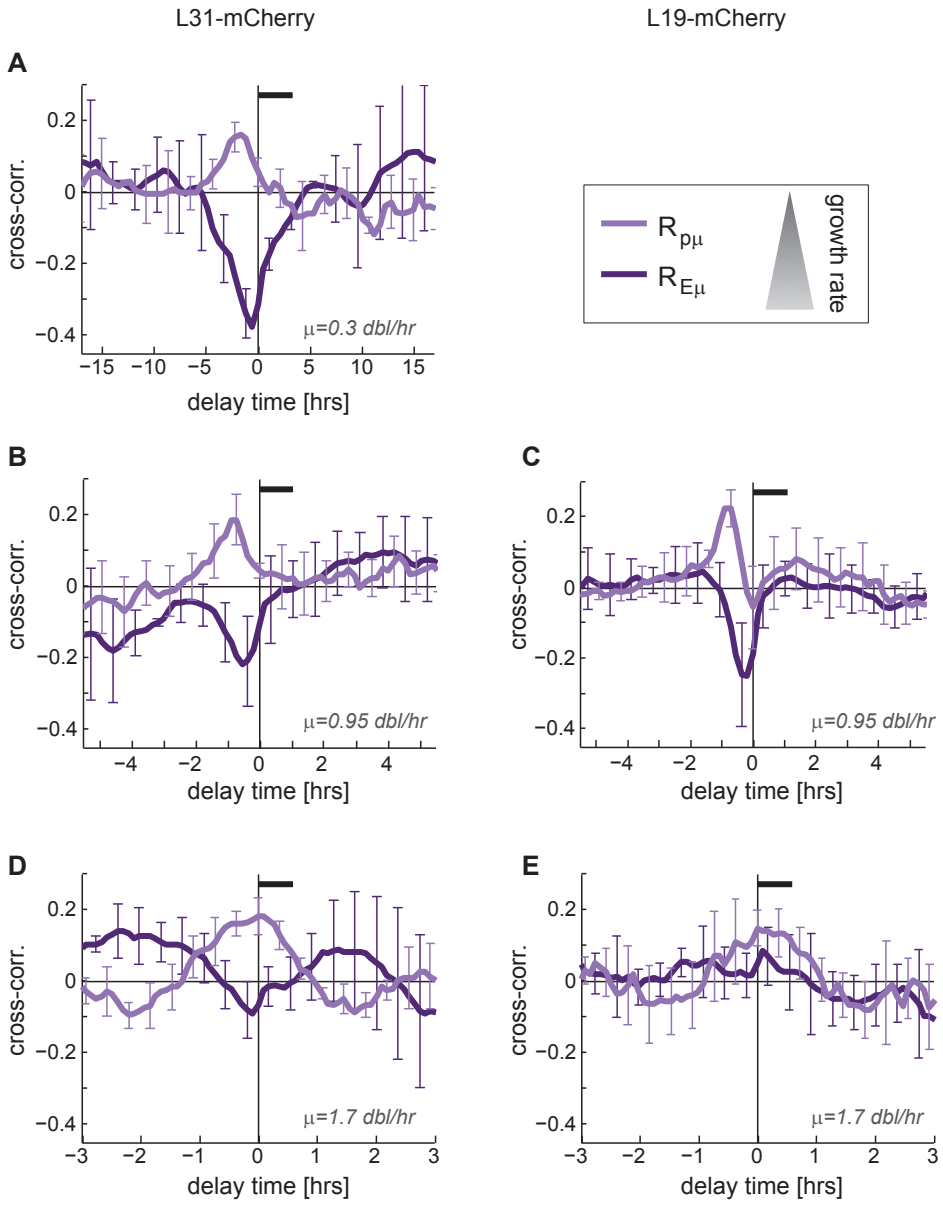


Figure 7.9: Cross-correlation of r-protein-mCherry with growth rate. Cross-correlations $R_{p\mu}(\tau)$ of protein production rate and growth rate (bright purple) and $R_{E\mu}(\tau)$ of protein concentration and growth (dark purple) are displayed for the L31-mCherry fusion (left) and L19-mCherry fusion (right). Average growth rate increases from top to bottom and the thick black lines indicate mean interdivision time. Error bars are standard deviations (see chapter 2.3.6), plotted for every fourth data point. (**A**) Growth on M9 + 0.2mM uracil + 30mM acetate. (**B**) and (**C**) Growth on M9 + 0.2mM uracil + 0.1% lactose. (**D**) and (**E**) Growth on EZ defined rich (which contains amino acids)+ 0.2% glycerol. Fluorescence concentration was determined by using the complete cell area instead of a center box because of the non-uniform protein localization (chapter 2.3.3). The cross-correlation shapes are likely an artifact of the mCherry label, see main text.

correlations were very similar at intermediate growth rates (Fig. 7.9B,C), even the time lag in $R_{p\mu}(\tau)$ scaled with growth rate. Only the negative correlation of $R_{E\mu}(\tau)$ was now less pronounced but still displayed the "dilution mode". At fast growth, when ribosomes constitute a major fraction of the cellular mass, the cross-correlations were qualitatively different (Fig. 7.9D,E): production rate and growth fluctuated (almost) simultaneously and were weakly correlated, while r-protein concentration and growth were practically uncorrelated. The two different L31-mCherry and L19-mCherry reporters produced identical cross-correlations (Fig. 7.9B-E) which increased confidence into the correctness of the signals. However, we note already here that although L31 and L19 fusions were also characterized above to be very similar, both lacked typical "ribosome behavior" such as scaling of concentration with growth rate (section 7.2) and might thus be similar but bad reporters.

Before interpreting these cross-correlations we first present the corresponding correlations of the rRNA reporter (Fig. 7.10). In slow growth conditions the rRNA cross-correlations differed strongly from the corresponding r-protein correlations: $R_{p\mu}(\tau)$ was now symmetric and the correlation strongly positive at $\tau = 0$ (Fig. 7.10A, bright line). The correlation between reporter concentration and growth rate had even switched sign (dark line). These correlations were quite similar to correlations observed for regulated but not limiting proteins, for example metabolic enzymes (chapter 6). The strong difference between the two reporter types was surprising since both were thought to be a proxy for the "same" ribosomes and we will investigate possible reasons for this difference further below. Cross-correlations for intermediate growth rate could not be measured due to the strange colony morphology. During fast growth, however, correlations were similar to results of the r-protein fusions (Fig. 7.10B): $R_{p\mu}(\tau)$ was symmetric and positive but small, while $R_{E\mu}(\tau)$ was indistinguishable from zero.

RIBOSOMES APPEAR NON-LIMITING

Despite the differences between the r-protein and rRNA cross-correlations, both data suggested that fluctuations in ribosomes are not dynamically limiting growth rate, not even in the fastest growth condition tested. We recall that if ribosomes were limiting, then the current ribosome concentration would dictate the current growth rate - $R_{E\mu}(\tau)$ should be positive and peaked at delay zero or close to zero. This was clearly not the case for the negatively correlated r-proteins. And while $R_{E\mu}(\tau)$ for rRNA was in fact positive in slow growth conditions, the time lag of the peak was large (hours). This strongly suggested that concentration and growth rate fluctuated jointly because of global noise sources ("common noise", see chapter 6), but not due to direct transmission of fluctuations. The non-limitation can also be seen in $R_{p\mu}(\tau)$: Neither fluctuations in r-protein production rate nor in rRNA production rate preceded fluctuations in growth rate.

TRANSLATIONAL REGULATION MIGHT CAUSE A TIME LAG BETWEEN REPORTERS

The different cross-correlations that were obtained with the seemingly redundant reporters for r-protein and rRNA expression were surprising. Especially striking was the observation that fluctuations in r-protein production rate lagged behind growth rate fluctuations (Fig. 7.9). Since the latter fluctuated in synchrony with rRNA production (Fig. 7.10), this meant that fluctuations in rRNA production were preceding fluctuations

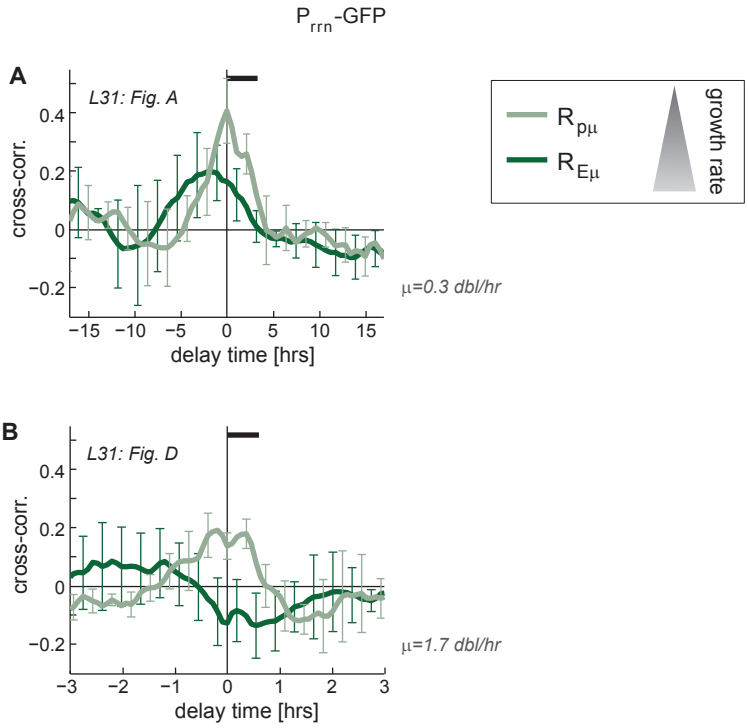


Figure 7.10: Cross-correlation of P_{rrn}-GFP with growth rate. Cross-correlations $R_{p\mu}(\tau)$ of protein production rate and growth rate (bright green) and $R_{E\mu}(\tau)$ of protein concentration and growth (dark green) are displayed for the ribosomal RNA promoter reporter P_{rrn}-GFP . The thick black lines indicate mean interdivision time. Both datasets were obtained with dual-label strains containing also L31-mCherry and corresponding r-protein cross-correlations (in Fig. 7.9), obtained from the same microcolony, are indicated in the figure. Error bars denote the standard deviation (see chapter 2.3.6) and are indicated for a fraction of the data points. (**A**) Growth on M9 + 0.2mM uracil + 30mM acetate. (**B**) Growth on EZ defined rich (which contains amino acids)+ 0.2% glycerol. Fluorescence concentration of this uniformly distributed reporter was determined with the standard procedure (center box, see chapter 2.3.3).

in r-protein production. As expression of rRNAs and r-proteins is co-regulated and these components are needed in a one-to-one stoichiometry during ribosome biogenesis, we would have expected synchronous fluctuations. Therefore, we investigated what biological process could cause such a delay and whether we can believe this delay to be real or an artifact caused by labeling.

If the time delay between rRNA and r-protein expression is real, then it must stem from a regulatory process which affects r-protein expression differently than rRNA expression. Transcription of both components is regulated very similarly and can thus be excluded. However, r-protein expression is additionally controlled during translation by the amount of free rRNA, and this control could cause a delay (Fig. 7.11): Con-

Figure 7.11: Translational regulation can cause a delay of r-protein production in $R_{p\mu}(\tau)$. rRNA production rate and growth rate are simultaneously affected by a common noise source (circle with tilde), in the figure a down-fluctuation is illustrated. The concentration of free rRNA then decreases because of dilution. This process is slow, with a time scale in the order of the doubling time. Free r-proteins do not find rRNA to bind any more, therefore bind to their own mRNA and inhibit translation (ribosome in gray). The slow dilution process of rRNA causes the delay in $R_{p\mu}(\tau)$ of r-proteins in this scenario.

sider for example a random down-fluctuation in rRNA production rate. Growth rate will down-fluctuate simultaneously (Fig. 7.10), and maybe also r-protein expression down-fluctuates to some extent. With a delay on the order of the interdivision (dilution) time free rRNA concentration decreases. Then, free r-proteins will repress their own translation due to lack of free rRNA and r-protein production down-fluctuates with a delay. The delay would be caused by the time needed for changes in rRNA production rate to significantly affect rRNA concentration.

To directly test this scenario, we measured a transcriptional mCherry reporter under control of the L31-promoter (P_{rpmE}, strain ASC789), which should not be affected by translational regulation. Unfortunately, the cross-correlations were ambiguous, showing different results for different microcolonies. The experiment was not pursued further because at this time issues with the mCherry label became evident (see below), and a repeat will only be useful with a better r-protein reporter.

As side remark, we also tested whether ppGpp, synthesized by RelA, might play a role in any of the correlations and differences observed. To this end, we constructed a knockout Δ*relA* in the L31-mCherry strain (ASC689, a knockout in the rRNA reporter background was not tested here but might be interesting). Cross-correlations, as well as average r-protein expression, were identical in the parent and the knockout strain. This suggested that fluctuations were either not mediated via ppGpp or that SpoT instead of RelA is the relevant ppGpp synthase in this context. Creating a *spoT* deletion is, however, not trivial because a single knockout is lethal and a double knockout of *spoT* and *relA* creates a viable, but sick phenotype.

R-PROTEIN-MCHERRY CROSS-CORRELATIONS MIGHT BE AN ARTIFACT

A different explanation for the observed cross-correlations and time lags is that they are artifacts produced by the r-protein-mCherry label. Indeed, mCherry has proven to be problematic in several cases: When for example placing mCherry under control of an exogenous, constitutive promoter, the correlation between protein production and growth rate sometimes, but not always, displayed very similar time delays (Fig. 7.12). The delay appeared to be larger at slower growth rate (was for example smaller in direct maturation time measurements at faster growth, chapter 3.6). This suggests that

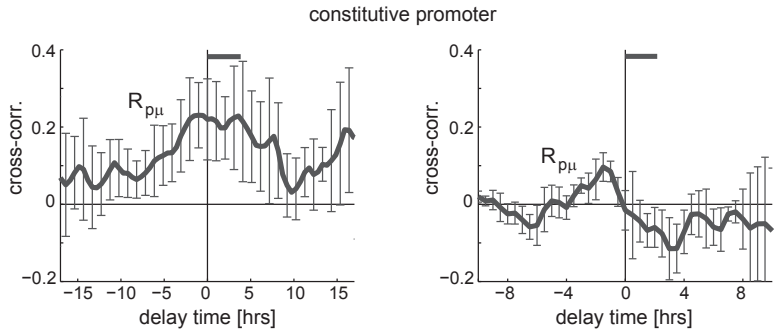


Figure 7.12: Cross-correlation $R_{p\mu}(\tau)$ of constitutively expressed mCherry. mCherry is expressed from an exogenous constitutive promoter (P_{bla}, ASC640). Two repeats of the experiment are shown and the thick bars indicate mean interdivision time. Error bars denote the standard deviation. Cells were grown on M9 + 0.2mM uracil + 0.1% lactulose + 6μM Iptg. The large change in growth rate is likely due to organic contamination of the gel pad used for the right dataset. We note that error bars are large and therefore cross-correlations are barely significant.

a slow maturation time could be the cause of the delay and thus cause differences in P_{rrn}-GFP and r-protein-mCherry signal. The maturation rate and therefore oxygen concentration would then need to be growth medium dependent (slower at slower growth) and would need to vary between experiments because the delay was not always present. This is in principle possible because external oxygen concentration could change with for example the number of applied cells. The internal oxygen reservoir might be dependent on cellular metabolism [112], which is itself dependent on the supplied carbon source.

A second issue was that, in general, cross-correlations of a variety of proteins (e.g. Icd) fused to a different fluorescent reporter (e.g. GFP) were difficult to reproduce when the reporter was exchanged for mCherry. Third, the lack of growth scaling investigated above could simply be due to the type of constructed fusion (e.g. no linker), but it cannot be excluded that it is due to an intrinsic problem of mCherry. Taken together, these problems motivated the systematic investigation of fluorescent proteins, which was presented in chapter 3. For this study here, we decided to construct a control fusion with a different fluorescent protein, L31-mCerulean, and to include a linker between the proteins (ASC810, Table 7.1). mCerulean was shown in chapter 3 to be a "good", recommended fluorescent protein. The results obtained with that construct will be presented in section 7.4.

7.3.2. LITTLE EFFECT OF ANTIBIOTICS ON LIMITATION

Translation inhibiting antibiotics render ribosomes less efficient and therefore limit the growth rate. Cells react to that perturbation by producing more ribosomes, at the cost of other proteins (Fig. 7.3B,C). We wondered whether, once cells are adjusted, ribosomes are still dynamically limiting growth or whether a new non-limiting equilibrium was found. To test this, we grew cells in rich medium at subinhibitory concentrations of

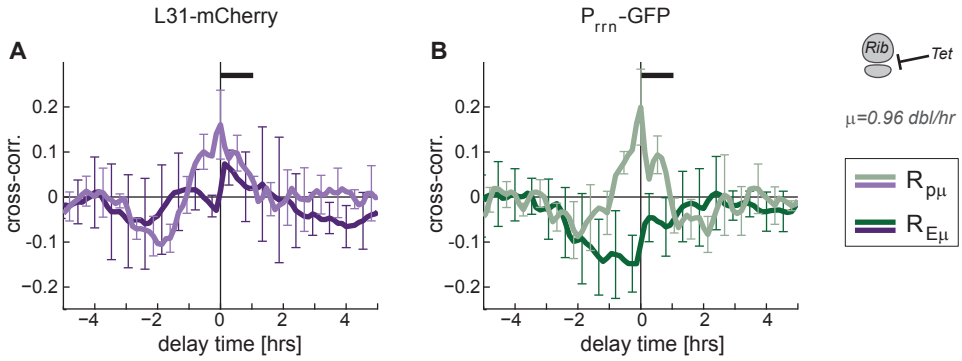


Figure 7.13: Cross-correlation of ribosomes and growth rate in the presence of antibiotics. Cross-correlations $R_{p\mu}(\tau)$ of protein production rate and growth rate (bright lines) and $R_{E\mu}(\tau)$ of protein concentration and growth rate (dark lines) are displayed for (**A**) the L31-mCherry fusion (purple) and (**B**) the P_{rrn}-GFP reporter (green). Error bars denote the standard deviation and are plotted for some data points only. Growth is in EZ defined rich + 0.2% glycerol + 0.5μM Tetracycline. The translationally inhibiting antibiotic doubled the interdivision time (thick black line). Corresponding cross-correlations without antibiotics are Fig. 7.9D and Fig. 7.10B.

Tetracycline. The antibiotic halved the growth rate, which was now quite similar to growth in minimal medium supplemented with lactose ("intermediate growth rate", for example Fig. 7.9B,C). We observed a strong upregulation of rRNA expression, but expression of r-protein-mCherry seemed not to be upregulated. While the general issues of mCherry were already discussed above, it may still be of interest to compare this dataset to cross-correlations obtained with the same reporter.

Both $R_{p\mu}(\tau)$ of the P_{rrn}-GFP reporter and r-protein-mCherry was slightly positive and peaked at a delay of zero (Fig. 7.13). Concentration fluctuations were not significantly correlated with growth (r-proteins, Fig. 7.13A), or negatively correlated and delayed to negative τ (rRNA, Fig. 7.13B, consistent with "dilution mode", see chapter 6 and Fig. 6.2). These cross-correlations suggest that ribosomes are not dynamically limiting growth rate, because $R_{E\mu}(\tau)$ is not positive. Cells seem to adjust to a new non-limiting steady state under the harsher environment. Interestingly, the cross-correlations resemble the correlation curves obtained in the same medium without antibiotics, that is at much faster growth rate (Fig. 7.9D,E, Fig. 7.10B). They differ from correlations in minimal medium supporting a similar growth rate (Fig. 7.9B,C). This suggests that the precise growth *medium* may be relevant when studying ribosomes and fluctuations.

7.3.3. RIBOSOMES DO NOT SEEM TO FLUCTUATE AROUND A NARROW OPTIMUM

We next tested whether the ribosome level was at an optimum, such that both up- and down-fluctuations would lead to a decrease in growth rate. Contrary to the scenario of "limiting ribosomes", cells expressing more ribosomes would not simply grow faster. The relation between growth and expression would be non-monotonous, peaking at an inter-

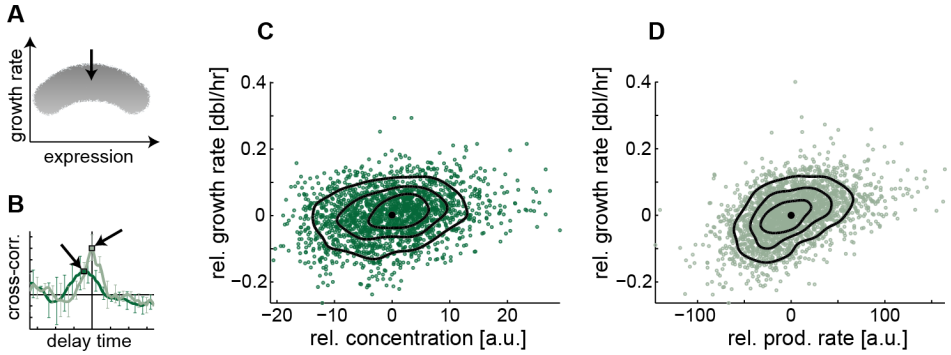


Figure 7.14: Scatter between rRNA expression and growth does not indicate an optimal expression level. (**A**) If ribosomes are tuned to an optimal level (arrow) both up- and down-fluctuations should decrease growth rate. Cross-correlations are not suitable to detect such non-linearity. (**B**) Cross-correlations $R_{p\mu}(\tau)$ (bright green) and $R_{E\mu}(\tau)$ (dark green) between P_{rrn}-GFP and growth rate. Image is identical to Fig. 7.10A. Growth was on M9 + 0.2mM uracil + 30mM acetate (0.3dbl/hr). A scatter plot as illustrated in (**A**) is determined for the time lag displaying the highest correlation coefficient (squares). (**C**) Scatter between protein concentration and growth resp. (**D**) protein production and growth at time lags indicated in (**B**) . Displayed quantities are mean subtracted. Using a different time lag produced very similar scatter plots. Lines are contours of equal probability, obtained by kernel density estimate. Growth rate is not maximal at an intermediate ribosome level. Script for plotting was developed by Martijn Wehrens.

mediate expression level (Fig. 7.14A). Cross-correlations cannot detect non-linearities, and we therefore plotted the full distribution of growth rates against (rRNA) expression (Fig. 7.14C,D). Arguably, ribosome concentration may be the more relevant parameter, compared to production rate, to test whether growth rate is maximized at an intermediate value. Since concentration fluctuations are, however, complex (resulting from production and dilution fluctuations), we analyzed both concentration and the more direct measure production rate.

We found that growth rate increased monotonously with ribosome concentration, as well as with ribosome production rate (Fig. 7.14C,D). Therefore, the data suggests that ribosomes do not fluctuate around a narrow peak. The correlation between concentration and growth was, however, very small (Fig. 7.14C), suggesting that ribosome concentration has low control on growth rate. This would be consistent with a wide range of optimum ribosome levels (flat peak). We note that fluctuations of the observables can occur with a time delay relative to each other, leaving the chosen time difference between variables as free parameter when creating a scatter plot. The displayed figures show the distribution of growth in dependence of expression at the time delay for which the cross-correlation peaked (Fig. 7.14B). Different delays led to very similar results.

Conceptually, fluctuations around a steady state are fundamentally different to varying the steady states. Thus, this argumentation based on bulk level results for different steady states may have been too simply to transfer directly to dynamic fluctuations and adjustments may be needed (see also Outlook 7.6).

7.4. THE NEW R-PROTEIN-MCERULEAN REPORTER: CHARACTERIZATION AND RESULTS

7.4.1. GENERAL CHARACTERIZATION AND GROWTH SCALING

Since issues with the r-protein-mCherry fusions became apparent (see above), we decided to construct a control fusion L31-mCerulean (with linker, ASC810). In the simultaneously conducted study on fluorescent proteins mCerulean was shown to give good and reproducible results (chapter 3). The L31-mCerulean strain was growing at full wild type growth rate, labeled proteins localized in the cytoplasm as expected (suggesting correct insertion into the ribosome) and the colony morphology was normal (section 7.2).

The concentration of L31-mCerulean proteins increased strongly with growth rate (Fig. 7.15), in contrast to the data on mCherry (Fig. 7.7A). Such an increase is expected from growth scaling reported in literature and supported the quality of the new reporter. We note that on a quantitative level, the increase with growth would be expected to be larger (similar to rRNA data in Fig. 7.7B). Potential reasons are that experiments were performed on different days with differently aged lamps or that maturation time still decreases the apparent concentration at very fast growth (T_d=24min). Taken together, we concluded that the new L31-mCerulean reporter is suitable to report on ribosome expression.

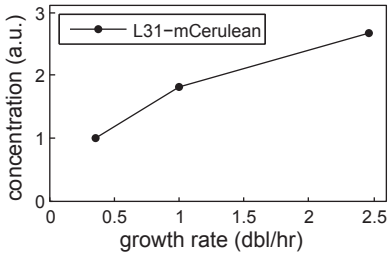


Figure 7.15: Estimate of r-protein-mCerulean concentration in dependence of growth rate. Concentrations were obtained as average values of growing microcolonies (different from Fig. 7.7). The medium for fastest growth is richer than in Fig. 7.7, the other media are identical. Experiments were performed on different days and with two different light bulbs. Concentration for intermediate growth was determined for both bulbs and thus used to normalize the concentrations of the other conditions. The data is useful for a qualitative check for growth-scaling but may have quantitative uncertainties. Concentration scales with growth rate, qualitatively consistent to Scott et al. [53] and in contrast to the previous label L31-mCherry.

7.4.2. CROSS-CORRELATIONS

We again measured cross-correlations between protein production and growth resp. protein concentration and growth for a wide span of growth rates (Fig. 7.16). It was immediately apparent that the cross-correlations differed strongly from previous correlations obtained with mCherry. For example, $R_{E\mu}(\tau)$ between concentration and growth rate was negative when growing at intermediate growth rates and expressing mCherry

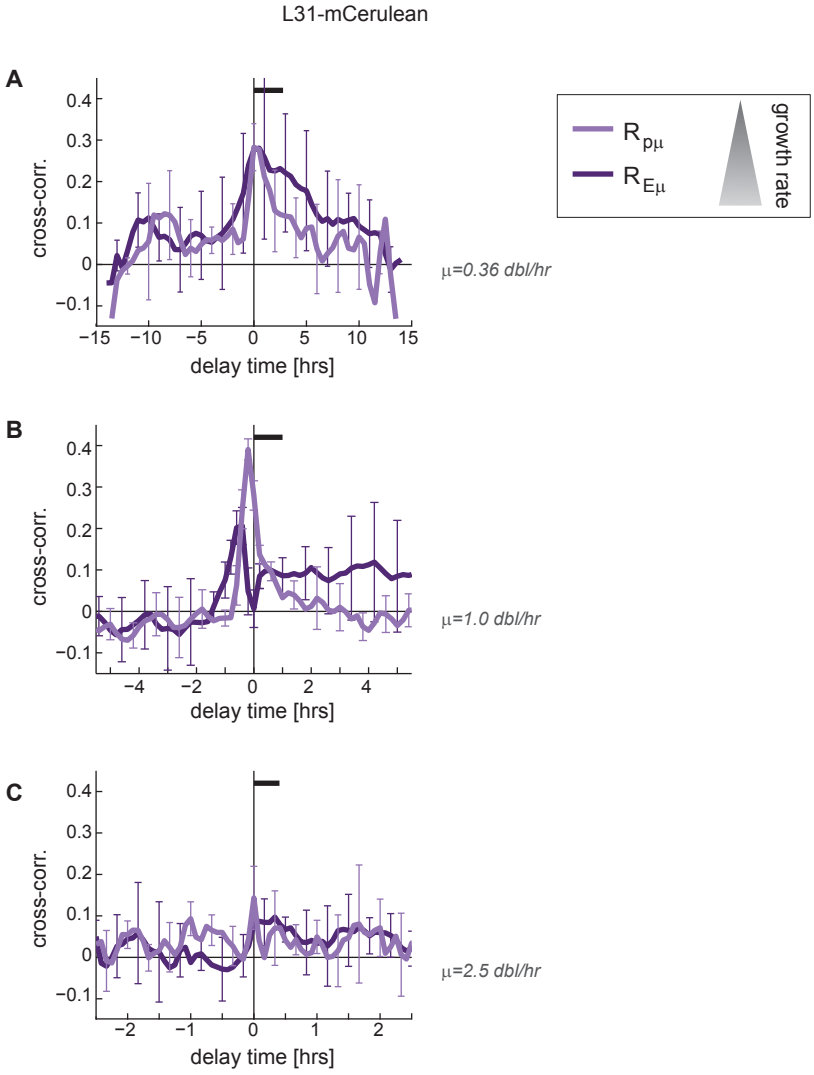


Figure 7.16: Cross-correlation of r-protein-mCerulean with growth rate. Cross-correlations $R_{p\mu}(\tau)$ of protein production rate and growth rate (bright purple) and $R_{E\mu}(\tau)$ of protein concentration and growth (dark purple) are displayed for the the L31-mCerulean fusion. Average growth rate increases from top to bottom and the thick black lines indicate mean interdivision time. Error bars are standard deviations and are indicated for a fraction of the data points. (**A**) Growth on M9 + 0.2mM uracil + 30mM acetate. Dataset is rather small: 140 cells as final colony size. (**B**) Growth on M9 + 0.2mM uracil + 0.1% lactose. (**C**) EZ defined rich + 0.2% glucose (different from above figures). Fluorescence concentration was determined as for Fig. 7.9. The cross-correlation shapes differ strongly from the theoretically equivalent label L31-mCherry (but without linker) of Fig. 7.9.

(Fig. 7.9B). Contrary to that, the same cross-correlation for L31-mCerulean was positive (Fig. 7.16B). Since the properties of the L31-mCerulean strain appeared in general better (growth rate scaling, colony morphology), this discrepancy points towards a potential issue with the cross-correlations obtained with L31-mCherry.

When investigating the cross-correlations of r-protein-mCerulean with growth rate for different conditions (μ=0.3-2.5dbl/hr), it was surprising that the functional shape was even qualitatively different in every condition. It also differed from results for P_{rrn}-GFP (Fig. 7.10). During slow growth, fluctuations in protein production rate, as well as in concentration, preceded growth fluctuations (Fig. 7.16A). This suggested that ribosomal r-proteins might be limiting, which was unexpected at such slow growth. A larger dataset would be needed to test this indication. At intermediate growth rates protein production and growth rate were strongly positively correlated and protein production was following behind with a small delay (ca 15min, Fig. 7.16B). This delay was confirmed in two independent experiments and did not occur for a constitutive promoter (Fig. 3.3). Therefore, it may be the result of active regulation and could be produced by slow translational feedback (Fig. 7.11). A transcriptional reporter could confirm this hypothesis. The concentration-growth correlation $R_{E\mu}(\tau)$ was positive and delayed to negative τ. Qualitatively, that shape resembled the cross-correlation that was observed for *lac* expression and which could be described with a model (chapter 6). Briefly, the model showed that $R_{E\mu}(\tau)$ would be positive if expression rate and growth fluctuations were coupled strong enough (consistent with the rather large correlation $R_{p\mu}(\tau)$ observed for ribosome production, Fig. 7.16B). However, quantitatively, $R_{E\mu}(\tau)$ of ribosomes was more complex: Before increasing at larger negative time delays, the cross-correlation reached a minimum at around $\tau = 0$. This decrease could be due to the time delay observed between protein production and growth rate, because it could render $R_{E\mu}(\tau)$ at $\tau \lesssim 0$ being dominated by dilution (see also model in Fig. 6.2). In addition, ribosomal proteins do not appear growth limiting at intermediate growth rate. During fast growth, cross-correlations were close to zero, similar to previous results (Fig. 7.16C). Whether the slightly positive $R_{E\mu}(\tau)$ points towards limitation would need to be investigated in independent experiments.

7.5. CONCLUSIONS

In the study presented in this chapter a lot of effort was spent on finding and characterizing good reporters for ribosome expression. We here present the conclusions of this search and preliminary conclusions on the biological question - whether ribosomes are limiting, optimal, and dynamically regulated.

7.5.1. CHOOSING THE BETTER REPORTER FOR RIBOSOMES

The r-protein-mCerulean fusion and P_{rrn}-GFP yielded different results, for example concerning cross-correlations. One thus has to choose which signal more likely reports on active ribosomes. Results of r-protein-mCherry were again different, but this fusion appeared unsuitable. Criteria for choosing are: The r-protein-mCerulean fusion did not affect growth rate, while an additional rRNA reporter caused a significant decrease (to be checked with a functional single-label strain). Both reporter concentrations, but rRNA

more accurately, scaled with growth rate. The fusion exactly reflects the r-protein copy number, while for rRNA concentration GFP is only a proxy. rRNA is, however, the more important component of a ribosome, regulation of rRNA expression is less complicated, and it is better studied in literature. Taken together, a strain containing both r-protein and rRNA reporters would be useful to shed light on the differences. When one reporter needs to be chosen, then an r-protein fusion seems better because of the absence of growth defects and the precise reporting on expression of the labeled ribosomal component.

7.5.2. RIBOSOMES ARE LIKELY NOT LIMITING, NOR FLUCTUATE AROUND A SHARP OPTIMUM

Cross-correlations showed that ribosomes are probably not dynamically limiting the growth rate because no synchronous up-fluctuations of ribosome concentration and growth rate were observed. That would suggest that ribosome content is not a limiting factor in gene expression, but further proof is needed. Showing non-limitation would be interesting because, contrary to our indications, it is commonly believed that fluctuations in the ribosome content directly cause noise in gene expression ("extrinsic noise") [5, 123].

The following experiments could corroborate our indications further (for more details, see the following Outlook section): To directly test whether a correlation between ribosome concentration and translation rate exists, expression of a constitutive gene can be measured together with ribosome concentration. Further, ribosomes could probably be made artificially limiting, which may result in different and contrasting cross-correlations.

Are our results contradicting results from literature? Likely not because all published results concern the relation between average ribosome concentration and population mean growth rate (for example [53]). This is conceptually very different from investigating fluctuations around an average. For example, we showed in chapter 6 that mean *lac* expression relates to mean growth rate very differently compared to fluctuations around that mean (Fig. 6.8 and 6.2). A more relevant argument is: To achieve a specific growth rate, cells need a corresponding ribosome concentration [53]. If they had less ribosomes they would grow more slowly (Fig. 7.3). In that sense ribosomes limit the growth rate. However, when cells express the required amount of ribosomes, an up-fluctuation may not increase growth because it can be at the cost of expression of other proteins. Thus, ribosomes would not appear limiting in a dynamic sense.

This argumentation suggests that ribosomes could fluctuate around an optimal level and both up- and down-fluctuations would yield lower growth rate. However, we did not find such a peaked optimum because the relation ribosomes-growth rate was flat or monotonous (Fig. 7.14). A possible reason is that global fluctuations ("common noise") may confound such a relation because faster growing cells always produce more proteins. Using a constitutive promoter to normalize for current cell state, that is, translation capacity, might disentangle these effects (see Outlook).

7.5.3. R-PROTEINS MIGHT LAG DUE TO TRANSLATIONAL FEEDBACK

Fluctuations in r-protein production lagged behind growth fluctuations in certain conditions (Fig.7.16B). Such a delay was not observed for a transcriptional r-protein promoter reporter supplied on a plasmid (data not shown). While data for rRNA is missing for that growth condition, the available data suggests the absence of delays for rRNA production. If this is the case, the delayed r-protein production might indicate that feedback on the translational level is active during dynamic fluctuations. This hypothesis can be tested by comparing results to a (chromosomally inserted) transcriptional r-protein reporter.

7.6. OUTLOOK

We conclude this chapter with suggestions on how the preliminary conclusions can be further tested and which future directions for research around the ribosome can be of interest. Our present results indicated that cells are quite robust in adjusting ribosome levels and balance is not easily disturbed. For example, cross-correlations did not change significantly after addition of antibiotics (Fig. 7.13) or deletion of RelA (data not shown). Therefore, if one wishes to search for clearly different phenotypes (e.g. "limiting ribosomes"), then more extreme perturbations such as nutrient up-/downshifts, addition of stress agents or manipulation of ribosome content might be needed. Next to that, the following list provides ideas for future projects.

7.6.1. INCREASE LIMITATION OF RIBOSOMES

An unexpected indication from our results was that during dynamic fluctuations ribosomes may be non-limiting for growth. To support this hypothesis, it would be helpful to construct a setting in which ribosomes are growth limiting and to contrast the results (for example cross-correlations) in these two scenarios.

A straightforward idea is to grow bacteria in the presence of low concentrations of translation inhibiting antibiotics because then efficiency of ribosomes decreases. However, cells adjust to this condition and increase the ribosome concentration, now growing at a lower rate (Fig. 7.3B). Indeed, our data indicates that in this new steady state ribosome concentration is again sufficient and not dynamically limiting (Fig. 7.13). We note that the effect of antibiotics was only tested in rich and not in minimal medium. To avoid adjusted cells, bacteria could be exposed to a small (sub-inhibitory) upshift of translation inhibiting antibiotics by using a microfluidic device. During the transient dynamics after the shift, cells will likely be unbalanced and ribosome concentration still needs to be adjusted. Thus, in this time window ribosomes may be growth limiting, and a randomly higher or lower ribosome concentration may result in faster or slower growing cells.

An alternative approach is to artificially increase the concentration of the alarmone ppGpp. ppGpp is a negative regulator of ribosome expression and therefore ribosome levels should decrease. Raising the ppGpp concentration can be achieved by overexpressing RelA from a plasmid or expressing a constitutively active mutant of RelA [233, 234]. Curiously, the absolute level of ppGpp seems to depend directly on the number of RelA proteins, without feedback from the ppGpp concentration [233]. Upon overexpression, activity of rRNA promoters and accumulation of stable rRNA indeed decreased, as

did the growth rate [233].

Expression from rRNA promoters can also be manipulated in a different way: The seven rRNA promoters are largely redundant and therefore up to six of these promoters can be deleted (usually denoted as $\Delta 6$) [187, 235, 236]. This reduces the concentration of rRNA in a cell and can thus make ribosomes growth limiting. Cells try to compensate the deletions with higher transcription from the remaining rRNA promoters. For up to roughly $\Delta 3$ no severe growth defect was observed, but for more deletions growth rate decreased significantly [187, 235, 236]. The severeness of the effect depended on the richness of the medium and thus the growth rate. However, it is also possible that the actual growth limitation is a result of the reduction of transfer RNA (tRNA) concentration instead of rRNA [235]. The tRNA and rRNA are jointly transcribed from the same promoters and tRNA expression is therefore also decreased upon P_{rrn} deletion. Providing tRNA on a plasmid indeed eliminated a large part of the growth defect for $\Delta 4$ and $\Delta 5$, thus suggesting that tRNA was growth limiting (Fig. 5 in [235]). However, for the severe $\Delta 6$ knockout a strong effect of rRNA limitation was observed [235].

Alternatively, ribosomes could be made growth-limiting by expressing large amounts of useless proteins (such as GFP) from plasmids. Then, growth rate is reduced [237], which is possibly caused by ribosomes being titrated away from the production of needed proteins. The precise expression burden may fluctuate over time due to variation in plasmid copy number, but cross-correlations between ribosome level and growth rate should represent a time-average of this additional expression load. According to a study of Shachrai et al. [238], the severeness of the growth burden seems, however, to depend on the growth state of the cell: During the first few generations after outgrowth of stationary phase or after a nutritional upshift, ribosomes were shown to be very limiting and the extra burden of gratuitous protein production is high. As consequence, ribosomes are produced at high rates. But after a couple of generations of growth, the burden significantly decreased. Probably ribosomes were not the main limiting factor any more but co-limiting growth with other factors. We note that the observed growth burden due to gratuitous protein expression differs in literature and was for example higher in work of Dong et al. [237], than in the study of Shachrai et al. [238]. Other studies observed variable growth defects upon overexpression, and the severeness of the decrease seemed to depend on the specific protein being overexpressed [112, 239].

Taken together, transient dynamics during an antibiotic upshift, or overexpressing the ppGpp synthase RelA or gratuitous proteins might be the most promising approaches to render ribosomes growth-limiting.

7.6.2. MEASUREMENT OF A CONSTITUTIVE PROMOTER

A complementary new possible experiment involves monitoring the expression from a constitutive promoter (e.g. a σ_{70} promoter) next to the ribosome reporter. This dual reporter construct would allow to investigate (at least) two open questions:

First, it could corroborate the indications that ribosome concentration is not limiting growth, or, more precisely, that it is not limiting the translation rate (i.e. rate of protein mass accumulation). To test this, fluctuations in ribosome concentration can be cross-correlated with fluctuations in production rate from the constitutive promoter. If ribosomes are non-limiting, this cross-correlation should not be positive at a delay of

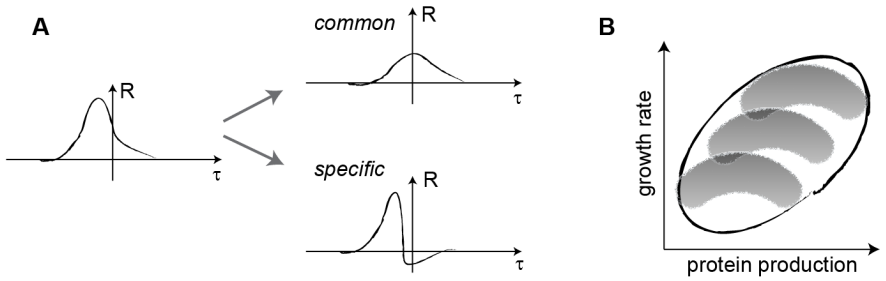


Figure 7.17: Potential benefits of a constitutive promoter. Monitoring expression from a constitutive promoter could allow disentangling of global and ribosome specific fluctuations by normalizing ribosome expression with the constitutive expression. (**A**) Hypothetical different contributions to the cross-correlation. (**B**) Even if the ribosome concentration was optimal for a fixed global state (three gray clouds), the observation may be confounded by averaging over all global states (black outline). Distinguishing for different global expression states with the constitutive promoter might provide a better test to investigate whether ribosomes fluctuate around an optimum.

zero.

Second, the constitutive promoter would allow to distinguish between general fluctuations in cellular state and ribosome specific effects. For example, production rate of many proteins is positively correlated with growth (for example chapter 3 and 6). Production of ribosomes is also positively correlated with growth rate, but appears to be related also by a regulatory interaction (Fig. 7.16B). To discriminate between common fluctuations and a regulatory level, ribosome expression can be normalized (divided) by expression from constitutive genes (Fig. 7.17A). Such an approach was already successfully used in bulk measurements to normalize for general expression activity [238]. It could also allow to readdress the question whether the ribosome level is optimal. This would manifest itself in a non-monotonous relation between ribosomes and growth rate (Fig. 7.17B). We note that if ribosomes were limiting and thus would affect expression of the constitutive gene, then this disentanglement might be more complicated.

7.6.3. MEASUREMENT AND MANIPULATION OF PPGPP

The alarmone ppGpp plays an important role in adjusting the ribosome synthesis to the demand (section 7.1.2). For example, the basal ppGpp level changes with mean growth rate, and upon stress a high level of ppGpp is transiently produced to decrease ribosome production [201]. Despite that importance and a variety of open questions, this study so far focused little on ppGpp. It is for example unknown whether and to what extent ppGpp signaling is also active during the more subtle fluctuations of ribosome content around its steady state. One question is whether cells transiently experience amino acid starvation and whether the ppGpp level fluctuates. Also, it was recently shown theoretically by Bosdriesz et al. [222] that ppGpp control is a robust and sufficient mechanism to maximize growth rate because it balances amino acid supply and demand. It would be interesting to test whether ppGpp is also involved in dynamically balancing amino acid

supply and demand and whether it is therefore involved in growth homeostasis.

To address these questions a sensor for ppGpp level is needed. Biochemical sensors suitable for time-lapse microscopy are to our knowledge not available. A good option is to use a fluorescent reporter controlled by a ppGpp-dependent promoter. Transcription of *rpoS*, the general stress response sigma factor, is strongly positively regulated by ppGpp level, and life-time of transcripts increases as well with ppGpp [10]. Therefore, a transcriptional or translational fluorescent reporter of *rpoS* expression could be used to monitor ppGpp concentration [10, 240].

Next to the questions mentioned above, such a reporter could also be used to test what type of stress (if any) cells experience during dynamic fluctuations: amino acid shortage triggers ppGpp synthesis by RelA, while many other stresses trigger synthesis by SpoT. Deleting either of the synthesis proteins could reveal which one was relevant during transient fluctuations (Δ*spoT* is however lethal in a wild type strain [241] and a knockout strain is only viable if *relA* is mutated as well). To artificially cause an amino acid starvation, the serine analog serine hydroxamate (SHX) can be added to the medium [202].

7.6.4. REGULATION OF R-PROTEIN SYNTHESIS AT THE TRANSCRIPTIONAL AND TRANSLATIONAL LEVEL

Potentially but not necessarily related to ppGpp regulation is the regulation of r-protein synthesis on transcriptional and translational level. It was suggested that one component may be more relevant during severe shifts while the other could do the fine-tuning during exponential growth [202]. This question could be addressed by measuring a transcriptional reporter and a translational fusion to r-proteins simultaneously. The measurement should be performed during balanced growth, as well as during severe up- and downshifts.

7.6.5. GROWTH MEDIUM DEPENDENCE - AMINO ACID VERSUS RIBOSOME LIMITATION

In recent research growth rate was often successfully used to parametrize the environment [53, 116], see for example scaling of ribosome concentration with the growth rate (Fig. 7.3A). In these works, details on the chemical composition of the medium were omitted and two media were basically equivalent if they yielded the same growth rate. However, conflicting work states that not only growth rate but the growth *medium* is important for control of physiological parameters of the cell [225]. In minimal medium, when typically a sugar is the only supplied carbon source, the flux of amino acids (and not ribosomes!) is suggested to be limiting the translational capacity and thus growth [225, 226]. A question posed by Ehrenberg et al. in [225] is: "Why are the concentrations of ribosomes (demand) and the activity of amino acid synthetic enzymes (supply) not balanced to each other during growth in poor media?" The situation is different in rich medium when amino acids are supplied: Then ribosomes themselves may be limiting growth. Also the ribosome concentration might depend on the environment in a more complex fashion than simple growth rate scaling: When a minimal medium and a rich medium (containing amino acids) support the same growth rate, then ribosome

concentration may be lower in the rich medium [225].

Therefore, investigating minimal and rich media yielding the same growth rate would be a very interesting (and very feasible) experiment to shed more light on the limiting factors of cell growth. A first step could be to investigate the shape of cross-correlations at the same mean growth rate, but with different or no precursors supplied. The study could be complemented with a reporter for amino acid synthesis. Medium shifts that add or remove amino acids (partly done in [225]) could for example reveal whether the ribosome pool in minimal medium has spare capacity to immediately process a higher influx of amino acids.

8

CONCLUSION AND OUTLOOK

In this thesis we studied stochasticity of gene expression and growth rate on the single-cell level. So, in addition to simple curiosity, why should one care about origins and consequences of stochasticity in cells? Or not simply accept the noise as given, because cells seem to be doing just fine [128]? A vivid motivation and comparison was given in Lestas et al. [128], who studied signal transmission fidelity, and which we here summarize: Consider gravity - it poses strong constraints on movement, but neither does it make flight impossible, nor do birds just happen to fly sufficiently well. Similarly, molecular stochasticity poses strong constraints on cellular control, but neither does it make any level of control impossible, nor do cells just happen to operate sufficiently deterministically. The task for scientists is to understand the trade-offs involved and to discern when noise is acceptable to cells and when not. Moreover, understanding properties of cellular noise is important for understanding robustness of cell proliferation [161], phenotypic heterogeneity (such as persistence to antibiotics) [9], cell fate decisions [12], and for designing robust synthetic gene circuits [242].

In this work, we found that the noise intensities of growth and expression fluctuations are connected by a remarkably simple linear relation (chapter 4). Thus, despite the inherent complexity of a living cell, rather simple quantitative relations can be found [53, 117, 121]. Such phenomenological relations can be very useful to quantify and predict behavior of complex biological systems before a full understanding of the processes on the molecular level is available [53]. With the help of a linear noise model we also showed that the intensities of different generic sources of cellular noise are connected by one global parameter. In future work, it would be interesting to explore how such a coupling could arise biologically, as well as which precise biological processes can be associated with the different model noise sources.

In the next chapter 5 we investigated the influence of the bacterial cell cycle on gene expression. We showed that about half of the noise measured in gene expression rate is caused by progression through the bacterial cell-cycle, or, more specifically, chromosome replication. Of that cell-cycle noise, more than half was attributed to the population averaged characteristics of cell cycle progression. Hence, large fluctuations in

cells can be caused by a deterministic process, but would appear random to an observer who is lacking the relevant information on the cells (here: their age). In biology, it is a long standing open question to what extent cellular variability is due to stochastic processes, and to what fraction it is caused by deterministic (but maybe unknown) factors [243, 244]. An interesting recent study by Balaban and coworkers [245], for example, showed that seemingly random variations in interdivision times of mammalian cells are actually rather deterministic and controlled by few parameters. In contrast to that, by applying the same algorithm [246] to *E. coli* expression data, they found that fluctuations in protein concentrations are reminiscent of random noise.

Our results showed that protein concentration, contrary to the production rate, was hardly affected by the cell cycle (consistent with the randomness detected in [245]; the focus of the analysis methods is, however, very different and results cannot be directly compared). This indicated a noise canceling mechanism: The twofold increase in production rate is automatically compensated for by a concurrent acceleration of volume increase. Thus, a passive homeostasis mechanism, for which active control is not required, balances protein concentration. If proteins were actively degraded, the noise canceling should be less efficient, which would be interesting to test experimentally. Next to that, as replication of the whole chromosome takes a significant amount of time [158], future work could test whether a regulatory interaction between two proteins is influenced by the relative distance of their genes on the chromosome.

In the next chapter 6 we showed that fluctuations in enzyme levels can propagate and cause growth fluctuations, and that fluctuations in growth rate can propagate back again to affect enzyme concentrations. This suggested that cellular metabolism is inherently stochastic. Because of the large number of reactions involved and the presence of stabilizing regulatory mechanisms [34], metabolism and growth were often considered to be constant [162, 163]. Our work suggests that, in future, stochasticity should be included in models, and be considered in experiments. It is still largely unknown and remains to be discovered what the consequences of fluctuations in metabolism are. Recent work in the group of Teusink [154] suggests that the effects can be large: They showed that heterogeneity purely on the metabolic level determined the cell fate of yeast cells upon exposure to a glucose upshift. Metabolism in a, genetically indistinguishable, subpopulation was imbalanced, which resulted in growth arrest of these cells.

A second interesting observation in our study was the existence of a noise-canceling mechanism that stabilized protein concentrations. Up-fluctuations in the global state of a cell (metabolic activity) increased protein production rate as well as growth rate. These two parameters act antagonistically on protein concentration and thus partly cancel each other, resulting in only small fluctuations in the enzyme level. Hence, a second balancing mechanism, next to the cell-cycle buffering discussed above, acts on protein concentration. The existence of these passive (automatic) buffering mechanisms may promote growth homeostasis. They could be a first explanation to why cells can grow so robustly, despite the stochasticity on the molecular level.

Still, cellular growth rate is not completely steady but fluctuates, as was shown by us and [31, 32]. The properties of growth fluctuations are poorly understood and could be addressed in future projects. Open questions are for example: Why does the growth noise intensity depend on the environment? Why is the dominant time scale of growth

fluctuations faster than the time scale of fluctuations in protein concentration (note that active transmission from enzymes as studied in chapter 6 caused additional fluctuations at slower time scales)? Is growth noise an evolvable property [247] and does lower noise convey a fitness advantage? To approach the first of these questions, one could try to disentangle possible influences such as average growth rate, richness of the growth medium (e.g. supplied amino acids) and cell size. The second topic might be addressed with modeling. Specifically, one could test whether the combined action (additive or multiplicative) of multiple slow noise sources can result in fast fluctuations as output.

In the last study presented, we investigated the influence of dynamic fluctuations in the concentration of a central molecular machine, the ribosome, on the growth rate (chapter 7). The results of this study are preliminary but indicate that ribosomes are neither dynamically limiting growth rate, nor fluctuating around a sharply peaked optimum concentration. If this result can be further supported, it naturally opens up the question which other components are the main factors that limit growth, or cause fluctuations in growth rate and gene expression. In chapter 6 we already found that some central metabolism proteins and lowly expressed *lac* proteins can transmit fluctuations. But the influence of many other factors such as RNA polymerase concentration [120], metabolite or ATP concentrations, or activity-regulated proteins is largely unknown. Potential approaches could include the use of spinach aptamers [248] for visualizing small molecules and FRET sensors or mutants for detecting, respectively manipulating, protein activity. More generally, also limitations occurring during fluctuating environments (e.g. during nutrient shifts), next to the here studied steady-state fluctuations, would be interesting to investigate.

As last point, the studies presented in chapter 6 and 7 also showed the fundamental difference between measuring population averages in different steady states, or single-cell fluctuations within one population in one environment (see also [32]). Relations obtained from bulk experiments may be quite different when measured on the single-cell level, hence intuition may be misleading and one should be open for surprises.

BIBLIOGRAPHY

[1] H. Bremer and P. P. Dennis, *Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate,* in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, edited by F. C. Neidhardt (ASM Press, Washington, DC, 1996) 2nd ed., pp. 1553–1569.

[2] H. C. Berg, *Bacterial behaviour,* Nature **254**, 389 (1975).

[3] M. B. Elowitz, M. G. Surette, P. E. Wolf, J. B. Stock, and S. Leibler, *Protein mobility in the cytoplasm of Escherichia coli,* J. Bacteriol. **181**, 197 (1999).

[4] T. Mignot and J. W. Shaevitz, *Active and passive mechanisms of intracellular transport and localization in bacteria,* Curr. Opin. Microbiol. **11**, 580 (2008).

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

[6] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, *Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells.* Science **329**, 533 (2010).

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

[8] P. Choi, L. Cai, K. Frieda, and X. Xie, *A stochastic single-molecule event triggers phenotype switching of a bacterial cell,* Science **322**, 442 (2008).

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

[10] E. Maisonneuve, M. Castro-Camargo, and K. Gerdes, *(p)ppGpp controls bacterial persistence by stochastic induction of toxin-antitoxin activity,* Cell **154**, 1140 (2013).

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

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

[13] S. Artavanis-Tsakonas, M. D. Rand, and R. J. Lake, *Notch Signaling : Cell Fate Control and Signal Integration in Development,* Signal Transduct. **284**, 770 (1999).

[14] A. Colman-Lerner, A. Gordon, E. Serra, T. Chin, O. Resnekov, D. Endy, C. G. Pesce, and R. Brent, *Regulated cell-to-cell variation in a cell-fate decision system.* Nature **437**, 699 (2005).

[15] G. Balázsi, A. van Oudenaarden, and J. J. Collins, *Cellular decision making and biological noise: from microbes to mammals.* Cell **144**, 910 (2011).

[16] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, *Stochasticity in gene expression: from theories to phenotypes.* Nat. Rev. Genet. **6**, 451 (2005).

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

[18] M. Thattai and A. van Oudenaarden, *Intrinsic noise in gene regulatory networks.* Proc. Natl. Acad. Sci. U.S.A. **98**, 8614 (2001).

[19] P. S. Swain, M. B. Elowitz, and E. D. Siggia, *Intrinsic and extrinsic contributions to stochasticity in gene expression.* Proc. Natl. Acad. Sci. U.S.A. **99**, 12795 (2002).

[20] J. T. Mettetal, D. Muzzey, J. M. Pedraza, E. M. Ozbudak, and A. van Oudenaarden, *Predicting stochastic gene expression dynamics in single cells.* Proc. Natl. Acad. Sci. U.S.A. **103**, 7304 (2006).

[21] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, and N. Barkai, *Noise in protein expression scales with natural protein abundance.* Nat. Genet. **38**, 636 (2006).

[22] M. L. Simpson, C. D. Cox, and G. S. Sayler, *Frequency domain analysis of noise in autoregulated gene circuits.* Proc. Natl. Acad. Sci. U.S.A. **100**, 4551 (2003).

[23] D. W. Austin, M. S. Allen, J. M. McCollum, R. D. Dar, J. R. Wilgus, G. S. Sayler, N. F. Samatova, C. D. Cox, and M. L. Simpson, *Gene network shaping of inherent noise spectra.* Nature **439**, 608 (2006).

[24] M. Scott, T. Hwa, and B. Ingalls, *Deterministic characterization of stochastic genetic circuits.* Proc. Natl. Acad. Sci. U.S.A. **104**, 7402 (2007).

[25] W. K. Smits, O. P. Kuipers, and J.-W. Veening, *Phenotypic variation in bacteria: the role of feedback regulation,* Nat. Rev. Microbiol. **4**, 259 (2006).

[26] J. M. Pedraza, *Noise Propagation in Gene Networks,* Science **1965**, 1965 (2005).

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

[28] B. Munsky, G. Neuert, and A. van Oudenaarden, *Using gene expression noise to understand gene regulation.* Science **336**, 183 (2012).

[29] H. J. E. Beaumont, J. Gallie, C. Kost, G. C. Ferguson, and P. B. Rainey, *Experimental evolution of bet hedging,* Nature **462**, 90 (2009).

[30] E. Kussell and S. Leibler, *Phenotypic diversity, population growth, and information in fluctuating environments.* Science **309**, 2075 (2005).

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

[32] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun, *Cell-Size Control and Homeostasis in Bacteria,* Curr. Biol. **25**, 1 (2014).

[33] S. Iyer-Biswas, C. S. Wright, J. T. Henry, K. Lo, S. Burov, Y. Lin, G. E. Crooks, S. Crosson, A. R. Dinner, and N. F. Scherer, *Scaling laws governing stochastic growth and division of single bacterial cells,* Proc. Natl. Acad. Sci. U.S.A. **111**, 15912 (2014).

[34] F. C. Neidhardt, J. L. Ingraham, and M. Schaechter, *Physiology of the bacterial cell: a molecular approach* (Sinauer Associates, Sunderland, MA, 1990).

[35] B. Volkmer and M. Heinemann, *Condition-Dependent Cell Volume and Concentration of Escherichia coli to Facilitate Data Conversion for Systems Biology Modeling,* PLoS One **6**, e23126 (2011).

[36] F. Blattner, I. Plunkett G, C. Bloch, N. Perna, V. Burland, M. Riley, J. Collado-Vides, J. Glasner, C. Rode, G. Mayhew, J. Gregor, N. Davis, H. Kirkpatrick, M. Goeden, D. Rose, B. Mau, and Y. Shao, *The Complete Genome Sequence of Escherichia coli K-12,* Science **277**, 1453 (1997).

[37] G. Banfalvi, *Structural Organization of Dna B-Dna,* Biochem. Educ. **14**, 50 (1986).

[38] R. J. Ellis, *Macromolecular crowding: an important but neglected aspect of the intracellular environment,* Curr. Opin. Struct. Biol. **11**, 114 (2001).

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

[40] S. Itzkovitz and A. van Oudenaarden, *Validating transcripts with probes and imaging technology,* Nat. Methods **8**, S12 (2011).

[41] F. Tang, C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B. B. Tuch, A. Siddiqui, K. Lao, and M. A. Surani, *mRNA-Seq whole-transcriptome analysis of a single cell,* Nat. Methods **6**, 377 (2009).

[42] R. Zenobi, *Single-cell metabolomics: analytical and biological perspectives.* Science **342**, 1243259 (2013).

[43] M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher, *Green fluorescent protein as a marker for gene expression,* Science **263**, 802 (1994).

[44] P. Nghe, S. Boulineau, S. Gude, P. Recouvreux, J. S. van Zon, and S. J. Tans, *Microfabricated Polyacrylamide Devices for the Controlled Culture of Growing Cells and Developing Organisms,* PLoS One **8**, 1 (2013).

[45] W. A. Bonner, H. R. Hulett, R. G. Sweet, and L. A. Herzenberg, *Fluorescence activated cell sorting,* Rev. Sci. Instrum. **43**, 404 (1972).

[46] L. A. Herzenberg, J. Tung, W. A. Moore, L. A. Herzenberg, and D. R. Parks, *Interpreting flow cytometry data: a guide for the perplexed.* Nat. Immunol. **7**, 681 (2006).

[47] A. K. White, M. VanInsberghe, O. I. Petriv, M. Hamidi, D. Sikorski, M. A. Marra, J. Piret, S. Aparicio, and C. L. Hansen, *High-throughput microfluidic single-cell RT-qPCR,* Proc. Natl. Acad. Sci. U.S.A. **108**, 13999 (2011).

[48] A. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. Weitz, and M. Kirschner, *Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells,* Cell **161**, 1187 (2015).

[49] T. Nawy, *Single-cell sequencing,* Nat. Methods **11**, 18 (2014).

[50] D. B. Kell and J. D. Knowles, *The Role of Modeling in Systems Biology,* in *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts*, edited by Z. Szallasi, J. Stelling, and V. Periwal (MIT Press, Cambridge, Massachusetts, 2006) 1st ed., Chap. 1, pp. 3–18.

[51] M. Kollmann, L. Løvdok, K. Bartholomé, J. Timmer, and V. Sourjik, *Design principles of a bacterial signalling network.* Nature **438**, 504 (2005).

[52] A. van Ooyen, *Using theoretical models to analyse neural development,* Nat. Rev. Neurosci. **12**, 311 (2011).

[53] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa, *Interdependence of Cell Growth and Gene Expression: Origins and Consequences,* Science **330**, 1099 (2010).

[54] M. Osella, E. Nugent, and M. Cosentino Lagomarsino, *Concerted control of Escherichia coli cell division.* Proc. Natl. Acad. Sci. U.S.A. **111**, 3431 (2014).

[55] M. Schaechter, O. Maaloe, and N. O. Kjeldgaard, *Dependency on medium and temperature of cell size and chemical composition during balanced grown of Salmonella typhimurium.* J. Gen. Microbiol. **19**, 592 (1958).

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

[57] K. F. Jensen, *The Escherichia coli K-12 'wild types' W3110 and MG1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyrE expression levels,* J. Bacteriol. **175**, 3401 (1993).

[58] K. A. Datsenko and B. L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proc. Natl. Acad. Sci. U.S.A. **97**, 6640 (2000).

[59] K. Shigekawa and W. J. Dower, *Electroporation of eukaryotes and prokaryotes: a general approach to the introduction of macromolecules into cells,* Biotechniques **6**, 742 (1988).

[60] J. Yuan and H. C. Berg, *Resurrection of the flagellar rotary motor near zero load.* Proc. Natl. Acad. Sci. U.S.A. **105**, 1182 (2008).

[61] D. Kiviet, *The lac Operon: Fluctuations, Growth and Evolution*, Phd thesis, Technical University Delft (2010).

[62] W. J. Chi, Y. K. Chang, and S. K. Hong, *Agar degradation by microorganisms and agar-degrading enzymes,* Appl. Microbiol. Biotechnol. **94**, 917 (2012).

[63] A. Edelstein, N. Amodaj, K. Hoover, R. Vale, and N. Stuurman, *Computer control of microscopes using μ manager,* Curr. Protoc. Mol. Biol. , 14 (2010).

[64] E. Itan, G. Carmon, A. Rabinovitch, I. Fishov, and M. Feingold, *Shape of nonseptated Escherichia coli is asymmetric,* Phys. Rev. E **77**, 1 (2008).

[65] J. S. Bendat and A. G. Piersol, *Random data: analysis and measurement procedures*, 4th ed. (John Wiley & Sons, New York, 2011).

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

[67] V. Shahrezaei, J. F. Ollivier, and P. S. Swain, *Colored extrinsic fluctuations and stochastic gene expression.* Mol. Syst. Biol. **4**, 196 (2008).

[68] A. Sigal, R. Milo, A. Cohen, N. Geva-Zatorsky, Y. Klein, Y. Liron, N. Rosenfeld, T. Danon, N. Perzov, and U. Alon, *Variability and memory of protein levels in human cells.* Nature **444**, 643 (2006).

[69] C. W. J. Granger, *Investigating Causal Relations by Econometric Models and Cross-spectral Methods,* Econometrica **37**, 424 (1969).

[70] S. Guo, A. K. Seth, K. M. Kendrick, C. Zhou, and J. Feng, *Partial Granger causality-Eliminating exogenous inputs and latent variables,* J. Neurosci. Methods **172**, 79 (2008).

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

[72] N. C. Shaner, P. A. Steinbach, and R. Y. Tsien, *A guide to choosing fluorescent proteins.* Nat. Methods **2**, 905 (2005).

[73] M. Ehrenberg, *The green fluorescent protein: discovery, expression and development.* Scientific Background on the Nobel Prize in Chemistry. Retrieved from https://www.nobelprize.org/nobel_prizes/chemistry/laureates/2008/advanced-chemistryprize2008.pdf (2008).

[74] Nobel Media AB 2014, *The Nobel Prize in Chemistry 2008,* Press Release. Retrieved from http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2008/press.html (2008).

[75] O. Shimomura, F. H. Johnson, and Y. Saiga, *Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea,* J. Cell. Comp. Physiol. **59**, 223 (1962).

[76] R. Heim, A. B. Cubitt, and R. Y. Tsien, *Improved green fluorescence,* Nature **373**, 663 (1995).

[77] A. Crameri, E. A. Whitehorn, E. Tate, and W. P. Stemmer, *Improved green fluorescent protein by molecular evolution using DNA shuffling.* Nat. Biotechnol **14**, 315 (1996).

[78] T. T. Yang, L. Cheng, and S. R. Kain, *Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein,* Nucleic Acids Res. **24**, 4592 (1996).

[79] R. Heim, D. C. Prasher, and R. Y. Tsien, *Wavelength mutations and posttranslational autoxidation of green fluorescent protein.* Proc. Natl. Acad. Sci. U.S.A. **91**, 12501 (1994).

[80] R. Heim and R. Y. Tsien, *Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer.* Curr. Biol. **6**, 178 (1996).

[81] M. A. Rizzo, G. H. Springer, B. Granada, and D. W. Piston, *An improved cyan fluorescent protein variant useful for FRET.* Nat. Biotechnol. **22**, 445 (2004).

[82] M. Ormö, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien, and S. J. Remington, *Crystal structure of the Aequorea victoria green fluorescent protein,* Science **273**, 1392 (1996).

[83] T. Nagai, K. Ibata, E. S. Park, M. Kubota, K. Mikoshiba, and A. Miyawaki, *A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications.* Nat. Biotechnol. **20**, 87 (2002).

[84] B. P. Cormack, R. H. Valdivia, and S. Falkow, *FACS-optimized mutants of the green fluorescent protein (GFP).* Gene **173**, 33 (1996).

[85] J. Lippincott-Schwartz and G. H. Patterson, *Development and use of fluorescent protein markers in living cells.* Science **300**, 87 (2003).

[86] A. Miyawaki, T. Nagai, and H. Mizuno, *Mechanisms of protein fluorophore formation and engineering,* Curr. Opin. Chem. Biol. **7**, 557 (2003).

[87] R. Iizuka, M. Yamagishi-Shirasaki, and T. Funatsu, *Kinetic study of de novo chromophore maturation of fluorescent proteins.* Anal. Biochem. **414**, 173 (2011).

[88] J. Megerle, G. Fritz, U. Gerland, K. Jung, and J. O. Rädler, *Timing and dynamics of single cell gene expression in the arabinose utilization system.* Biophys. J. **95**, 2103 (2008).

[89] R. Rizzuto, M. Brini, F. De Giorgi, R. Rossi, R. Heim, R. Y. Tsien, and T. Pozzan, *Double labelling of subcellular structures with organelle-targeted GFP mutants in vivo.* Curr. Biol. **6**, 183 (1996).

[90] R. Y. Tsien, *The Green Fluorescent Protein,* Annu. Rev. Biochem. **67**, 509 (1998).

[91] O. Griesbeck, G. S. Baird, R. E. Campbell, D. A. Zacharias, and R. Y. Tsien, *Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications,* J. Biol. Chem. **276**, 29188 (2001).

[92] M. V. Matz, A. F. Fradkov, Y. A. Labas, A. P. Savitsky, A. G. Zaraisky, M. L. Markelov, and S. A. Lukyanov, *Fluorescent proteins from nonbioluminescent Anthozoa species.* Nat. Biotechnol. **17**, 969 (1999).

[93] J. Wiedenmann, A. Schenk, C. Röcker, A. Girod, K.-D. Spindler, and G. U. Nienhaus, *A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from Entacmaea quadricolor (Anthozoa, Actinaria).* Proc. Natl. Acad. Sci. U.S.A. **99**, 11646 (2002).

[94] L. A. Gross, G. S. Baird, R. C. Hoffman, K. K. Baldridge, and R. Y. Tsien, *The structure of the chromophore within DsRed, a red fluorescent protein from coral.* Proc. Natl. Acad. Sci. U.S.A. **97**, 11990 (2000).

[95] E. M. Merzlyak, J. Goedhart, D. Shcherbo, M. E. Bulina, A. S. Shcheglov, A. F. Fradkov, A. Gaintzeva, K. A. Lukyanov, S. Lukyanov, T. W. J. Gadella, and D. M. Chudakov, *Bright monomeric red fluorescent protein with an extended fluorescence lifetime.* Nat. Methods **4**, 555 (2007).

[96] G. S. Baird, D. A. Zacharias, and R. Y. Tsien, *Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral.* Proc. Natl. Acad. Sci. U.S.A. **97**, 11984 (2000).

[97] N. C. Shaner, R. E. Campbell, P. A. Steinbach, B. N. G. Giepmans, A. E. Palmer, and R. Y. Tsien, *Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein.* Nat. Biotechnol. **22**, 1567 (2004).

[98] D. Shcherbo, E. M. Merzlyak, T. V. Chepurnykh, A. F. Fradkov, G. V. Ermakova, E. A. Solovieva, K. A. Lukyanov, E. A. Bogdanova, A. G. Zaraisky, S. Lukyanov, and D. M. Chudakov, *Bright far-red fluorescent protein for whole-body imaging.* Nat. Methods **4**, 741 (2007).

[99] D. Shcherbo, C. S. Murphy, G. V. Ermakova, E. A. Solovieva, T. V. Chepurnykh, A. S. Shcheglov, V. V. Verkhusha, V. Z. Pletnev, K. L. Hazelwood, P. M. Roche, S. Lukyanov, A. G. Zaraisky, M. W. Davidson, and D. M. Chudakov, *Far-red fluorescent tags for protein imaging in living tissues.* Biochem. J. **418**, 567 (2009).

[100] D. M. Chudakov, M. V. Matz, S. Lukyanov, and K. A. Lukyanov, *Fluorescent proteins and their applications in imaging living cells and tissues.* Physiol. Rev. **90**, 1103 (2010).

[101] Y. Hiraoka, T. Shimi, and T. Haraguchi, *Multispectral imaging fluorescence microscopy for living cells.* Cell Struct. Funct. **27**, 367 (2002).

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

[103] O. K. Silander, N. Nikolic, A. Zaslaver, A. Bren, I. Kikoin, U. Alon, and M. Ackermann, *A genome-wide analysis of promoter-mediated phenotypic noise in Escherichia coli,* PLoS Genet. **8** (2012).

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

[105] C. D. Webb, A. Decatur, A. Teleman, and R. Losick, *Use of Green Fluorescent Protein for Visualization of Cell-Specific Gene Expression and Subcellular Protein Localization during Sporulation in Bacillus subtilis,* J. Bacteriol. **177**, 5906 (1995).

[106] X. Ma, D. W. Ehrhardt, and W. Margolin, *Colocalization of cell division proteins FtsZ and FtsA to cytoskeletal structures in living Escherichia coli cells by using green fluorescent protein.* Proc. Natl. Acad. Sci. U.S.A. **93**, 12998 (1996).

[107] R. S. Cox, M. J. Dunlop, and M. B. Elowitz, *A synthetic three-color scaffold for monitoring genetic regulation and noise.* J. Biol. Eng. **4**, 10 (2010).

[108] D. A. Zacharias, J. D. Violin, A. C. Newton, and R. Y. Tsien, *Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells.* Science **296**, 913 (2002).

[109] A. Gordon, A. Colman-Lerner, T. E. Chin, K. R. Benjamin, R. C. Yu, and R. Brent, *Single-cell quantification of molecules and rates using open-source microscope-based cytometry,* Nat. Methods **4**, 175 (2007).

[110] A. M. Adiciptaningrum, *Phase Variation of Type 1 Fimbriae : a Single Cell Investigation*, Phd thesis, University of Amsterdam (2009).

[111] G. Charvin, F. R. Cross, and E. D. Siggia, *A microfluidic device for temporally controlled gene expression and long-term fluorescent imaging in unperturbed dividing yeast cells.* PLoS One **3**, e1468 (2008).

[112] E. Hebisch, J. Knebel, J. Landsberg, E. Frey, and M. Leisner, *High Variation of Fluorescence Protein Maturation Times in Closely Related Escherichia coli Strains.* PLoS One **8**, e75991 (2013).

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

[114] S. Tănase-Nicola and P. R. ten Wolde, *Regulatory control and the costs and benefits of biochemical noise.* PLoS Comput. Biol. **4**, e1000125 (2008).

[115] A. S. Kennard, M. Osella, A. Javer, J. Grilli, P. Nghe, S. J. Tans, P. Cicuta, and M. C. Lagomarsino, *Individuality and universality in the growth-division laws of single e. coli cells,* Phys. Rev. E **93**, 012408 (2016).

[116] S. Klumpp, Z. Zhang, and T. Hwa, *Growth rate-dependent global effects on gene expression in bacteria.* Cell **139**, 1366 (2009).

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

[118] V. Shahrezaei and S. Marguerat, *Connecting growth with gene expression: of noise and numbers.* Curr. Opin. Microbiol. **25**, 127 (2015).

[119] C. J. Zopf, K. Quinn, J. Zeidman, and N. Maheshri, *Cell-cycle dependence of transcription dominates noise in gene expression.* PLoS Comput. Biol. **9**, e1003161 (2013).

[120] S. Yang, S. Kim, Y. Rim Lim, C. Kim, H. J. An, J.-H. Kim, J. Sung, and N. K. Lee, *Contribution of RNA polymerase concentration variation to protein expression noise.* Nat. Commun. **5**, 4761 (2014).

[121] L. Keren, D. Van-Dijk, S. Weingarten-Gabbay, D. Davidi, G. Jona, A. Weinberger, R. Milo, and E. Segal, *Noise in gene expression is coupled to growth rate,* Genome Res. **25**, 1893 (2015).

[122] C. S. Foote, *Mechanisms of Photosensitized Oxidation,* Science **162**, 963 (1968).

[123] C. G. Bowsher and P. S. Swain, *Identifying sources of variation and the flow of information in biochemical networks.* Proc. Natl. Acad. Sci. U.S.A. **109**, E1320 (2012).

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

[125] A. Schwabe and F. J. Bruggeman, *Contributions of cell growth and biochemical reactions to nongenetic variability of cells.* Biophys. J. **107**, 301 (2014).

[126] D. C. Laporte, K. Walsh, and D. E. Koshland, *The branch point effect. ultrasensitivity and subsensitivity to metabolic control,* J. Biol. Chem. **259**, 14068 (1984).

[127] L. Gerosa and U. Sauer, *Regulation and control of metabolic fluxes in microbes,* Curr. Opin. Biotechnol. **22**, 566 (2011).

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

[129] C. C. Govern and P. R. ten Wolde, *Optimal resource allocation in cellular sensing systems,* Proc. Natl. Acad. Sci. U.S.A. **111**, 17486 (2014).

[130] G. M. Süel, R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, and M. B. Elowitz, *Tunability and noise dependence in differentiation dynamics.* Science **315**, 1716 (2007).

[131] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, *Kernel density estimation via diffusion,* Ann. Stat. **38**, 2916 (2010).

[132] N. Walker, P. Nghe, and S. J. Tans, *Generation and filtering of gene expression noise by the bacterial cell cycle,* BMC Biol. **14**, 1 (2016).

[133] A. Raj and A. van Oudenaarden, *Single-molecule approaches to stochastic gene expression.* Annu. Rev. Biophys. **38**, 255 (2009).

[134] T. Kalisky, E. Dekel, and U. Alon, *Cost-benefit theory and optimal design of gene regulation functions.* Phys. Biol. **4**, 229 (2007).

[135] T. Mora and A. M. Walczak, *Effect of Phenotypic Selection on Stochastic Gene Expression,* J. Phys. Chem. B **117**, 13194 (2013).

[136] N. A. Cookson, S. W. Cookson, L. S. Tsimring, and J. Hasty, *Cell cycle-dependent variations in protein concentration.* Nucleic Acids Res. **38**, 2676 (2010).

[137] J. D. Wang and P. A. Levin, *Metabolism, cell growth and the bacterial cell cycle.* Nat. Rev. Microbiol. **7**, 822 (2009).

[138] J. Yu, J. Xiao, X. Ren, K. Lao, and X. S. Xie, *Probing Gene Expression in Live Cells, One Protein Molecule at a Time,* Science **311**, 1600 (2006).

[139] R. Marathe, V. Bierbaum, D. Gomez, and S. Klumpp, *Deterministic and Stochastic Descriptions of Gene Expression Dynamics,* J. Stat. Phys. **148**, 608 (2012).

[140] M. Acar, B. F. Pando, F. H. Arnold, M. B. Elowitz, and A. van Oudenaarden, *A general mechanism for network-dosage compensation in gene circuits.* Science **329**, 1656 (2010).

[141] A. Becskei and L. Serrano, *Engineering stability in gene networks by autoregulation.* Nature **405**, 590 (2000).

[142] L. H. Koppes and C. L. Woldringh, *Size Variations and Correlation of Different Cell Cycle Events in Slow-Growing Escherichia coli,* J. Bacteriol. **134**, 423 (1978).

[143] S. Cooper and C. E. Helmstetter, *Chromosome replication and the division cycle of Escherichia coli B/r.* J. Mol. Biol. **31**, 519 (1968).

[144] S. Cooper, *What Is the Bacterial Growth Law during the Division Cycle?* J. Bacteriol. **170**, 5001 (1988).

[145] S. Cooper, *Distinguishing between linear and exponential cell growth during the division cycle: single-cell studies, cell-culture studies, and the object of cell-cycle research*. Theor. Biol. Med. Model. **3**, 10 (2006).

[146] M. Mir, Z. Wang, Z. Shen, M. Bednarz, R. Bashir, I. Golding, S. G. Prasanth, and G. Popescu, *Optical measurement of cycle-dependent cell growth*. Proc. Natl. Acad. Sci. U.S.A. **108**, 13124 (2011).

[147] G. Reshes, S. Vanounou, I. Fishov, and M. Feingold, *Cell shape dynamics in Escherichia coli*. Biophys. J. **94**, 251 (2008).

[148] H. E. Kubitschek, *Bilinear cell growth of Escherichia coli*. J. Bacteriol. **148**, 730 (1981).

[149] H. E. Kubitschek, *Increase in cell mass during the division cycle of Escherichia coli B/rA*. J. Bacteriol. **168**, 613 (1986).

[150] H. E. Kubitschek and S. R. Pai, *Variation in Precursor Pool Size during the Division Cycle of Escherichia coli : Further Evidence for Linear Cell Growth,* J. Bacteriol. **170**, 431 (1988).

[151] K. Skarstad, E. Boye, and H. B. Steen, *Timing of initiation of chromosome replication in individual Escherichia coli cells,* EMBO J. **5**, 1711 (1986).

[152] D. White, J. Drummond, and C. Fuqua, *The Physiology and Biochemistry of Prokaryotes*, 4th ed. (Oxford University Press, New York, 2012).

[153] H. P. Lu, L. Xun, and X. S. Xie, *Single-molecule enzymatic dynamics*. Science **282**, 1877 (1998).

[154] J. H. van Heerden, M. T. Wortel, F. J. Bruggeman, J. J. Heijnen, Y. J. M. Bollen, R. Planqué, J. Hulshof, T. G. O'Toole, S. A. Wahl, and B. Teusink, *Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells*. Science **343**, 1245114 (2014).

[155] H. Kempe, A. Schwabe, F. Cremazy, P. J. Verschure, and F. J. Bruggeman, *The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise,* Mol. Biol. Cell **26**, 797 (2014).

[156] O. Padovan-Merhar, G. P. Nair, A. G. Biaesch, A. Mayer, S. Scarfone, S. W. Foley, A. R. Wu, L. S. Churchman, A. Singh, and A. Raj, *Single Mammalian Cells Compensate for Differences in Cellular Volume and DNA Copy Number through Independent Global Transcriptional Mechanisms,* Mol. Cell **58**, 1 (2015).

[157] F. J. Trueba and C. L. Woldringh, *Changes in cell diameter during the division cycle of Escherichia coli.* J. Bacteriol. **142**, 869 (1980).

[158] M. Bipatnath, P. P. Dennis, and H. Bremer, *Initiation and Velocity of Chromosome Replication in Escherichia coli B / r and K-12,* J. Bacteriol. **180**, 265 (1998).

[159] D. J. Wilkinson, *Stochastic modelling for quantitative description of heterogeneous biological systems,* Nat. Rev. Genet. **10**, 122 (2009).

[160] A. Eldar and M. B. Elowitz, *Functional roles for noise in genetic circuits.* Nature **467**, 167 (2010).

[161] M. G. Vander Heiden, L. Cantley, and C. Thompson, *Understanding the Warburg effect: the metabolic requirements of cell proliferation.* Science **324**, 1029 (2009).

[162] D. Fell and A. Cornish-Bowden, *Understanding the control of metabolism*, Vol. 2 (Portland Press, London, 1997).

[163] M. J. Herrgård, M. W. Covert, and B. Ø. Palsson, *Reconstruction of microbial transcriptional regulatory networks,* Curr. Opin. Biotechnol. **15**, 70 (2004).

[164] M. L. Ferguson, D. Le Coq, M. Jules, S. Aymerich, O. Radulescu, N. Declerck, and C. A. Royer, *Reconciling molecular regulatory mechanisms with noise patterns of bacterial metabolic promoters in induced and repressed states.* Proc. Natl. Acad. Sci. U.S.A. **109**, 155 (2012).

[165] M. Rodríguez, T. A. Good, M. E. Wales, J. P. Hua, and J. R. Wild, *Modeling allosteric regulation of de novo pyrimidine biosynthesis in Escherichia coli,* J. Theor. Biol. **234**, 299 (2005).

[166] Y. Hart, D. Madar, J. Yuan, A. Bren, A. E. Mayo, J. D. Rabinowitz, and U. Alon, *Robust control of nitrogen assimilation by a bifunctional enzyme in E. coli,* Mol. Cell **41**, 117 (2011).

[167] H. S. Yun, J. Hong, and H. C. Lim, *Regulation of Ribosome Synthesis in Escherichia coli: Effects of Temperature and Dilution Rate Changes,* Biotechnol. Bioeng. **52**, 615 (1996).

[168] E. M. El-Mansi and W. H. Holms, *Control of carbon flux to acetate excretion during growth of Escherichia coli in batch and continuous cultures.* J. Gen. Microbiol. **135**, 2875 (1989).

[169] W. A. Wilson, P. J. Roach, M. Montero, E. Baroja-Fernández, F. J. Muñoz, G. Eydallin, A. M. Viale, and J. Pozueta-Romero, *Regulation of glycogen metabolism in yeast and bacteria,* FEMS Microbiol. Rev. **34**, 952 (2010).

[170] E. Levine and T. Hwa, *Stochastic fluctuations in metabolic pathways.* Proc. Natl. Acad. Sci. U.S.A. **104**, 9224 (2007).

[171] A. M. Dean, *A molecular investigation of genotype by environment interactions,* Genetics **139**, 19 (1995).

[172] B. Görke and J. Stülke, *Carbon catabolite repression in bacteria: many ways to make the most out of nutrients.* Nat. Rev. Microbiol. **6**, 613 (2008).

[173] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, *Network motifs in the transcriptional regulation network of Escherichia coli.* Nat. Genet. **31**, 64 (2002).

[174] K. Walsh and D. E. Koshland, *Characterization of rate-controlling steps in vivo by use of an adjustable expression vector,* Proc. Natl. Acad. Sci. U.S.A. **82**, 3577 (1985).

[175] A. Wagner, R. Zarecki, L. Reshef, C. Gochev, R. Sorek, U. Gophna, and E. Ruppin, *Computational evaluation of cellular metabolic costs successfully predicts genes whose expression is deleterious.* Proc. Natl. Acad. Sci. U.S.A. **110**, 19166 (2013).

[176] M.-K. Oh, L. Rohlin, K. C. Kao, and J. C. Liao, *Global expression profiling of acetate-grown Escherichia coli.* J. Biol. Chem. **277**, 13175 (2002).

[177] Z. Wang and J. Zhang, *Impact of gene expression noise on organismal fitness and the efficacy of natural selection.* Proc. Natl. Acad. Sci. U.S.A. **108**, E67 (2011).

[178] R. Knaus and H. Bujard, *PL of coliphage lambda: an alternative solution for an efficient promoter,* EMBO J. **7**, 2919 (1988).

[179] C. M. Hamilton, M. Aldea, B. K. Washburn, P. Babitzke, and S. R. Kushner, *New method for generating deletion and gene replacements in Escherichia coli,* J.Bacteriol. **171**, 4617 (1989).

[180] J. W. Bigger and J. H. Nelson, *The growth of coliform bacilli in distilled water,* J. Pathol. Bacteriol. **53**, 189 (1941).

[181] J. R. Postgate and J. R. Hunter, *The survival of starved bacteria,* J. Gen. Microbiol. **29**, 233 (1962).

[182] T. E. Shehata and A. G. Marr, *Effect of nutrient concentration on the growth of Escherichia coli,* J. Bacteriol. **107**, 210 (1971).

[183] Z. Shajani, M. T. Sykes, and J. R. Williamson, *Assembly of bacterial ribosomes.* Annu. Rev. Biochem. **80**, 501 (2011).

[184] M. Kaczanowska and M. Rydén-Aulin, *Ribosome biogenesis and the translation process in Escherichia coli.* Microbiol. Mol. Biol. Rev. **71**, 477 (2007).

[185] P. P. Dennis, M. Ehrenberg, and H. Bremer, *Control of rRNA synthesis in Escherichia coli: a systems biology approach.* Microbiol. Mol. Biol. Rev. **68**, 639 (2004).

[186] J.-P. Waller and J. I. Harris, *Studies on the composition of the protein from Escherichia coli ribosomes,* Proc. Natl. Acad. Sci. U.S.A. **47**, 18 (1961).

[187] N. Condon, D. Liveris, C. Squires, I. R. A. Schwartz, and C. L. Squires, *rRNA operon multiplicity in Escherichia coli and the physiological implications of rrn inactivation.* J. Bacteriol. **177**, 4152 (1995).

[188] C. Condon, S. French, C. Squires, and C. L. Squires, *Depletion of functional ribosomal RNA operons in Escherichia coli causes increased expression of the remaining intact copies.* EMBO J. **12**, 4305 (1993).

[189] J. M. Zengel, D. Mueckl, and L. Lindahl, *Protein L4 of the E. coli ribosome regulates an eleven gene r protein operon,* Cell **21**, 523 (1980).

[190] M. Nomura, R. Gourse, and G. Baughman, *Regulation of the synthesis of ribosomes and ribosomal components.* Annu. Rev. Biochem. **53**, 75 (1984).

[191] L. Lindahl, *Intermediates and time kinetics of the in vivo assembly of Escherichia coli ribosomes.* J. Mol. Biol. **92**, 15 (1975).

[192] S. S. Chen, E. Sperling, J. M. Silverman, J. H. Davis, and J. R. Williamson, *Measuring the dynamics of E. coli ribosome biogenesis using pulse-labeling and quantitative mass spectrometry,* Mol. Biosyst. **8**, 3325 (2012).

[193] L. Noller, Harry F and Hoffarth, Vernita and Zimniak, *Unusual resistance of peptidyl transferase to protein extraction procedures,* Science **256**, 1416 (1992).

[194] K. Lang, M. Erlacher, D. N. Wilson, R. Micura, and N. Polacek, *The Role of 23S Ribosomal RNA Residue A2451 in Peptide Bond Synthesis Revealed by Atomic Mutagenesis,* Chem. Biol. **15**, 485 (2008).

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

[196] A. Sanamrad, F. Persson, and J. Elf, *Isotropic diffusion of the small ribosomal subunit in Escherichia coli,* arXiv Preprint , 1 (2012), arXiv:1205.5857 .

[197] M. Nomura, *Regulation of Ribosome Biosynthesis in Escherichia coli and Saccharomyces cerevisiae : Diversity and Common Principles,* J. Bacteriol. (1999).

[198] A. Pulk, A. Liiv, L. Peil, Ü. Maiväli, K. Nierhaus, and J. Remme, *Ribosome reactivation by replacement of damaged proteins,* Mol. Microbiol. **75**, 801 (2010).

[199] T. Giessen and M. Marahiel, *The tRNA-Dependent Biosynthesis of Modified Cyclic Dipeptides,* Int. J. Mol. Sci. **15**, 14610 (2014).

[200] L. U. Magnusson, A. Farewell, and T. Nyström, *ppGpp: a global regulator in Escherichia coli.* Trends Microbiol. **13**, 236 (2005).

[201] K. Potrykus and M. Cashel, *(p)ppGpp: still magical?* Annu. Rev. Microbiol. **62**, 35 (2008).

[202] J. Lemke, P. Sanchez-Vazquez, H. Burgos, G. Hedberg, W. Ross, and R. Gourse, *Direct regulation of Escherichia coli ribosomal protein promoters by the transcription factors ppGpp and DksA,* Proc. Natl. Acad. Sci. U.S.A. **108**, 5712 (2011).

[203] J. A. Gallant, *Stringent control in E. coli,* Annu. Rev. Genet. **13**, 393 (1979).

[204] M. M. Barker, T. Gaal, C. A. Josaitis, and R. L. Gourse, *Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro.* J. Mol. Biol. **305**, 673 (2001).

[205] A. Perederina, V. Svetlov, M. N. Vassylyeva, T. H. Tahirov, S. Yokoyama, I. Artsimovitch, and D. G. Vassylyev, *Regulation through the secondary channel - Structural framework for ppGpp-DksA synergism during transcription,* Cell **118**, 297 (2004).

[206] B. J. Paul, M. M. Barker, W. Ross, D. A. Schneider, C. Webb, J. W. Foster, R. L. Gourse, and H. Mall, *DksA: A critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP,* Cell **118**, 311 (2004).

[207] W. Ross, J. F. Thompson, J. T. Newlands, and R. L. Gourse, *E.coli Fis protein activates ribosomal RNA transcription in vitro and in vivo.* EMBO J. **9**, 3733 (1990).

[208] T. Gaal, M. S. Bartlett, W. Ross, C. L. Turnbough, and R. L. Gourse, *Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria.* Science **278**, 2092 (1997).

[209] H. D. Murray, D. A. Schneider, and R. L. Gourse, *Control of rRNA expression by small molecules is dynamic and nonredundant,* Mol. Cell **12**, 125 (2003).

[210] J. R. Cole and M. Nomura, *Changes in the half-life of ribosomal protein messenger RNA caused by translational repression,* J. Mol. Biol. **188**, 383 (1986).

[211] B. J. Paul, W. Ross, T. Gaal, and R. L. Gourse, *rRNA transcription in Escherichia coli.* Annu. Rev. Genet. **38**, 749 (2004).

[212] J. Justesen, T. Lund, F. S. Pedersen, and N. O. Kjeldgaard, *The physiology of stringent factor (ATP: GTP 3'-diphosphotransferase) in Escherichia coli,* Biochimie **68**, 715 (1986).

[213] H. Xiao, M. Kalman, K. Ikehara, S. Zemel, and M. Cashel, *Residual Guanosine 3 ', 5 ' -Bispyrophosphate Synthetic Activity of relA Null Mutants Can Be Eliminated by spoT Null Mutations.* Biol. Chem. **266**, 5980 (1991).

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

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

[216] S. Metzger, G. Schreiber, E. Aizenman, M. Cashel, and G. Glaser, *Characterization of the relA1 Mutation and a Comparison of relA1 with new relA Null Alleles in Escherichia coli,* J. Biol. Chem. **264**, 21146 (1989).

[217] A. Battesti and E. Bouveret, *Acyl carrier protein/SpoT interaction, the switch linking SpoT-dependent stress response to fatty acid metabolism,* Mol. Microbiol. **62**, 1048 (2006).

[218] A. Bougdour and S. Gottesman, *ppGpp regulation of RpoS degradation via anti-adaptor protein IraP.* Proc. Natl. Acad. Sci. U.S.A. **104**, 12896 (2007).

[219] J. Ryals, R. Little, and H. Bremer, *Control of rRNA and tRNA syntheses in Escherichia coli by guanosine tetraphosphate,* J. Bacteriol. **151**, 1261 (1982).

[220] T. Gaal and R. L. Gourse, *Guanosine 3'-diphosphate 5'-diphosphate is not required for growth rate-dependent control of rRNA synthesis in Escherichia coli.* Proc. Natl. Acad. Sci. U.S.A. **87**, 5533 (1990).

[221] K. Potrykus, H. Murphy, N. Philippe, and M. Cashel, *ppGpp is the major source of growth rate control in E. coli.* Environ. Microbiol. **13**, 563 (2011).

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

[223] G. Bokinsky, E. E. K. Baidoo, S. Akella, H. Burd, D. Weaver, J. Alonso-Gutierrez, H. García-Martín, T. S. Lee, and J. D. Keasling, *HipA-triggered growth arrest and β-lactam tolerance in Escherichia coli are mediated by RelA-dependent ppGpp synthesis,* J. Bacteriol. **195**, 3173 (2013).

[224] G.-W. Li, D. Burkhardt, C. Gross, and J. S. Weissman, *Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources.* Cell **157**, 624 (2014).

[225] M. Ehrenberg, H. Bremer, and P. P. Dennis, *Medium-dependent control of the bacterial growth rate.* Biochimie **95**, 643 (2013).

[226] J. Elf and M. Ehrenberg, *Near-critical behavior of aminoacyl-tRNA pools in E. coli at rate-limiting supply of amino acids.* Biophys. J. **88**, 132 (2005).

[227] B. Bohrmann, W. Villiger, R. Johansen, and E. Kellenberger, *Coralline shape of the bacterial nucleoid after cryofixation,* J. Bacteriol. **173**, 3149 (1991).

[228] P. P. Dennis and H. Bremer, *Differential rate of ribosomal protein synthesis in Escherichia coli B/r.* J. Mol. Biol. **84**, 407 (1974).

[229] H. J. Weber, *Stoichiometric measurements of 30S and 50S ribosomal proteins from Escherichia coli.* Mol. Gen. Genet. **119**, 233 (1972).

[230] S. J. Hardy, *The stoichiometry of the ribosomal proteins of Escherichia coli.* Mol. Gen. Genet. **140**, 253 (1975).

[231] M. Tal, I. Weissman, and A. Silberstein, *A new method for stoichiometric analysis of proteins in complex mixture–reevaluation of the stoichiometry of E. coli ribosomal proteins.* J. Biochem. Biophys. Methods **21**, 247 (1990).

[232] S. T. Liang, M. Ehrenberg, P. Dennis, and H. Bremer, *Decay of rplN and lacZ mRNA in Escherichia coli.* J. Mol. Biol. **288**, 521 (1999).

[233] G. Schreiber, S. Metzger, E. Aizenman, S. Roza, M. Cashel, and G. Glaser, *Overexpression of the relA Gene in Escherichia coli,* J. Biol. Chem. **266**, 3760 (1991).

[234] S. Liang, M. Bipatnath, Y. Xu, S. Chen, P. Dennis, M. Ehrenberg, and H. Bremer, *Activities of constitutive promoters in Escherichia coli.* J. Mol. Biol. **292**, 19 (1999).

[235] S. Quan, O. Skovgaard, and C. L. Squires, *Escherichia coli Ribosomal RNA deletion strain,* Unpublished. Available online at Coli Genetic Stock Center. Retrieved from http://cgsc.biology.yale.edu/Reference.php?ID=122787 (2010).

[236] T. Bollenbach, S. Quan, R. Chait, and R. Kishony, *Nonoptimal microbial response to antibiotics underlies suppressive drug interactions.* Cell **139**, 707 (2009).

[237] H. Dong, L. Nilsson, and C. G. Kurland, *Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction.* J. Bacteriol. **177**, 1497 (1995).

[238] I. Shachrai, A. Zaslaver, U. Alon, and E. Dekel, *Cost of Unneeded Proteins in E. coli Is Reduced after Several Generations in Exponential Growth,* Mol. Cell **38**, 758 (2010).

[239] M. Kitagawa, T. Ara, M. Arifuzzaman, T. Ioka-Nakamichi, E. Inamoto, H. Toyonaga, and H. Mori, *Complete set of ORF clones of Escherichia coli ASKA library (A Complete Set of E. coli K-12 ORF Archive): Unique Resources for Biological Research,* DNA Res. **12**, 291 (2006).

[240] H. Funabashi, T. Haruyama, M. Mie, Y. Yanagida, E. Kobatake, and M. Aizawa, *Non-destructive monitoring of rpoS promoter activity as stress marker for evaluating cellular physiological status.* J. Biotechnol. **95**, 85 (2002).

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

[242] B.-S. Chen, C.-Y. Hsu, and J.-J. Liou, *Robust design of biological circuits: evolutionary systems biology approach.* J. Biomed. Biotechnol. **2011**, 304236 (2011).

[243] S. P. Mizrahi, O. Sandler, L. Lande-Diner, N. Q. Balaban, and I. Simon, *Distinguishing between stochasticity and determinism: Examples from cell cycle duration variability,* BioEssays **38**, 8 (2016).

[244] B. Snijder and L. Pelkmans, *Origins of regulated cell-to-cell variability,* Nat. Rev. Mol. Cell Biol. **12**, 119 (2011).

[245] O. Sandler, S. P. Mizrahi, N. Weiss, O. Agam, I. Simon, and N. Q. Balaban, *Lineage correlations of single cell division time as a probe of cell-cycle dynamics,* Nature **519**, 468 (2015).

[246] P. Grassberger, *Characterization of Strange Attractors,* Phys. Rev. Lett. **50**, 346 (1983).

[247] D. L. Jones, R. C. Brewster, and R. Phillips, *Promoter architecture dictates cell-to-cell variability in gene expression.* Science **346**, 1533 (2014).

[248] R. L. Strack, W. Song, and S. R. Jaffrey, *Using Spinach-based sensors for fluorescence imaging of intracellular metabolites and proteins in living bacteria.* Nat. Protoc. **9**, 146 (2014).

SUMMARY

Life of single cells is not deterministic, but virtually all processes in a cell are subject to stochastic fluctuations. As consequence, genetically identical cells, which are living in the same environment, can behave differently. They can, for example, produce different amounts of proteins, grow at different rates and can even specialize into completely different phenotypes. The discovery of cell-to-cell variability opened up many questions, ranging from what are the origins of the fluctuations, to whether its consequences are detrimental or beneficial for cells.

The aim of this thesis is to better understand stochasticity in gene expression and growth rate of bacterial cells. To this end, we investigate whether, next to gene expression, also growth rate of single cells fluctuates dynamically, which we will show to be the case. We then focus on the following questions about gene expression and growth: What are the origins of fluctuations? Can fluctuations propagate from one to the other? To address these questions, we work with a model organism, the bacterium *E. coli* . This organism is simple enough to allow for quantitative experiments, but already so complex that many processes inside these cells have not been understood yet.

Studying stochastic fluctuations in gene expression and growth rate in single cells requires precise and rather high throughput tools to measure these quantities. In chapter 2 and 3 we describe these methods. By using automated time-lapse microscopy, we acquire movies of growing *E. coli* cells which are expressing fluorescent proteins. We then use automated analysis software to segment, track and analyze cells. To accurately measure gene expression, a fluorescent protein reporter of high quality is needed. Therefore, we present a comparison of different fluorescent proteins which assesses their suitability for time-lapse experiments.

In chapter 4 we investigate the general interdependence of fluctuations in gene expression and growth rate. We find that the noise intensities of both fluctuations are very strongly correlated and that they scale linearly. In apparent contrast to that, the fluctuating time traces are shown to be only modestly correlated. We develop a linear noise model to explain these observations and analyze what constraints the linear scaling imposes on such models. A central result is that the intensities of different cellular noise sources do not change independently, but are set by one single global parameter.

In chapter 5 we study the effects of a specific source of fluctuations in gene expression: the bacterial cell cycle. As cells grow and prepare for division, they copy their chromosome. With two copies of each gene and an increasing cell size, the production rate of proteins increases. We show that about half of the noise in protein production rate is caused by gene duplication. In contrast to that, protein concentration is hardly affected by the cell cycle because the exponential volume increase almost perfectly cancels the increase in protein production rate.

The fact that gene expression fluctuates has been known for many years. However, it was unclear whether such fluctuations also affect the growth rate of cells. This ques-

tion is addressed in chapter 6. We show that fluctuations in enzymes can indeed propagate and cause growth fluctuations. Conversely, growth fluctuations also propagate back to disturb protein concentration. We develop an analytical model to accurately predict noise transmission. Our results indicate that fluctuations can not only propagate through gene networks but also through metabolic reactions. They also suggest that cellular metabolism is inherently stochastic.

Finally, in chapter 7 we investigate the influence of fluctuations in ribosome concentration on the growth rate. Because ribosomes are the molecular machines responsible for protein production, they play a central role in cellular growth. Each cell needs thousands of ribosomes and as producing these large numbers is costly, the ribosome production is tightly regulated and adjusted to the need. Still, their concentration in single cells fluctuates, and we here investigate whether these fluctuations affect growth rate. Our data suggests that fluctuations in ribosome concentration do not propagate to growth, which indicates that ribosomes are not dynamically limiting growth rate.

SAMENVATTING

Het leven van individuele cellen is niet deterministisch, in praktisch alle processen in de cel spelen stochastische fluctuaties een rol. Hierdoor kunnen cellen die genetisch identiek zijn en zich in dezelfde omgeving bevinden, zich toch anders gedragen. Ze kunnen bijvoorbeeld verschillende hoeveelheden van bepaalde eiwitten produceren, op verschillende snelheden groeien, en zich zelfs tot compleet verschillende fenotypen ontwikkelen. De ontdekking van deze variabiliteit tussen cellen leidde tot veel vragen: variërend van waar deze fluctuaties vandaan komen, tot of dit goede of slechte consequenties heeft voor de cel.

Het doel van dit proefschrift is beter te leren begrijpen welke invloed stochasticiteit heeft op genexpressie en de groeisnelheid van bacteriecellen. Daarom onderzoeken we of - naast genexpressie – ook groeisnelheid van individuele cellen dynamisch fluctueert. Dit blijkt inderdaad het geval te zijn. Vervolgens kijken we naar de volgende vragen rond genexpressie en groei: wat zijn de oorzaken van de fluctuaties? Kunnen fluctuaties propageren van de ene naar de andere parameter? Om deze vragen te onderzoeken gebruiken we het modelorganisme *E. coli* . Dit organisme is simpel genoeg om kwantitatieve experimenten te faciliteren, maar al zo complex dat veel processen in deze cellen onbegrepen zijn.

Het bestuderen van stochastische fluctuaties in genexpressie en groeisnelheid in individuele cellen vereist precieze, *high throughput* meetinstrumenten. In hoofdstuk 2 en 3 beschrijven we de methoden om deze precieze, *high throughput* metingen te doen. We gebruiken geautomatiseerde *time-lapse* microscopie om filmpjes te verkrijgen van groeiende *E. coli* cellen die fluorescente eiwitten tot expressie brengen. Een computeralgoritme segmenteert, volgt en analyseert de cellen vervolgens. Om genexpressie accuraat te kunnen meten is een fluorescent reporter eiwit van hoge kwaliteit vereist. Daarom vergelijken we verschillende fluorescente eiwitten wat betreft hun geschiktheid voor *time-lapse* experimenten.

In hoofdstuk 4 onderzoeken we de onderlinge afhankelijkheid van fluctuaties in genexpressie en groeisnelheid. We laten zien dat de intensiteit van de ruis in deze beide grootheden sterk is gecorreleerd. Bovendien is er een lineair verband tussen de intensiteit van de expressieruis en de intensiteit van de groeiruis. In ogenschijnlijk contrast hiermee zijn de signalen over de tijd slechts matig gecorreleerd. We hebben een lineair model ontwikkeld dat deze observaties verklaart, en bekijken tot welke beperkingen de lineaire schaling leidt binnen deze modellen. Een belangrijk resultaat is dat de intensiteiten van verschillende bronnen van ruis niet onafhankelijk van elkaar veranderen, maar bepaald worden door één centrale parameter.

In hoofdstuk 5 bestuderen we de effecten op genexpressie van een specifieke bron van ruis: de bacteriële celcyclus. Wanneer cellen groeien en zich voorbereiden op de celdeling, kopiëren zij hun chromosoom. Ten gevolge van de verdubbeling van genen en de toename in celgrootte, neemt de productiesnelheid van eiwitten toe. Wij tonen hier

aan dat ongeveer de helft van de ruis in eiwitexpressie veroorzaakt wordt door genduplicatie. Daarentegen wordt de concentratie van eiwitten nauwelijks beïnvloed door de celcyclus omdat de toegenomen productiesnelheid van eiwitten vrij nauwkeurig wordt gecompenseerd door de exponentiele toename in celvolume.

Dat genexpressie fluctueert is al jaren bekend. Het was echter niet duidelijk of deze fluctuaties ook een effect op de groeisnelheid van cellen hebben. In hoofdstuk 6 gaan we in op deze vraag. We laten zien dat fluctuaties in enzymconcentraties zich inderdaad kunnen propageren en fluctuaties in de groeisnelheid veroorzaken. Omgekeerd kunnen schommelingen in de groeisnelheid van de cel ook de eiwitconcentraties in de cel verstoren. We hebben een analytisch model ontwikkeld waarmee we accuraat de voortplanting van de ruis kunnen voorspellen. Onze resultaten tonen aan dat fluctuaties zich niet alleen via gen-netwerken propageren maar ook via metabole reacties. Daarnaast duiden onze resultaten erop dat het metabolisme inherent stochastisch is.

Als laatste, in hoofdstuk 7, onderzoeken we het effect van fluctuaties in de ribosoomconcentratie op de groeisnelheid. Omdat ribosomen de moleculaire machines zijn die verantwoordelijk zijn voor eiwitproductie, spelen ze een belangrijke rol bij de celgroei. Elke cel heeft duizenden ribosomen nodig en de productie van deze grote aantallen is zeer kostbaar. Daarom is de ribosoomproductie strikt gereguleerd en wordt deze op de vraag afgestemd. Desondanks fluctueren ribosoomconcentraties in individuele cellen en hier onderzoeken we of deze fluctuaties de groeisnelheid beïnvloeden. Onze data toont de afwezigheid van propagatie van fluctuaties in ribosoomconcentraties naar celgroei. Dit suggereert dat ribosomen niet dynamisch limiterend voor de groeisnelheid zijn.

Graag wil ik Joris Paijmans en Martijn Wehrens bedanken voor hun Nederlandse vertaling van de samenvatting.

ACKNOWLEDGEMENTS

Doing a PhD is an intense experience. Lots of ups and downs and quite often it is unclear where this whole thing is actually heading. But luckily, there were a lot of great people along this journey who helped with support and advice or simply with making life fun and good, so let me take this chance and thank you here.

Amolf is a great place to work at and this definitely had its own contribution to successfully completing this thesis: many motivated people, an open and constructive atmosphere, lots of technical and administrative support and, not to forget, the amazing green building. I would like to thank my promoter Sander for giving me the chance to switch from theory and to start working in an experimental lab. I much appreciated your talent to view projects from a new angle, to see the bigger picture and to find the story in the data. Our sometimes intense discussions taught me to think beyond my original ideas and to work out proper argumentation lines. Next, a big thank you goes to all the biophysics group members, from the cell side and the tweezers side (and please don't forget that you are one group), for the good times that we shared over the last years.

To the bacteria people, especially Martijn and Dmitry and also Philippe (who is somehow still involved): it was good to team up with you in the effort to understand more about cells and to have you around when once again struggling with non-cooperating bacteria. Katja, thanks for all the open and critical discussions, scientific and non-scientific. I much appreciated your common sense and independent thoughts. Vanda, thank you for your incredible patience and endurance when working yourself through long strain-construction-to-do-lists, and for all your stories. Fatemeh, we started this PhD adventure almost at the same time and even in the same office. Thanks a lot for your kindness and the chance to share the latest little scientific dramas. Sebastian, the lunch conversations definitely profited from your unconventional input. Joris en Martijn, bedankt dat jullie in het Nederlands met mij praten. And for the birthday-office-decorations, late-hour coffee breaks and long random conversations.

There are many people who make Amolf and Amsterdam a great place to be at, during and after work: Aditya I., Galja, Simone, Agata, Nicola, Giorgos, Sebastian, Mario, Olga, Cristina, Stef, Yuval, Joris, Martijn, Aditya V.. Plus the "old" people that left Amolf already: Stephen, Katja, Maga, Jeanette, Núria, Ioana, Lutz, Stephan, Felipe, José, Tomek, Milena, David. Plus the people that I forgot to mention but whose names should be written here (which I will probably realize once the thesis is in the printer): Thanks for the coffee breaks, squash sessions, board game evenings, dinners, concerts and everything else! I guess I should at this place also acknowledge the coffee machine, as well as Wikipedia, who both contributed significantly to the success of this thesis.

Philipp G. and Philipp S., forcing me to still have a social life during the last and intense PhD phase kept me going during that time and is much appreciated! Thanks for all the relevant and not so relevant discussions during the last years, lots of fun and of course our amazing Wuidsau uitjes (best decision ever). To you two and to my climbing

girls Eleni and Kersti: I much enjoyed some great climbing and traveling with you and your enthusiasm for good ideas and for bringing them into action.

And finally, a very big thank you to my family! Philipp, we share the same sense of humor and irony, so, among many other things, thanks for some refreshing comments! Papa, you were always there for the family and I miss you. Mama, thanks a lot for always listening to the stories and giving advice when things here got a bit stressful and chaotic. To my parents and my brother, thanks for everything!

LIST OF PUBLICATIONS

RELATED TO THIS THESIS

Chapter 4
N. Walker, P. Nghe, and S. J. Tans
Interdependence of gene expression and growth rate fluctuations in bacteria
manuscript in preparation

Chapter 5
N. Walker, P. Nghe, and S. J. Tans
Generation and filtering of gene expression noise by the bacterial cell cycle
BMC Biology **14**, 1 (2016)

Chapter 6
D. J. Kiviet*, P. Nghe*, **N. Walker**, S. Boulineau, V. Sunderlikova, and S. J. Tans
Stochasticity of metabolism and growth at the single-cell level
Nature **514**, 376 (2014) * equal contribution

OTHER PUBLICATIONS

D. Ershov, **N. Walker**, R. Rozendaal, and S. J. Tans
Size control in *Escherichia coli* at cell division
manuscript in preparation

CURRICULUM VITÆ

Noreen Walker was born on 9 September 1985 in Munich, Germany. From 1996 to 2005 she attended the high school ("Gymnasium") in Geretsried. After graduation, she started studying physics at the Ludwig-Maximilians-Universität in Munich. During her studies, she spent from 2008 to 2009 one year at the École Polytechnique Fédéderale de Lausanne in Switzerland. There, she pursued two research projects, next to attending lectures. One project concerned the modeling of reinforcement learning in the context of neural networks and was carried out in the lab of Computational Neuroscience, led by prof. Wulfram Gerstner. In the second project she developed simulations to model the scaling behavior of polymers in different dimensions. This project was under supervision of Prof. Paolo de los Rios in the lab of Statistical Biophysics. After returning to Munich in 2009, she did a summer internship on evolutionary game theory at the chair of Statistical and Biological Physics under supervision of Prof. Erwin Frey. During her studies, she tutored multiple classes in mathematics. From 2010 until 2011 she worked on her Diplom (master) thesis, in which she studied heterogeneous timing of gene induction as a potential regulation strategy. The work was carried out under the supervision of Prof. Ulrich Gerland at the chair of Statistical and Biological Physics. She obtained her Diplom (M.Sc.) degree in 2011 and graduated with the best achievable grade. Her areas of specialization were mathematics and biophysics. In fall 2011 she started her PhD in the Biophysics group of prof. Sander Tans at the FOM institute AMOLF in Amsterdam.