
Cell-to-cell variability of gene expression dynamics in inducible regulatory networks

Judith Megerle



München 2011

Cell-to-cell variability of gene expression dynamics in inducible regulatory networks

Judith Megerle

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Judith Megerle
aus Regensburg

München, den 25. Januar 2011

Erstgutachter: Prof. Dr. Joachim Rädler

Zweitgutachter: Prof. Dr. Ulrich Gerland

Tag der mündlichen Prüfung: 2. März 2011

Contents

Zusammenfassung	1
Summary	3
1 Introduction	5
2 Basic concepts	9
2.1 Gene expression and transcriptional regulation	9
2.2 Gene regulatory networks	10
2.3 Stochasticity in gene expression	10
2.4 Noise: Use or nuisance?	13
2.5 Modeling gene expression	14
2.6 Fluorescent proteins as reporters for gene expression	17
3 Quantitative time-lapse fluorescence microscopy	19
3.1 Fluorescence microscopy	20
3.2 Sample environment and flow system	21
3.3 Image analysis of time-lapse movies	23
3.4 Fluorescence signal and bacterial autofluorescence	29
4 Determination of the GFP maturation time on the single cell level	31
5 Timing and dynamics of gene expression in the arabinose system	33
5.1 The arabinose utilization system	34
5.2 Heterogeneous timing of single cell gene induction	37
5.2.1 Single cell induction kinetics	37
5.2.2 Deterministic gene expression function	39
5.2.3 Distribution of GFP expression rate and intrinsic delay time	41
5.2.4 Stochastic model for the uptake module	43
5.2.5 Single copy reporter	47
5.2.6 Heterogeneous timing in a strain capable of arabinose degradation	48
5.3 Variation of the uptake protein expression	51
5.3.1 Single cell induction kinetics	52
5.3.2 Differences between native and arabinose independent uptake protein expression	57

5.4	A quantitative model of the arabinose system	59
5.4.1	Deterministic rate equations	59
5.4.2	Arabinose efflux	62
5.4.3	Comparison of modeling results and induction kinetics with native and arabinose independent transporter expression	66
5.4.4	Comparison of modeling results and the expression dynamics following arabinose pulses	66
5.4.5	Scope of the model and future experiments	69
5.5	Cell density dependent gene expression	71
5.5.1	Distributions of single cell fluorescence	71
5.5.2	Causes of density dependence	74
5.6	Conclusions	76
6	Dynamics of AHL mediated quorum sensing	79
6.1	Introduction to quorum sensing	80
6.2	Gene expression dynamics	80
6.2.1	Experimental system	81
6.2.2	Fluorescence decrease	82
6.2.3	Induction with external AHL	84
6.2.4	Comparison of the data and a rate equation model	86
6.2.5	Colony-to-Colony and Cell-to-Cell variability	88
6.3	Discussion	90
7	Outlook	91
A	Arabinose uptake	95
B	Experimental details	99

Zusammenfassung

Induzierbare Genregulationsnetzwerke in Bakterien schalten auf einen Reiz aus der Umgebung hin die Transkription von Genen an. Die inhärente Stochastizität der Genexpression führt zu einer Variabilität von Zelle zu Zelle, welche Funktionen wie die Arbeitsteilung zwischen genetisch identischen Zellen ermöglicht. Über den Einfluss der Stochastizität auf den zeitlichen Ablauf der Proteinproduktion ist jedoch nur wenig bekannt. In dieser Arbeit wurde die Zelle-zu-Zelle-Variabilität der Dynamik der Genexpression eines metabolischen Systems untersucht sowie eines Systems, das zu einer Änderung der Morphologie der Bakterienkultur führt. Auf der Basis quantitativer Daten wurden mathematische Modelle aufgestellt, die die Funktionsweise aufgrund der Netzwerkstruktur vorhersagbar machen.

Fluoreszierende Proteine dienen als Reporter um die Dynamik der Genexpression einzelner Zellen mit Hilfe von quantitativer zeitaufgelöster Fluoreszenzmikroskopie zu messen. Für diese Technik wurde ein Aufbau entwickelt, der die gleichzeitige Aufnahme einer großen Anzahl von Einzelzellkurven ermöglicht. Zudem kann die Konzentration von Signalsubstanzen in der verwendeten Probenumgebung zeitlich variiert werden. Dies erlaubt die Bestimmung der Maturationszeit des fluoreszierenden Proteins in einzelnen Bakterien. Die Kenntnis dieser Größe war nötig um die Dynamik des Netzwerks von der des Reporters zu unterscheiden. Für *GFPmut3* wurde eine Maturationszeit von 6.5 ± 0.6 min bestimmt.

In *Escherichia coli* wird die Produktion von Proteinen für den Abbau und die Aufnahme des Zuckers Arabinose durch ein eigenes System gesteuert. In Abwesenheit des Zuckers ist die Produktionsrate der Aufnahme Proteine klein, sie steigt aber stark an, wenn die intrazelluläre Arabinosekonzentration einen Schwellwert übersteigt. Bei der Zugabe von Arabinose zu Bakterien, die zuvor ohne diesen Zucker gewachsen waren, wurde eine Variabilität von Zelle zu Zelle bezüglich des Einsetzens der Genexpression beobachtet. Für diese Stochastizität im zeitlichen Ablauf wurde der Begriff "heterogenes Zeitverhalten" geprägt. Die Verteilung der Verzögerungszeiten zwischen der Zugabe des Induktors und dem Beginn der Expression skaliert invers mit der externen Arabinosekonzentration und wird durch ein einfaches stochastisches Modell der Arabinoseaufnahme erklärt. Diese Ergebnisse weisen darauf hin, dass das heterogene Zeitverhalten auf eine breite Verteilung der Aufnahme Proteine zum Zeitpunkt der Arabinosezugabe zurückzuführen ist.

Das Netzwerk wurde genetisch manipuliert um weitere Hinweise auf diesen Zusammenhang zu erhalten. In der Mutante ist die Anzahl der Aufnahme Proteine pro Zelle ebenfalls unterschiedlich, bleibt aber zeitlich konstant. Interessanterweise begann die Genexpression in allen Zellen gleichzeitig, aber ihre Rate, die im nativen Netzwerk konstant war, sank mit abnehmender Arabinosekonzentration. Ein Ratenmodell wurde aufgestellt, welches die

Genexpressionsdynamiken der beiden Netzwerke sowie das Antwortverhalten auf aufeinander folgende Arabinosepulse konsistent beschreibt. Die Dynamik des modifizierten Netzwerks kann nur erklärt werden, wenn das Ausströmen der Arabinose im Modell berücksichtigt wird. Die Berücksichtigung dieses Prozesses wurde durch die Beobachtung veranlasst, dass die Genexpression endet, sobald die Arabinose aus der Umgebung vorher induzierter Zellen entfernt wird.

Da einige stochastische Effekte zu optimalen Wachstumsraten der Bakterienkulturen führen oder Teilpopulationen vor seltenen schädlichen Effekten schützen, werden mögliche Vorteile des heterogenen Zeitverhaltens diskutiert. Vermutlich treten solche Effekte auf, wenn die Arabinosekonzentration zeitlichen Fluktuationen unterworfen ist.

Im *PPU*-System, welches die Bildung von Biofilmen in *Pseudomonas putida* steuert, wird der Induktor von den Bakterien selbst produziert und kann durch die Membran diffundieren. Es wurde untersucht, ob heterogene räumliche Verteilungen der Zellen oder der Induktorkonzentration die Geneexpressionsdynamik beeinflussen. Wurde der Induktor von außen zu einer Bakterienkultur hinzugefügt, die anschließend nicht verändert wurde, sodass die von den Zellen produzierten Induktoren sich ansammeln konnten. Die beobachtete Dynamik stimmt gut mit einem Modell überein, welches eine gut gemischte Umgebung voraussetzt. Unterschiede zwischen den Vorhersagen dieses Modells und den Daten bei konstanter externer Konzentration deuten darauf hin, dass die Konzentration in den Zellen größer ist als in der Umgebung. Ein Modell mit zwei räumlich getrennten Bereichen soll entscheiden, welcher von mehreren vorgeschlagenen Mechanismen zur Induktoransammlung führt.

Zwischen verschiedenen Kolonien und zwischen einzelnen Zellen innerhalb einer Kolonie wurden erhebliche Variationen der Genexpression beobachtet. Diese werden vermutlich durch die Kombination von stochastischer Genexpression und räumlicher Heterogenität der Induktorkonzentration verursacht. Die Variabilität könnte zu der bereits beschriebenen funktionalen Differenzierung von Zellen in Biofilmen beitragen.

Summary

Inducible gene regulatory networks in bacteria switch on the transcription of genes upon an environmental stimulus. The inherent stochasticity of gene expression leads to cell-to-cell variability that enables functions such as the division of labor between genetically identical cells. However, little is known about the impact of stochasticity on the timing of protein production. In this thesis the cell-to-cell variability of the dynamics of gene expression in a metabolic system and a system which mediates a change in the morphology of the bacterial culture was studied. Quantitative data were used to establish mathematical models which allow predicting the functionality from the network structure.

Fluorescent proteins are used as reporters to measure single cell gene expression dynamics by quantitative time-lapse fluorescence microscopy. A setup for this technique was established that allows for the simultaneous acquisition of a large number of single cell traces. In addition, the sample environment enables temporal variations of signaling molecule concentrations. It was thus possible to measure the maturation time of a fluorescent protein in single bacteria. This quantity is necessary to distinguish between network and reporter dynamics. For *GFPmut3* a maturation time of 6.5 ± 0.6 min was determined.

In *Escherichia coli* the production of proteins for the degradation and the uptake of the sugar arabinose is controlled by a specific system. In the absence of the sugar, the production rate of the uptake proteins is small, but it increases strongly once the intracellular arabinose concentration crosses a threshold level. When arabinose was added to bacteria which had previously grown without this sugar, cell-to-cell variability was found for the time at which gene expression starts. The phrase coined for this temporal stochastic effect is heterogeneous timing. The distribution of the delay times between inducer addition and expression onset scales inversely with the external arabinose concentration and can be explained by a simple stochastic model of arabinose uptake. These results indicate that heterogeneous timing is causally related to a broad distribution of arabinose uptake proteins at the time of inducer addition.

The network was genetically modified in order to obtain further evidence for this relationship. In the mutant, the number of uptake proteins varies between cells, too, but remains constant over time. Interestingly, gene expression started simultaneously in all cells, but the rate of gene expression, which was constant in the native network, became smaller with decreasing inducer concentration. A rate equation model was developed, which consistently explains the gene expression dynamics in both networks and also the response to subsequent arabinose pulses. Importantly, the dynamics of the modified network are only predicted correctly when arabinose efflux is considered in the model. The

inclusion of this process was prompted by the observation that gene expression ceases when arabinose is removed from previously induced cells.

As some stochastic effects lead to an optimal growth rate of bacterial cultures or protect a fraction of a population from rare detrimental events, possible benefits resulting from heterogeneous timing are discussed. Such effects will probably be observed when the arabinose concentration varies over time.

In the *PPU* system, which controls biofilm formation in *Pseudomonas putida*, the inducer is produced by the cells themselves and can diffuse through the membrane. It is investigated whether heterogeneous spatial distributions of the cells or the inducer concentration have an impact on the gene expression dynamics. Dynamics were measured when the inducer was applied externally to a subsequently undisturbed sample, so that inducer molecules released by the cells could accumulate. These data were in good agreement with a model assuming a well-mixed environment. Differences between the predictions of this model and data acquired when the external concentration was kept constant indicated a higher inducer concentration within the cells than in the surrounding medium. A two compartment model will be developed in order to analyze which of the proposed mechanisms leads to accumulation.

Significant variations of gene expression between single colonies and between single cells within colonies were found. Presumably, these result from a combination of stochasticity in gene expression and spatially heterogeneous inducer concentrations. The variations might contribute to the previously observed functional differentiation of cells in biofilms.

1 Introduction

The use of genetically modified microorganisms is a highly promising approach for applications such as the production of biofuels from waste biomass and the removal of environmental contaminations. Many examples prove its technological potential: A classical application of genetically manipulated bacteria is the production of insulin. More recently, the cost of highly effective anti-malaria drugs could be reduced strongly by the use of modified yeast cells [1].

However, in contrast to disciplines such as mechanical or electrical engineering there is no framework which allows designing a microbe with a predefined functionality. Instead, a large number of trial and error steps are necessary that are tedious, costly and time-consuming. Thus, the goal is to develop an engineering framework for genetic reprogramming [2]. A prerequisite for this effort is more knowledge about the fundamental working principles of cells.

Systems biology [3] aims at unraveling these principles by establishing a holistic description of biological systems. To this end, the involved molecules and their interactions as well as the response to perturbations are characterized quantitatively. These data are used to establish mathematical models that facilitate predicting the behavior of a system from its structure.

A very important cellular function is signal transduction whereby the cellular protein content is changed upon an external stimulus. Single celled organisms are thus able to adapt to environmental changes. To alter the protein content, the transcription rate of certain genes is modified. These modifications are mediated by regulatory networks that comprise several genes and proteins and their regulatory interactions. It has turned out that these networks have a modular structure and that certain regulatory motifs recur frequently [4]. An example of such a motif is positive feedback, i.e. the stimulation of a gene by its own gene product, which also characterizes the networks analyzed in this study. In many cases, positive feedback leads to the conversion of a graded input to a binary signal, thus constituting a switch.

Even though this framework resembles man-made signaling tools at first glance, there is an important additional influence: The response is subject to a high degree of stochasticity as many of the relevant molecules are present in very low copy numbers per cell, leading to significant cell-to-cell variations [5]. In some cases, genetically identical cells within one population are in physiologically completely different states. For example, under certain conditions a fraction of the cells express some genes at a high level, while other cells do not express these genes at all. This behavior is called all-or-nothing gene expression and

has been observed in two catabolic systems, the arabinose and lactose system [6] [7].

Many fundamental questions, such as the generation of noise, the transmittance of noise through gene cascades or the stochastic switching into distinct states have been addressed in recent years [5] [8]. This analysis has been facilitated by the discovery of fluorescent proteins [9]. These can be used as reporters for gene expression since the gene encoding a fluorescent protein can be put under the control of a transcriptional regulatory network. As the fluorescence signal of each individual cell can be measured, the cell-to-cell variations in gene expression can be analyzed. Furthermore, time-lapse fluorescence microscopy can be used to study gene expression dynamics on the single cell level [10]: Ensembles of bacteria are imaged in regular time intervals and fluorescence time series of individual cells are determined by quantitative image analysis.

The shaping of gene networks by evolution, is another topic which is currently addressed both experimentally and by using theoretical concepts such as cost-benefit and game theory [11] [12]. Particularly, it is investigated under which environmental conditions a certain network is optimal, meaning that it yields maximal fitness of the bacteria. In this context it is suggested that phenotypic variability created by noise is optimal under certain conditions [13].

In this thesis the response dynamics of bacterial regulatory networks are analyzed on the single cell level. The investigated systems switch on the expression of certain genes and share a positive feedback architecture. The general behavior of such a system as well as the influence of noise on the distribution of cells between the off and on state are characterized experimentally [14] [15], but the transition of single cells between the off and on state has only been predicted theoretically [16]. Thus, the influence of noise on the timing of the transition remains unclear. This information is necessary in order to establish a framework that allows predicting the behavior of gene networks from their architecture. This predictive power will in turn greatly facilitate the construction of artificial networks.

The thesis is organized as follows: An introduction to gene regulation, noise and further basic concepts is given in chapter 2. Fundamental concepts, specific solutions developed in this work and advanced considerations on quantitative time-lapse fluorescence microscopy are presented in Chapter 3. Detailed experimental protocols can be found in Appendix B.

In the arabinose system the influence of gene expression noise on the time evolution of the switching between the off and on state is analyzed (Chapter 5). To this end, expression kinetics of the native system, as well as modified networks are measured. Heterogeneous timing, a temporal stochastic effect, is observed. A stochastic model indicates that this effect is due to a broad distribution of arabinose uptake proteins. A rate equation based model is developed which consistently explains the gene expression dynamics measured for different network architectures and different time courses of the inducer concentration. Furthermore, it is analyzed whether cell-to-cell variations in the timing of the response are beneficial under conditions where inducer availability is limited. GFP maturation has to

be accounted for when modeling dynamic gene expression data. Therefore, the maturation time was measured in single *E.coli* cells (Chapter 4).

In contrast to the arabinose system, which requires transport proteins to take up the added sugar, the *Pseudomonas putida* *PPU* system is regulated by membrane diffusible molecules, which are produced by the bacteria themselves. The *PPU* system is a model for the formation of biofilms, which are layers of surface attached bacteria that are surrounded by a protective matrix. It is tested whether gene expression dynamics are influenced by inducer accumulation in the vicinity of the cells by comparing results (i) under undisturbed conditions in which the molecules can accumulate and (ii) under fixed environmental conditions achieved by a constant flow of medium. Furthermore, the stochastic variability that arises while single cells grow into microcolonies is addressed.

2 Basic concepts

2.1 Gene expression and transcriptional regulation

The production of a protein by a cell, which is called gene expression, comprises two steps: The transcription of a gene into mRNA and its subsequent translation into a protein (see Figure 2.1). Regulation of this process is possible at any stage. Thus, the protein content of the cell can be adjusted, when bacteria are faced with variations in the environmental conditions, like the increase in the concentration of a certain nutrient or the sudden presence of an antibiotic. Here, we focus on transcriptional regulation in bacteria which was studied

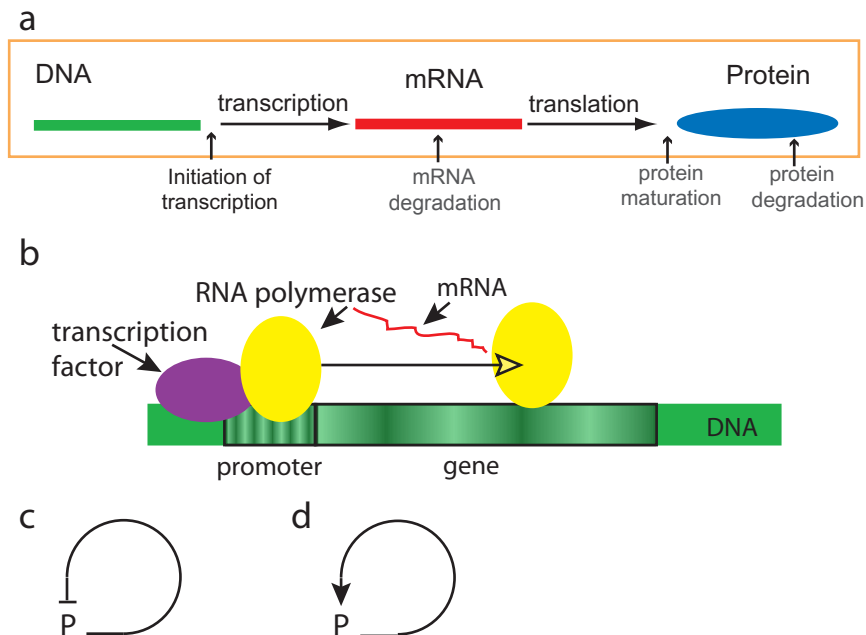


Figure 2.1: Gene Regulation (a) The production of a protein from a gene comprises transcription and translation. In addition to the transcription and translation rate the amount of protein present in a cell also depends on the processes indicated by vertical arrows. (b) Transcription factors facilitate or prevent binding of the RNA polymerase to the promoter. Once bound, the polymerase moves along the DNA and copies the DNA into mRNA. Schematic of negative (c) and positive autoregulation (d) of a protein P. In regulation schemes blunt ends indicate inhibition, while arrows indicate activation.

in this work. Genes are transcribed by the enzyme RNA polymerase, which initially binds to the promoter, a site upstream of the gene on the DNA (initiation of transcription). This association can be inhibited or facilitated by other proteins, so called transcription factors (TF), which occupy specific binding sites on the DNA. Transcription factors thus influence the overall rate at which mRNA molecules are produced, which is mainly determined by the frequency of initiation events. TFs often have a signal-sensing-domain, such that they can be activated or deactivated via ligand binding, phosphorylation or other modification. For example, the protein AraC, which regulates the arabinose system, only stimulates gene expression when bound to arabinose. In signal transduction mechanisms, the modification of a TF which is already present allows for a very fast response.

2.2 Gene regulatory networks

Transcription factors are proteins which can themselves be transcriptionally regulated and often they regulate more than one gene. Furthermore, there are additional regulatory influences, for example by other proteins, which degrade activating compounds. The entity of transcriptional regulatory interactions in one cell thus comprises a large network, which is characterized by a high degree of modularity, robustness and the use of recurring elements [4] [17] [18]. Two simple and frequently occurring modules are positive and negative autoregulation. A positively autoregulated gene stimulates its own expression, while a negatively autoregulated one inhibits its own expression (see Figure 2.1). Positive autoregulation, which is the central element of the regulatory networks studied in this thesis is known to increase fluctuations, resulting in significant variability of the protein level. Furthermore, it is the simplest implementation of a switch, when it is combined with a non-linear response function [19]: If a signal leads to the creation of a sufficient number of proteins it is amplified, resulting in a large response. For small signals there is no response at all.

2.3 Stochasticity in gene expression

The expression level of one protein can vary significantly between cells even though they are genetically identical. These variations stem from stochasticity in gene expression [5] [8] [20] [21]. The variability can be visualized using cells which express two different fluorescent proteins (see Figure 2.2 a): The different levels of the two fluorophores result in different colors of the cells when the two channels are overlaid.

Gene expression noise is generated by the interplay of bursting, time averaging and noise propagation ([21], see figure 2.2 c):

Most proteins are only present in few to 100 copies per cell [23]. These low numbers do not arise from a continuous, low production of protein molecules. Instead, from time to time a large number of proteins, a so called **burst**, is expressed.

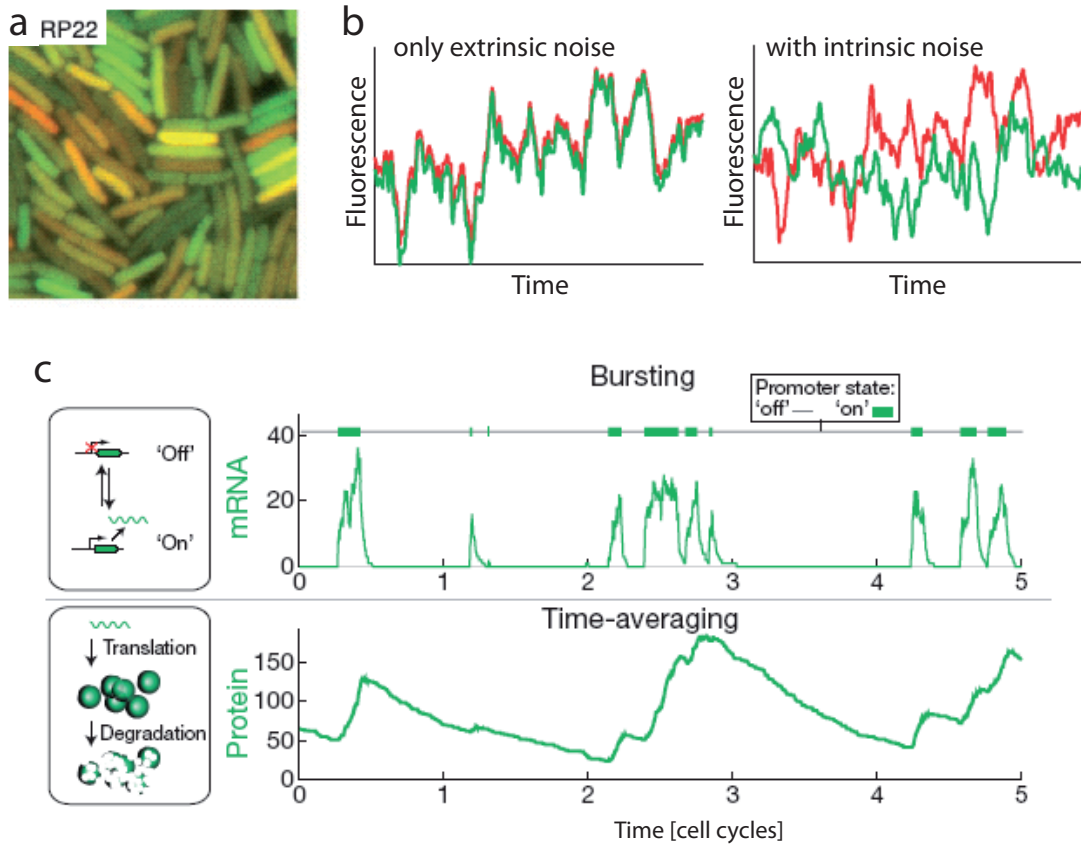


Figure 2.2: Noise in gene expression (a) Visualization of noise by the expression of two fluorescent proteins: The overlay of the two fluorescence channels shows that the expression level of the two fluorescent proteins differs between genetically identical bacteria. (b) Schematic illustration of intrinsic and extrinsic noise: If only extrinsic noise was present, the temporal fluctuations of the expression level of two fluorescent proteins controlled by two copies of the same promoter would be correlated. With intrinsic noise the variations become uncorrelated. (c) Random and transient derepression of a promoter gives rise to mRNA bursts. Fluctuations in protein concentration are less sharp, as the mRNA bursts are smeared out by the lifetime of proteins, which is typically longer than the time between bursts. (a and b are from [22]. Reprinted with permission from AAAS. c is from [21]. Reprinted by permission from Macmillian Publishers Ltd.)

The underlying reason is that **bursts** of mRNA molecules are produced [24] [25] due to the transient and random derepression of the promoter. Each mRNA molecule is amplified to many proteins as it is translated multiple times. As an example, consider a repressive transcription factor (TF), which unbinds from the DNA when it is associated with a certain signaling molecule. Still, even in the absence of the signal there is a small probability for the TF to fall off the DNA. Thus, occasionally mRNA molecules can be produced for a short time. With increasing numbers of signaling molecules the probability for the TF to unbind from the DNA and thus the frequency of derepression events increases, resulting in

increasingly continuous mRNA production. Noise thus mainly originates from transcription, while translation has only a minor influence. This was also shown in experiments, in which noise was found to be larger with low transcription and high translation rate than with high transcription and low translation rate [26]. The protein concentration at a given time results from **time averaging** over several expression bursts as the lifetime of proteins is usually longer than the time between bursts. Thus, the variability caused by mRNA bursts is partially averaged out by the protein lifetime. Finally, there is **propagation** of noise: The molecular machines responsible for gene expression, as well as the regulatory molecules are proteins themselves. Thus, their variations give rise to variations in the expression of other genes.

The overall stochasticity is frequently subsumed into two classes: Intrinsic noise, which originates from the expression of a given gene and extrinsic noise, which subsumes cell to cell variations in the numbers of eg. polymerases and other cellular components and affects all genes in a cell [27] [22]. The two types can be analyzed experimentally by expressing two different fluorescent proteins from two copies of one promoter [22] (See Figure 2.2 b). Fluctuations of the two proteins due to extrinsic noise are correlated, while the intrinsic fluctuations are independent.

Under certain conditions the variations are amplified, so that the cells are in entirely different physiological states. For example, only a fraction of the population becomes competent for DNA uptake in *Bacillus subtilis* [28] and in *E.coli* only a small fraction of bacteria is in a dormant state, which is resistant to antibiotics [29]. At certain sugar concentrations the genes of inducible catabolic networks are expressed at a high level, while the rest of the cells do not express the system at all. This so-called all-or-nothing gene expression [7] [6] will be discussed in detail in chapter 5.1, as it concerns the systems investigated in this thesis. Note that there is an important difference between the all-or-nothing behavior and the other effects: While in the other cases the cells differentiate into distinct states by stochastic switching, external inducer is necessary for transition into the highly expressing state.

2.4 Noise: Use or nuisance?

As more and more stochastic effects were discovered the question arose whether noise in gene expression is just an inevitable nuisance or useful for the cells [5] [21]. Even though certain network architectures can reduce noise, it cannot be shut out completely and theoretical work has shown that noise suppression by feedback systems has a lower boundary [30]. Increasing the number of molecules generically reduces the effect of noise, but the production of molecules is energetically costly. Still, proteins which fulfill essential cellular tasks show significantly lower variability than proteins which are only occasionally used [5].

For some cases the advantage conveyed to a bacterial population by stochastic differentiation is already qualitatively obvious: A small fraction of the population is committed to a costly task, or prepared to cope with rare or detrimental effects. This is best illustrated by the example of persister cells, which have a strongly decreased growth rate, but can resist antibiotics [29]. Quantitative considerations regarding the advantages of stochasticity can be done in the framework of game theory. In its context, each molecular implementation of a regulatory task including its noise characteristics is considered as a regulation strategy. As the regulation strategies we find in cells have obviously been favored by evolution it can be assumed that they are optimal under typical environmental conditions. To assess the suitability of a regulation strategy its costs and benefits under different environmental conditions are computed [11] [31]. To this end it is assumed that the quantity which is optimized is the growth rate of the population.

Noise generated phenotypic variations within a population represent so-called "mixed strategies", which combine pure strategies in a probabilistic manner. From game theory it is known that in many cases these mixed strategies are optimal (eg [32]). For several systems cost-benefit analysis revealed that stochastic gene expression effects are optimized for certain kinds of environmental fluctuations: The analysis of a general model, capturing the essential features of stochastic regulation mechanisms, showed that population heterogeneity leads to an increased population growth rate in a broad range of environmental variations [12]. Spontaneous switching of a small fraction of a bacterial population to a slow growing persistent state, which can resist antibiotic treatment pays off when antibiotic stress is a rare event [33]. Yeast strains were engineered to stochastically switch between two phenotypic states with different rates. Each state conferred a growth advantage under a specific environmental composition. As predicted, slow switching was beneficial in slowly fluctuating environments, while fast switching allowed for faster growth under rapidly changing environmental conditions [13].

Thus it seems that noise is exploited for particular purposes, while it has been minimized in instances where fluctuations impede functions by evolutionary network shaping. Additional evidence in this direction comes from a study in which an artificial network was created, which can generate dynamics similar to the natural network, but is nonetheless functionally different due to significantly different noise characteristics [34].

2.5 Modeling gene expression

Mathematical models quantitatively connect observed phenomena and their causes. They enable us to predict the behavior of existing and new systems, which is a crucial precondition for any engineering effort. A huge challenge when developing models of cellular functionality is the complexity of living systems. For the treatment of gene expression governed by regulatory networks several modeling approaches exist, which include various degrees of molecular detail [35] [36]. Some of the models are able to grasp the inherent stochasticity, while others are purely deterministic. Very detailed models explicitly consider the biochemical details of all processes, such as binding reactions between transcription factors and promoters. In contrast, the most abstract models describe a gene regulatory network by interconnected objects which can only be in one of two states (either off or on).

When choosing between these modeling approaches, the size of the network and the question which is addressed need to be considered. For example, many molecular details are included in models of small regulatory units, while larger pathways are frequently described by more abstract models.

Usually, a direct quantitative coincidence of experimental data and modeling results is desired. To this end reliable values of the rates governing the involved processes are crucial. However, these are often lacking, as the rates are numerous and strongly dependent on specific experimental conditions. Hopefully, this obstacle can be overcome in the near future by advances in large scale screening technologies.

In this work, models are used to understand and describe the time evolution of gene expression governed by inducible regulation networks. For this purpose, rate equations and stochastic descriptions are employed.

Rate Equations

The time evolution of biochemical reactions is classically modeled by rate equations. Mean particle numbers and reaction rates are used in these deterministic equations. Thus, stochasticity is not considered and the mean trajectory of an ensemble of molecules is computed. As a basic example from gene expression, we consider the expression of a stable protein via transcription and translation: mRNA molecules R are produced at a constant transcription rate ν from one DNA molecule, and actively degraded with rate λ

$$\frac{dR}{dt} = \nu - \lambda \cdot R \quad (2.1)$$

Proteins P are translated at a constant rate μ

$$\frac{dP}{dt} = \mu \cdot R \quad (2.2)$$

In this simple case the time evolutions can be computed:

$$R(t) = \frac{\nu}{\lambda}(1 - a^{-\lambda t}) \quad (2.3)$$

$$P(t) = \frac{\mu \cdot \nu}{\lambda} \cdot t + \frac{\mu \cdot \nu}{\lambda^2} e^{-\lambda t} \quad (2.4)$$

However, as soon as regulatory processes are considered, the system becomes more complex. In this work transcriptional regulation is at the center of interest, meaning that the transcription rate ν is not constant, but depends on the intracellular concentration of transcription factors which in turn change with time. This dependence, called *gene regulation function*, can either be measured (see for example [37] [38] [39] [40]) or predicted, for example from thermodynamic models of transcription factor binding [41]. Often, regulation is mediated by the cooperative binding of several molecules of a transcription factor. These mechanisms give rise to Hill-type regulation functions, where the dependence of ν on the transcription factor concentration A can be approximated by

$$\nu(A) = \nu_{max} \cdot \frac{A^n}{K^n + A^n}. \quad (2.5)$$

Here, K is the ligand concentration at which transcription proceeds with half-maximal rate and $n > 1$ gives the degree of cooperativity. This non-linear term has important consequences for the dynamics of the regulation network: In combination with feedback within the system it can give rise to complex characteristics such as multiple stable states or oscillations [19] [42].

In many cases the equation system has no analytical solution. Still, steady state levels can often be computed. The time evolution can be determined or even fitted to data by numerical integration. Furthermore, stability or bifurcation analysis can reveal important characteristics.

When formulating rate equations, one assumes that the reaction environment is well-mixed and homogeneous. Still, natural systems are often influenced by effects such as compartmentalization or diffusion. To account for these aspects the equations can be appended by additional terms. Compartments can also be treated by defining different states, which indicate in which compartment a molecule is.

Modeling stochastic events

On a fundamental level, noise in gene expression and its effects can be addressed by probability theory [43]. Using this methodology, the variability of mRNA and protein levels and the shape of their distributions over a population can be determined. The origin of noise and its propagation can be understood from fundamental statistical principles. In this work, we find the signature of a statistically derived protein distribution in our data (see Chapter 5.2.4 [44] [45]).

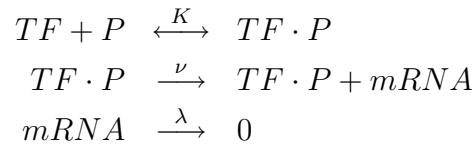
Rate equations are deterministic, but can be appended by noise terms, resulting in a

heuristic description of stochastic processes. The resulting formulation is called Langevin equation [46]. Another approach is the chemical master equation [46], which describes the probability for the system to be in a discrete state by the sum of entry and exit probabilities into and from this state. Apart from some very simple and small systems, like the birth-death process of one species, the master-equations is too complicated to be solved.

Usually, the Langevin and master equation are analyzed using Monte-Carlo methods. Each run of a stochastic simulation yields one possible realization of the systems dynamics under given initial conditions. Thus, one obtains an approximation of the probability distribution of the outcome when the simulation is repeated many times. This distribution can in turn be compared to experimental data, which is for example done in Chapter 5.2.4. Even though computationally costly and time-consuming stochastic simulations are the only feasible approach in many instances.

Commonly, the simulations employ an algorithm developed by D. Gillespie [47]. For a reaction system with a given number of molecules of each species and given reaction rates the next reaction and the time interval until this reaction happens are determined by a random number generator. As the exact algorithm only accounts for mono- or bimolecular reactions all binding events have to be included explicitly. The probability for a reaction to be chosen linearly depends on the product of the number of molecules of the involved species and the reaction rate.

As a simple example, we write down the reaction system used to simulate the number of mRNA molecules for a simple case. Their production is controlled by a promoter P . Transcription proceeds at a rate ν when a single molecule of a transcription factor TF is bound to the promoter. $TF \cdot P$ denotes a promoter to which the transcription factor is bound. mRNA is degraded at a rate λ .



The first line consists of two reactions, association and dissociation of promoter and transcription factor. The actual reaction rates for the association k_{on} and k_{off} dissociation of transcription factor-promoter binding are related to the affinity K as $K = \frac{k_{off}}{k_{on}}$. For the choice of the association rate the diffusion-limited value, which is $k_{on} = 2nM^{-1}min^{-1}$ for a typical transcription factor in *E. coli* serves as a guideline [48].

2.6 Fluorescent proteins as reporters for gene expression

Fluorescent probes for microscopic detection are characterized by strong absorption in the ultraviolet and/or visible spectral range. In addition, the excited molecules relax predominantly via radiative decay. The absorption and emission spectra show a mirror-symmetry and the maximum of the emission spectrum is red-shifted compared to the maximum of the absorption spectrum. The red-shift is a consequence of energy dissipation via non-radiative relaxation processes. The fluorescent probes can thus be detected with high sensitivity, as the excitation wavelength can be blocked in the detection light path. In addition, they allow visualizing structures which cannot be resolved otherwise due to low contrast or sizes below the optical resolution limit. Staining a sample with a classical fluorescent probe can be challenging, particularly when targeting intracellular substructures. In contrast, when using fluorescent proteins (FPs [9] [49]) cell components with an attached fluorophore can be produced within cells. Originally, the green fluorescent protein was discovered in the jellyfish *Aequorea Victoria* and its gene was isolated. It turned out that this gene can be cloned and functional GFP can be expressed in many kinds of cells and organisms. Their viability is usually not influenced by small amounts of GFP, as it is non-toxic. Via genetic engineering the properties of GFP were optimized and a large tool box of differently colored variants is now available. In addition, one can choose between stable and unstable variants [50] and mutants exist for special applications, for example pH sensitive ones [51].

In the majority of applications FPs are used to visualize cellular structures. FPs can also be used as quantitative tools to measure gene expression levels. The fundamental prerequisite for this application is the linearity of the fluorescent response, which is given as long as there is no saturation of the excited states of the fluorescent molecules or processes such as quenching due to a high density of fluorescent molecules.

For the analysis of gene expression the gene encoding a FP is controlled by the gene regulatory element of interest. More precisely, this means that the FP gene is inserted next to a promoter by cloning. GFP is much easier to detect based on its fluorescence than reporters classically used to monitor gene expression, which often rely on enzymatic properties. In contrast to most of the classical methods, GFP allows to follow the time evolution of expression processes, as detection is fast and does not require fixation or destruction of cells. Furthermore, signals from single cells can be measured easily.

When FPs are used to study gene expression photobleaching, FP maturation and the high stability of most FPs have to be considered carefully in studies: Fluorescent molecules are destroyed permanently after a number of excitation-deexcitation cycles (**photobleaching**). Besides the fact that the molecule can no longer be detected, photobleaching can be accompanied by the release of reactive compounds, which might cause problems in biological applications. The fluorescence photobleaching of GFP is reported to be slow in comparison to other fluorophores [52]. This property, however, differs also between mutants and needs to be tested under the chosen experimental conditions. Directly upon

protein translation GFP is not fluorescent. The protein and particularly the chromophore has to undergo **maturation**, a series of chemical modifications, the slowest of them being an oxidation step [9]. Thus, GFP cannot be produced in the absence of molecular oxygen. The duration of the maturation process varies strongly between GFP variants and depends on the temperature. For wild type GFP the maturation time was estimated to be 4 hours for expression in bacteria at 22 °C [53]. For *GFPmut1*, *GFPmut2* and *GFPmut3* it is reported that fluorescence can be detected already 8 min after induction of expression in bacteria at 37 °C [54]. Another fast and efficiently maturing variant is *Venus YFP* (yellow fluorescent protein) [55]. In experiments where GFP is used to monitor the time evolution of gene expression this maturation imposes a natural limitation on the time resolution and has to be taken into account when evaluating the data. Most FPs are **highly stable**. Thus, their concentration is only decreased by dilution due to cell growth. While the rate of gene expression and changes in this rate are faithfully reported by the FP, the accumulation of the protein can preclude effects on longer time scales. For example, it was shown that a protein distribution which was found to be homogeneous with a stable FP appeared to be much more heterogeneous with a destabilized FP [56].

Fluorescence is detected with plate readers, flow cytometers or microscopes in FP based studies of gene expression. Fluorescence plate readers monitor the fluorescence level of a large number of parallel cultures, which are grown in wells of microtiter plates which are incubated within the device. Thus, a large number of conditions can be investigated in one experiment [39]. In a flow cytometer (FC) the distribution of fluorescence values over a large population (≈ 5000 cells per second) is determined. Cell-to-cell variations can thus be resolved. Finally, the time course of expression of single cells can be analyzed by time-lapse fluorescence microscopy, which is discussed in detail in chapter 3. These data contain temporal information which cannot be captured by FC. The drawback is the high experimental effort necessary to obtain reasonable statistics (≈ 200 cells per day).

By pushing the limits of microscopy even further, FPs can be used to detect single molecules in bacterial cells [57]. In addition to high sensitivity of the equipment and a low level of background fluorescence, only molecules which move slowly or not at all can be detected. This has for example been done for transcription factors [58], membrane proteins [59] or molecules which have been anchored artificially to the cell membrane [60].

3 Quantitative time-lapse fluorescence microscopy

Circuit dynamics and stochastic effects of inducible regulation networks have been characterized extensively on the single cell level by imaging or flow cytometry [14] [15]. Using fluorescent proteins as reporters, these techniques reveal the distribution of gene expression levels over a population at one point in time. However, they do not allow following the fate of an individual cell over time. Still, the analysis of single cell dynamics is crucial to fully understand the functioning of regulatory networks and particularly the influence of noise.

Single cell gene expression dynamics can be measured by combining time-lapse microscopy, quantitative image analysis and fluorescent proteins as reporters [10]. Among many other examples, gene regulation functions have been resolved on the single cell level [37] and the differentiation dynamics of the *Bacillus subtilis* competence circuit [61] [62] have been studied in this way.

For a statistical analysis, a large number of single cell traces is necessary. Thus, the parallel acquisition of many cells and automated data analysis is highly desirable. The first experimental prerequisite is a fully automated fluorescence microscope which is equipped with an environmental control system in order to keep the temperature constant in the course of the experiment (Chapter 3.1). In addition, the bacteria need to be immobilized on a surface and the direct sample environment must provide sufficient nutrients. In addition, the possibility to change the environmental conditions in the course of the experiment is desirable (Chapter 3.2). Time-lapse movies are acquired by taking snapshots of a set of bacteria in regular time intervals. Quantitative image analysis consists of the determination of location and size of a cell (segmentation) and of a fluorescence value for each cell in each image. In addition each cell is tracked through the movie, in order to assemble a fluorescence time trace (Chapter 3.3). Finally, it is discussed for which applications a single-copy rather than a multi-copy fluorescent reporter is suited (Chapter 3.4). Detailed experimental protocols can be found in Appendix B.

3.1 Fluorescence microscopy

Fluorescent proteins (see Chapter 2.6) were used in this study to monitor gene expression by fluorescence microscopy. The variant *GFPmut3* was employed in the majority of the experiments. Therefore, its excitation and emission spectra are shown in figure 3.1 a.

Figure 3.1 c illustrates the principle of fluorescence microscopy and depicts our experimental setup: Illumination is provided by a lamp with an emission spectrum ranging from ultraviolet to infrared. Specific excitation of the fluorophore is guaranteed by the emission filter, which is a narrow band-pass. After passing through the emission filter the light is directed through the objective onto the sample by a dichroic mirror. The emitted light is collected by the objective and passes through the dichroic mirror and the emission filter, which blocks the excitation wavelength.

We use a fully automated inverted microscope (Axiovert 200M, Zeiss, Oberkochen, Germany) equipped with a motorized stage (Prior Scientific, Cambridge, UK). All devices are controlled by Andor IQ software (Andor, Belfast, Northern Ireland). Fluorescence illumination is provided by a xenon-mercury lamp, which is connected to the microscope via a liquid light guide (X-cite120, EXFO, Quebec, Canada). Filters for detection of GFP fluorescence are: excitation: transmission 450-490 nm; dichroic: transmission above 495nm; emission: transmission 500-550 nm (filter set 38HE, Zeiss, Oberkochen, Germany; see figure 3.1 b). The microscope also allows for the acquisition of bright field images. As the output of the fluorescence lamp can vary in the course of its life time Focalcheck fluorescence microspheres (Invitrogen, Karlsruhe, Germany) are used to correct for output variations. The temperature in the sample environment is maintained at 30°C or 37°C using a custom built heating box.

Prolonged illumination can lead to bleaching of the fluorophores or other photodamage, which can strongly decrease the growth rate. Therefore, we use a highly sensitive EMCCD camera (iXon DV885, Andor, Belfast, Northern Ireland) for image acquisition. Typical exposure times for fluorescence images are 0.1 s to 0.5 s. Fluorescence illumination is shuttered and bright-field illumination switched off between exposures and an orange filter is used in the bright field light path.

Either a 100x plan-neofluar (NA 1.3) or 40x plan-neofluar (NA 0.75) objective (both Zeiss, Oberkochen, Germany) are used. With the 40x objective each field of view is larger, such that a larger number of bacteria can be monitored. However, single cells in colonies cannot be separated, such that only time traces of the fluorescence of entire colonies can be determined. This analysis is feasible for the long time analysis of quorum sensing gene expression. In all other experiments the 100x objective is used to determined single cell time traces.

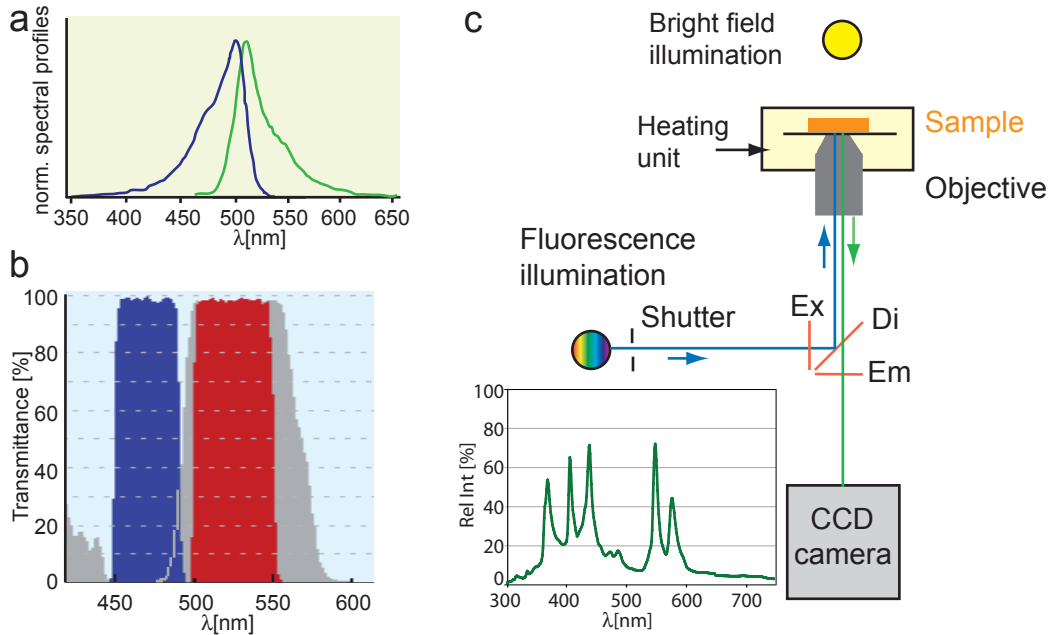


Figure 3.1: Fluorescence microscopy (a) Excitation (blue) and emission spectrum (green) of *GFPmut3* (adapted from [54]) (b) Transmission of the excitation (blue) and emission (red) filter and the dichroic mirror (grey) used for detection of *GFPmut3* in fluorescence microscopy (Filterset 38HE, Zeiss) (c) Fluorescence microscopy setup. The light path of fluorescent image formation is shown (blue and green lines). Ex: excitation filter, Di: Dichroic mirror, Em: Emission filter. The spectrum of the lamp used to provide fluorescence illumination is shown in the subgraph.

3.2 Sample environment and flow system

For the determination of fluorescence values via image analysis it is necessary to immobilize the bacteria on a surface. This guarantees that the entire cell is in focus and that the time evolution of one cell and its offspring can be followed over time. At the same time, adhesion may not interfere with cell growth. In many studies, gelled agarose is used for immobilization (e.g. [37] [61]). However, following sample preparation the chemical composition in the agarose gel cannot be changed. Microfluidic devices made from the polymer PDMS by using a structured silicon wafer as mold have been created to overcome this limitation [63]. They also allow analyzing the influence of limited or structured environments. However, it is a significant effort to manufacture and operate these devices.

In this study we established another sample environment, in which the environmental conditions can be changed in the course of the experiment: We use commercially available microscopy slides (μ -slideVI, Ibidi, Martinsried, see figure 3.2 a and b with several parallel, millimeter sized channels. Their surface is coated with Poly-L-Lysin (PLL), as *Escherichia coli* and *Pseudomonas putida* adhere to this coating. Still, their growth rate is not changed, compared to liquid cultures. The channels allow for manual rinsing and can also be combined with a flow system.

Sample preparation starts with filling a channel with 30 to 50 μl of the bacterial culture and incubating it for 5-15 min at the appropriate growth temperature. Subsequently, the channel is rinsed with medium to wash away non adherent cells and press the remaining cells to the surface. Following this procedure the vast majority of bacteria adhere with their long axis parallel to the surface. Induction is achieved by rinsing the sample with medium containing the appropriate inducer concentration. See Appendix B for detailed protocols.

We exploited the possibility to change the environmental conditions manually for the analysis of the GFP maturation time, which required the addition of an antibiotic which blocks translation. Furthermore, we applied inducer pulses, by flushing the channel with medium containing the inducer and after some time with pure medium.

As soon as nutrients are depleted or waste products accumulate significantly, the medium has to be exchanged. For example, *E.coli* cells can deplete a large amount of the sugar arabinose in the channel within 30-40 min. In other experiments it was observed that cells stop growing approximately 2 h after the experiment was started, probably due to nutrient depletion. Even though the medium can be exchanged by manually rinsing the channel, it is much more feasible to provide constant flow using a flow system. Constant flow provides the additional advantage of removing detached bacteria and detached parts of cells are pushed back to the surface. Furthermore, it can be used to keep the concentration of certain substances constant, which are degraded or produced by the bacteria themselves.

A simple flow system (figure 3.2 c), consisting of a syringe pump and connective tubing was already used in this work for the analysis of quorum sensing controlled gene expression. Here, the system served to rinse away signaling molecules and thus keep their concentration in the medium constant. The drawback of this system is that either the very beginning of the induction process cannot be monitored or the induction time is not very precise: Syringe and connected tubing are filled with medium containing the inducer. If the inducer is added to the channel prior to flow system connection the beginning of the process is missed. If the flow system is connected without previously adding inducer to the channel the starting time is ill defined, as low amounts of the inducer can diffuse into the channel before flow is started. As quorum sensing gene induction takes several hours, missing of the first few minutes can be tolerated.

Future experiments will require the temporal variation of inducer concentrations and monitoring of gene expression over many cell generations. To this end, the flow system will be appended by a second syringe pump and valves in the connective tubing (figure 3.2 d). Using this setup, the experiment can proceed under constant flow and the concentrations can be changed automatically by switching between the two reservoirs.

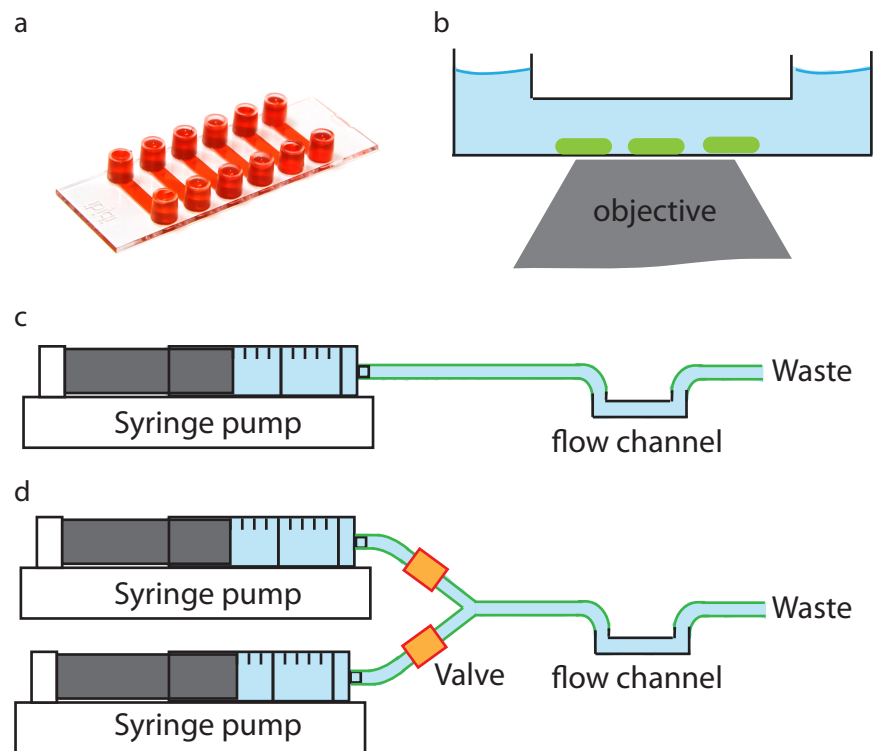


Figure 3.2: Flow channels and flow system (a) Commercially available microscopy slide with several millimeter sized channels (μ -slideVI, Ibidi). The bottom of the slide is 76x26 mm. (b) Bacteria (green) adhere to the Poly-L-Lysin coated channel surface and can be imaged. Medium can be exchanged by removing liquid from the reservoirs and adding fresh medium or via a connected flow system. (c) Our flow systems employs a syringe pump which is connected to the flow channel containing the sample via tubing. (d) In the future, the flow system will be extended by a second syringe pump in order to automatically switch from one concentration of a certain chemical to another one.

3.3 Image analysis of time-lapse movies

Bright field and fluorescence images of several fields in one sample were acquired regularly (eg. every 5 min), resulting in time lapse movies. Example image series are shown in figure 3.5 and 3.6. The evaluation of these movies includes the determination of fluorescence values for each cell in each image and the assembly of these values into time tracks (figure 3.3). ImageJ¹ is used for image analysis.

To measure the fluorescence value of a cell in one image an outline is determined on the bright field image and transferred to the back-ground corrected fluorescence image (Figure 3.3 a). For each cell either the total fluorescence (the sum over all pixel values within

¹Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, <http://rsb.info.nih.gov/ij/>, 1997-2009

the outline) or the mean fluorescence (total fluorescence divided by the pixel number) can be determined. The total fluorescence corresponds to the number of fluorescent molecules within a cell, while the mean fluorescence represents the concentration of fluorophores.

Background correction of the fluorescence images is done by subtracting the most frequent pixel value from all pixel values in each image. This is feasible as the background values lie within a small distribution and the number of background pixels is much larger than the number of pixels belonging to cells. As bleaching was found to be negligible no additional correction of the fluorescence values is necessary.

The simplest way to create the cell outline is thresholding of the bright field image, which was used in initial experiments. However, the separation of cells which are close together, particularly of daughter cells after division, is often not possible with this method.

The ImageJ PlugIn CellEvaluator [64], which was written and adapted to the specific needs of this project by S. Youssef (LMU), automatically creates outlines, tracks cells and measures the fluorescence and size of the cells. The creation of an outline starts wherever

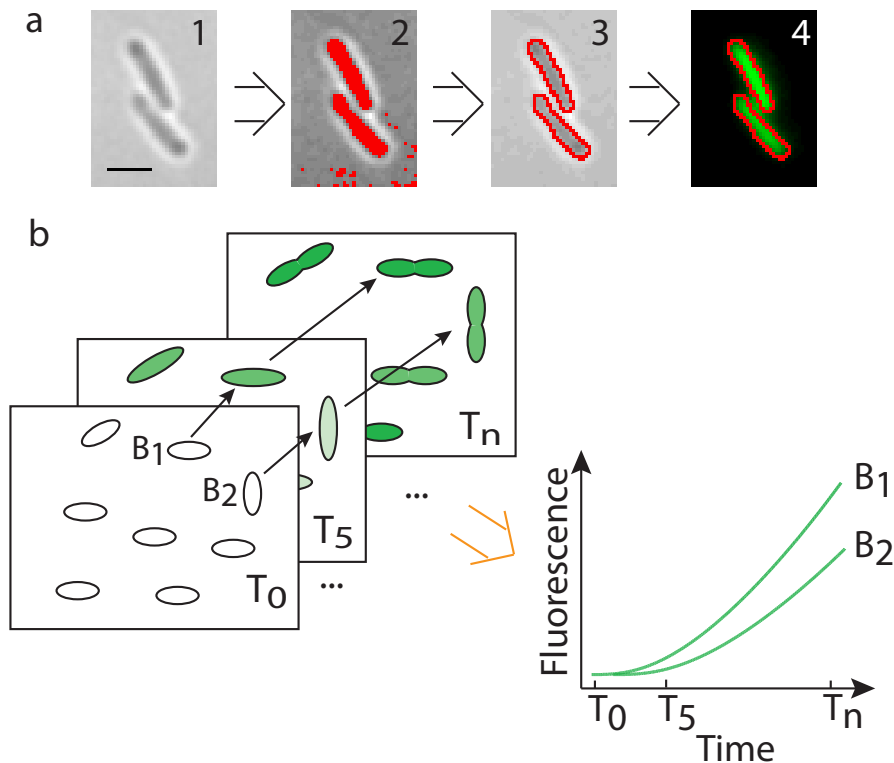


Figure 3.3: Illustration of image analysis (a) The bright field image (1) is used to create a cell outline. A threshold is applied to the image (2) and the rim of the cell marked (3). The outline is transferred to the fluorescence image (4) Note that, based on thresholding, advanced methods are employed to separate adjacent cells in automated image analysis. Scale bar size: $2\ \mu\text{m}$. (b) Fluorescence time series are assembled from the fluorescence values in single images by tracking cells through the movie.

a minimum number of pixels with values above a threshold are next to each other. Pixels are added on each border of these pixels, until either another object or the threshold, which indicates the background is met. For the separation of cells, the starting area for each cell is determined from the previous image in which the cells were apart. A weight matrix is used for image to image cell assignment (tracking), which includes among other parameters the position and outline of the cell. If a new track is started for each daughter cell after division, the program also displays a lineage tree, which includes a mother cell and all its offspring. Inspection by eye and manual selection of cells for evaluation is necessary, as a significant number of cells cannot be considered due to behavior such as partial detachment from the surface or piling up.

The bacteria grow and eventually divide in the course of the experiment. One can either start a new track for each daughter cell, or add the values of all daughter cells, eventually resulting in a "single colony" evaluation (figure 3.4).

In experiments with low magnification, cells within colonies cannot be separated, thus that the single colony evaluation is the only option. In this case, the single colony evalu-

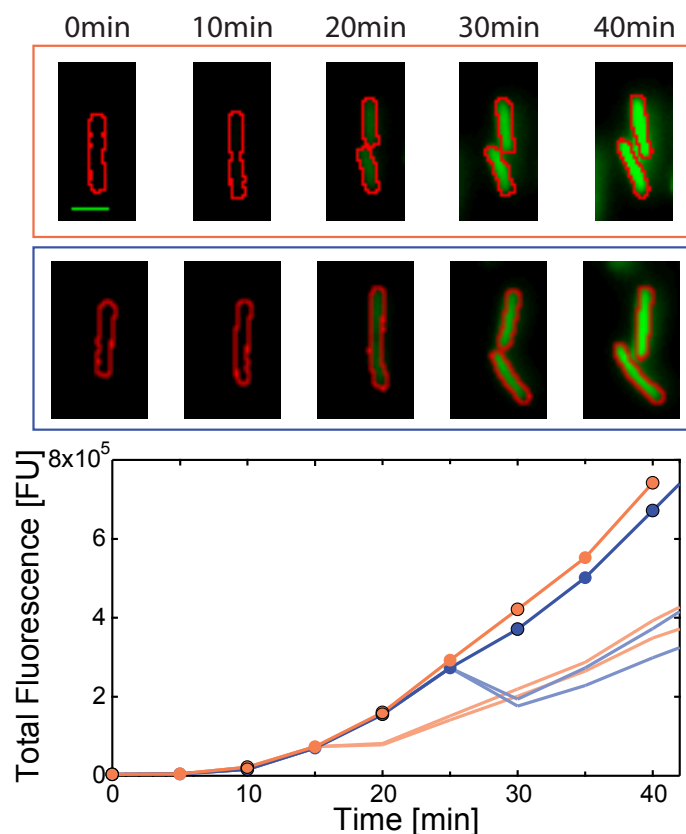


Figure 3.4: Image and fluorescence time series of *E. coli* cells containing the GFP gene under control of an arabinose inducible promoter which were induced with 0.2% arabinose at $t = 0$ min. Fluorescence traces without and with continuation after cell division are shown. Color of box corresponds to line in the graph. Time points which correspond to images are denoted with black circles. Scale bar size: $2 \mu\text{m}$

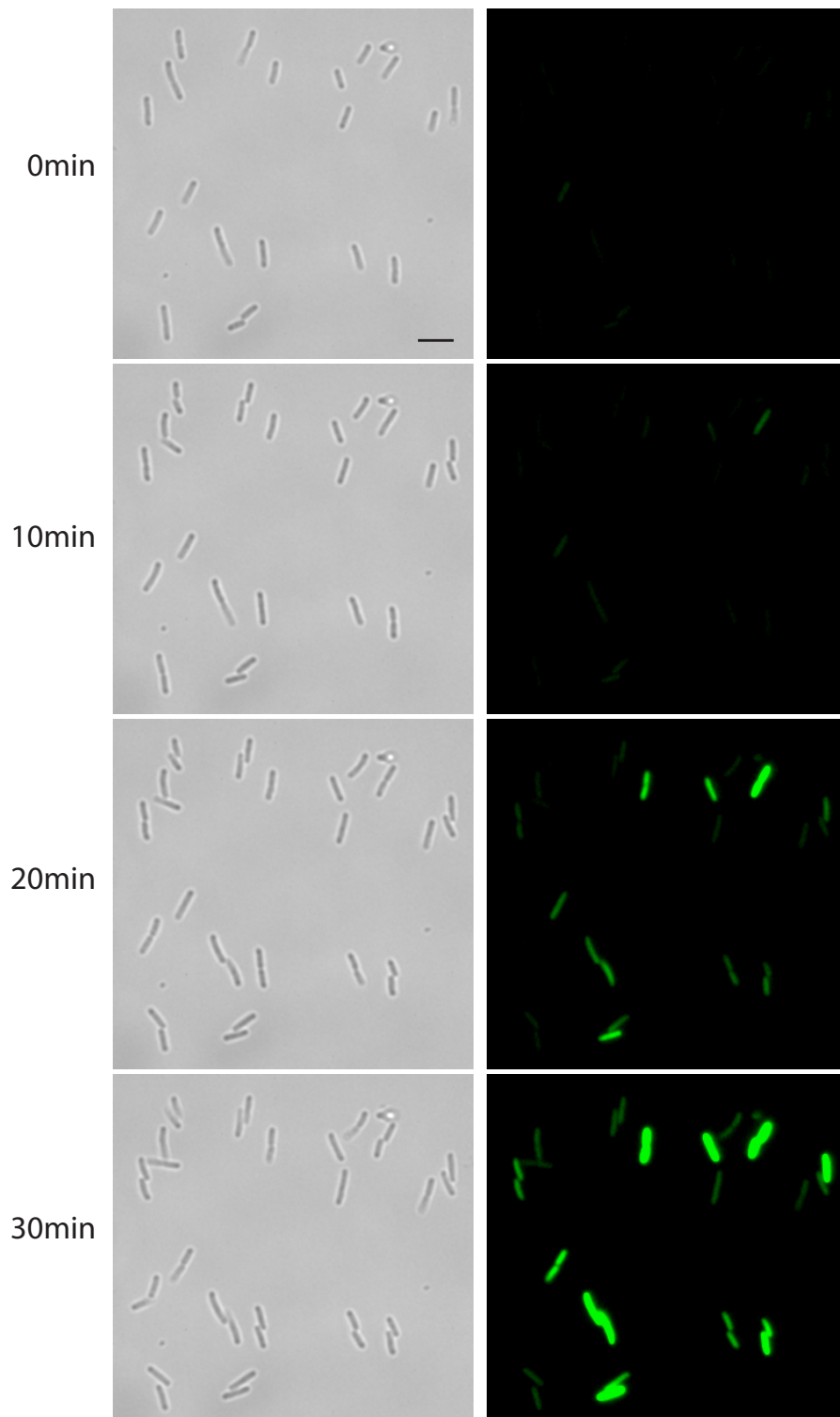


Figure 3.5: Time series of bright-field (*left column*) and fluorescence (*right column*) images. At $t=0$ min, 0.01% arabinose was added to *E.coli* cells containing the GFP gene under control of an arabinose inducible promoter. Scale bar size: $5\ \mu\text{m}$

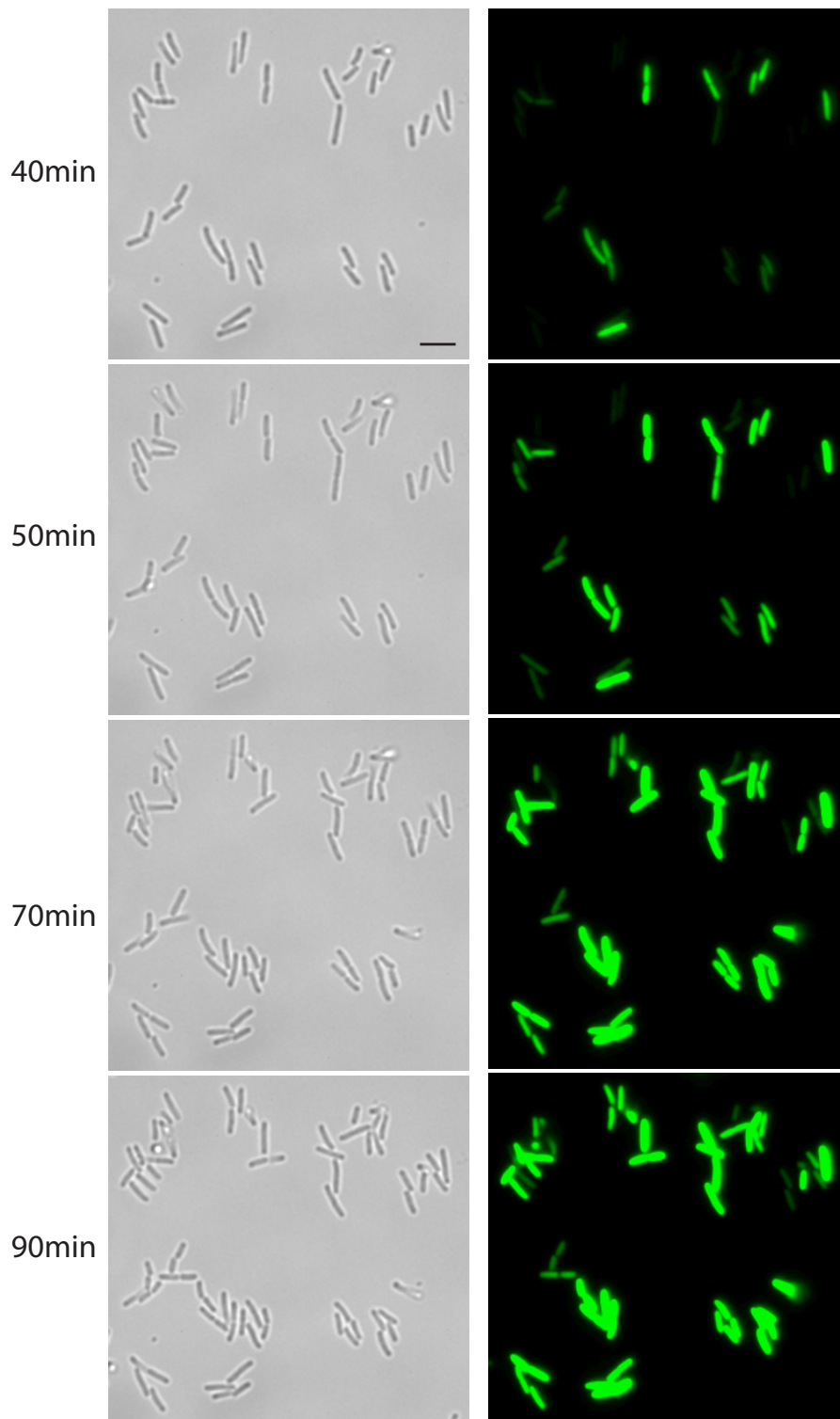


Figure 3.6: Time series of bright-field (*left column*) and fluorescence (*right column*) images (continued from 3.5). The contrast of the fluorescence images was changed after $t = 30$ min for better visualization of the fluorescence increase. Scale bar size: $5 \mu\text{m}$

ation is analogous to the single cell evaluation, with the one exception that an outline is created around a colony, not a cell.

If single cell evaluation is possible it depends on the question in focus whether to continue the track, or start new ones for daughter cells. For our analysis of timing in the arabinose system it turned out after the first set of experiments that continuation of the track is better suited, as the observed effects mainly depend on the initial state of the mother cell. Still, the results obtained with the two possibilities are very similar: For most cells, the effects of interest happen before the first division such that the curve is just continued slightly longer by adding the daughter cells. The fraction of cells for which the effects take place on a longer time scale (the experiments ends at the latest at the time of the second cell division) just change the distributions of the resulting parameters slightly.

Unfortunately, there is no calibration standard available which allows for the conversion of fluorescence signals to the number of fluorescent molecules. Thus, all fluorescence values are given in arbitrary fluorescence units (FU).

Two very neat, but indirect and time-consuming approaches for the calibration of the number of fluorophores are described in the literature: The partition of fluorophores between daughter cells was found to follow a binomial distribution. Thus, the relation that the mean value equals the standard deviation can be used for calibration [65]. Using an experimental setup with single-molecule resolution, a calibration curve can be generated by extrapolating the integrated fluorescence signal of a small, countable number of fluorophores [59].

3.4 Fluorescence signal and bacterial autofluorescence

The properties of fluorescent proteins and particularly the possibility to control the gene by any promoter of interest are summarized in Chapter 2.6. Here, we discuss the two ways in which a fluorescent reporter protein under control of a particular promoter can be present in a cell: It can reside on a plasmid, which is a short, circular DNA strand, or it can be incorporated on the bacterial chromosome. Plasmids are present in many copies per cell (typically 10 to 100), which leads to strong fluorescence signals. They can easily be engineered and introduced into bacteria. However, the increased copy number of the promoter can severely influence the regulation mechanism under investigation, for example by titration effects of transcription factor molecules [14]. This means that the regulation of each of the many promoters is different from the regulation of only one single promoter, as the number of transcription factors is too small. In addition, production of fluorescent proteins at a high rate can be a significant burden for a cell.

Integration of a reporter on the chromosome is more tedious than the engineering of a plasmid. The single copy of the reporter only gives rise to a small fluorescence signal, which can, at least partially, be compensated for by enhancing signal amplification. Still, in experiments with bacteria in which GFP was expressed from a single copy we found a

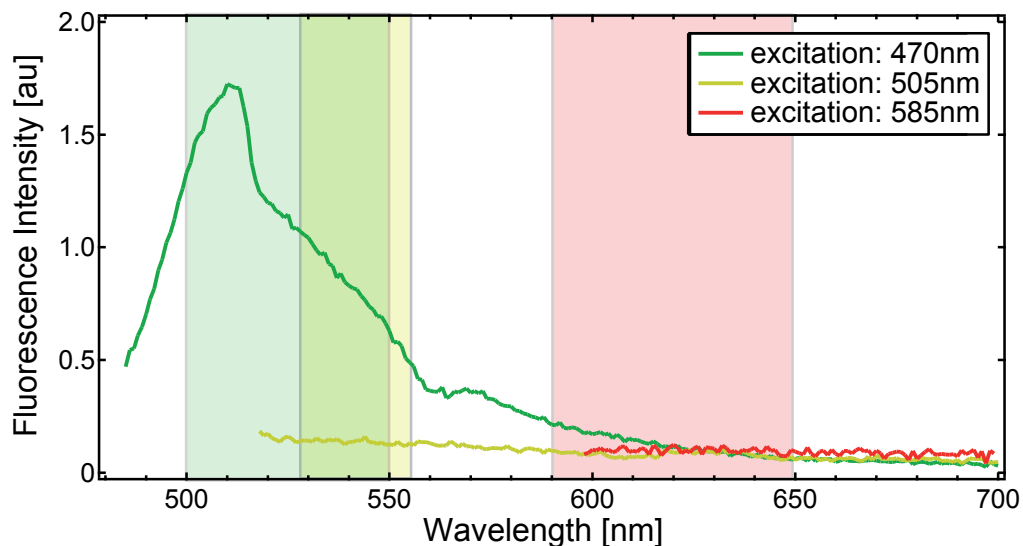


Figure 3.7: Autofluorescence emission spectra of bacteria at different excitation wavelength: Excitation at 470 nm corresponds to GFP (green), 505 nm to *Venus* YFP (yellow) and 585 nm to *mCherry* (red). Emission spectra were corrected for different lamp intensities at the excitation wavelength. To illustrate the significance of bacterial autofluorescence for each of the fluorescent proteins the transmission range (shaded regions, green GFP, yellow YFP, red *mCherry*) of typical filters used for microscopy are shown. The integral below the emission curve in the range of the filter is approximately $6 \cdot 10^5$ au (GFP), $4 \cdot 10^4$ au (YFP), $6 \cdot 10^4$ au (*mCherry*).

significant background signal, resulting from autofluorescence from each cell. If possible, it is desirable to circumvent the correction of this influence, as the autofluorescence differs significantly from cell to cell. Compared to the large signals resulting from a multi-copy reporter the autofluorescence is negligible. To check whether the autofluorescence level would be lower when using another fluorescent protein, fluorescence spectra of an *E. coli* strain containing no fluorescent reporter were recorded at wavelength corresponding to the excitation of GFP (green FP, maximal Ex/Em of *GFPmut3* at 501/511 nm), YFP (yellow FP, *Venus* maximal Ex/Em at 515/528 nm) and *mCherry* (red FP, maximal Ex/Em at 587/610 nm). The spectra show that the autofluorescence has a significant level in the emission range of GFP while it can be neglected for *Venus* YFP and *mCherry* (Figure 3.7). Thus, when a single copy reporter is necessary, it is advisable to use the latter ones. As their spectra are sufficiently distinct, they can also be combined to simultaneously monitor gene expression from two promoters. The most feasible and least disturbing way of chromosomal integration is placing an additional copy of the promoter which controls the fluorescent protein on the chromosome [10].

In this study, we use reporter plasmids as all regulatory interactions take place within isolated systems, meaning that there is no significant influence of global regulatory molecules. In all investigated cases, only one regulatory molecule is employed. To compensate for the additional promoter copies each of them is accompanied by a copy of the regulator. For the arabinose system we explicitly showed that our results are not different with a single copy reporter (Chapter 5.2.5).

However, for future studies, which will for example address the crossregulation of two systems, the use of a single copy reporter is necessary in order to conserve the native behavior.

4 Determination of the GFP maturation time on the single cell level

As discussed in Chapter 2.6, fluorescent proteins (FPs) have to undergo maturation. Depending on the FP variant, the duration of this process ranges from several minutes up to several hours [9]. Thus, this time needs to be considered when GFP is used as a reporter for gene expression dynamics. For our analysis on the single cell level, the average maturation time is not sufficient, as we need to know whether there is a large cell-to-cell variation associated with the maturation process. With our microfluidic setup, we can directly probe this cell-to-cell variation experimentally, under the same conditions as in the induction experiments. We use the *gfp* reporter plasmid pBAD24/*gfp* in strain E.coli LMG194 to determine the maturation time of *GFPmut3* [54], which is used in this study. On the plasmid, GFP is controlled by an arabinose inducible promoter. First, we induce bacteria with 0.2% arabinose and then inhibit protein synthesis *in situ* by flushing the channel with the antibiotic chloramphenicol. The resulting fluorescence trajectories cease to increase about 15 min after the addition of the antibiotic, see Fig. 4.1 a for a few representative trajectories. The fluorescence increase after the addition of chloramphenicol, which blocks translation, reflects the maturation dynamics of the remaining, non-fluorescent GFPs. A similar approach has been used to determine single cell maturation times in yeast [66]. The distribution of time-constants τ_m of GFP maturation shown in Fig. 4.1 b was obtained from exponential fits to the single-cell timeseries. We find an average maturation time of $\tau_m = 6.5$ min and a standard deviation of 0.6 min, i.e. a cell-to-cell variation of only about 10%.

Our finding of a relatively small cell-to-cell variation suggests that the maturation process is largely independent of the internal state of the cell in *E. coli*. This appears plausible, given that the oxidation reaction, which is rate limiting for the maturation process, does not depend on intracellular components [9]. For comparison, considerably longer maturation times of ~ 40 minutes were found for YFP and CFP in yeast [66], but only a slightly larger relative cell-to-cell variation (15 – 20 %). Moreover, from *in vitro* measurements of various YFP variants, oxidation timescales as low as 2-8 minutes were determined [55], indicating that the rapid maturation time detected in our experiment is conceivable *in vivo*.

These data have been published in [45].

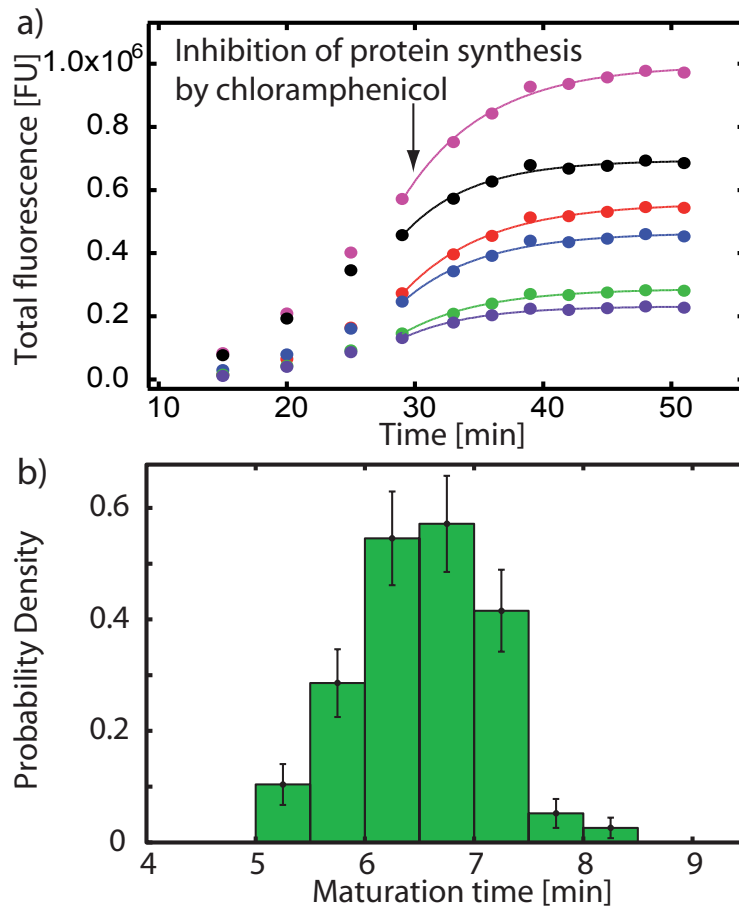


Figure 4.1: GFP maturation kinetics in single cells In (a) GFP expression was induced with 0.2% arabinose at $t=0$ min and protein synthesis was inhibited by addition of 200 $\mu\text{g}/\text{ml}$ chloramphenicol at $t=30$ min, as indicated by the arrow. Exponential fits to the fluorescent timeseries (*solid lines*) yield the maturation-time distribution in (b). The statistics was obtained from 77 cells.

5 Timing and dynamics of gene expression in the arabinose system

Sophisticated gene regulatory networks allow bacteria to optimize their proliferation in a wide range of environmental conditions. Inducible catabolic networks, which are activated when particular sugars become available, have been studied as model systems for a long time as they are fairly simple, even though they fulfill a crucial task. Their decisive regulation element is a positive feedback loop, in which expression of the sugar uptake proteins is enhanced by the sugar which in turn increases the sugar concentration. Positive feedback conveys the potential for multistability [19], which means that under certain conditions multiple distinct phenotypes can exist in parallel. In several inducible catabolic systems two states have been observed: One in which the genes are expressed at a high level (on state), while in the other one the genes are not expressed at all (off state).

Here the dynamics of the switching from the off to the on state in the arabinose system (Chapter 5.1) is resolved on the single cell level by time-lapse fluorescence microscopy. We find that the time at which expression of the operon starts varies significantly between the cells, a temporal stochastic effect which we denote as heterogeneous timing (Chapter 5.2). Using a mathematical model we conclude that this effect is causally related to a broad distribution of transporters at the time of arabinose addition. Subsequently, we analyze a modified network architecture in which the expression of the arabinose transporters is decoupled from arabinose (Chapter 5.3). From our model we expect to find heterogeneous timing in this case. However, we instead find scaling of the gene expression rate, indicating that the model is incomplete. As gene expression ceases rapidly in experiments in which arabinose is removed at a defined time after induction we include arabinose efflux in the model (Chapter 5.4). The extended model explains all experimental observations. Finally, we address the question whether heterogeneous timing might confer an advantage at high cell densities (Chapter 5.5).

5.1 The arabinose utilization system

The arabinose system [67] enables *E. coli* bacteria to take up and degrade the sugar arabinose. It consists of genes encoding proteins for arabinose uptake (AraE, AraFGH) and degradation (AraBAD), as well as the transcriptional regulator AraC (see fig 5.1). Each of these is controlled by a separate promoter. The transcriptional regulator AraC is an activator: When bound to arabinose it stimulates expression of the transporters AraE and AraFGH and the catabolic proteins AraBAD. At the promoter P_{BAD} AraC has an additional effect: In the absence of arabinose, two AraC molecules bind upstream of the promoter, thus creating a DNA loop which prevents gene expression. The expression of AraC is negatively regulated by itself with and without arabinose, resulting in a stable level of AraC molecules [68].

AraE, a proton symporter, is a low-affinity, high capacity transporter. In contrast, AraFGH is a high-affinity, low capacity system, which needs phosphate bond energy, eg from ATP hydrolysis for operation.

The system is also regulated by CRP (cyclic AMP receptor protein). This regulation strongly decreases the sensitivity to arabinose when a more favorable carbon source, such as glucose, is available. In this study the promoters are set to high sensitivity to arabinose

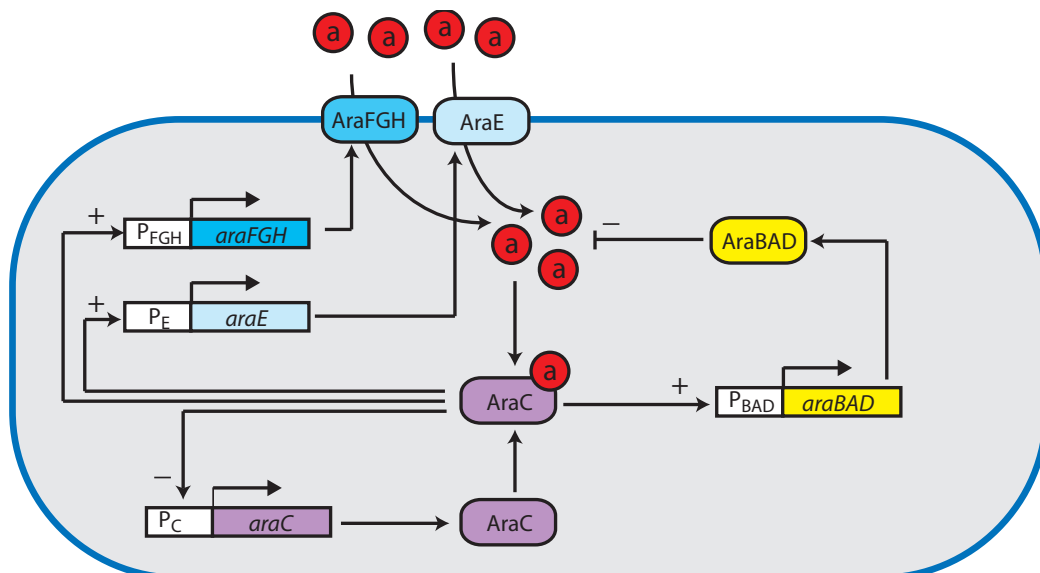


Figure 5.1: Regulatory network of the native arabinose utilization system [67] The system consists of proteins for arabinose uptake (AraE, AraFGH), arabinose metabolism (AraBAD) and the regulator AraC, along with their genes and respective promoters. When bound to arabinose AraC stimulates expression from the promoters P_{BAD} , P_E and P_{FGH} . In the absence of arabinose, AraC represses expression from P_{BAD} (not depicted). AraC negatively autoregulates itself in the absence and presence of arabinose [68]. Arabinose degradation causes a negative feedback on the internal arabinose level.

by using a minimal medium with glycerol as carbon source.

P_{BAD} promoter activity increases cubically with the intracellular arabinose concentration [69]. As the concentration usually increases fast, the cubic dependence mostly appears as a step function. This means that the promoter activity switches from zero to the maximal possible value once a threshold concentration is reached. The detailed dependence of P_E and P_{FGH} on internal arabinose concentration has not been determined, but seems to be very similar to P_{BAD} [68]. Arabinose exerts a positive feedback on itself, as the expression of the transporters AraE and AraFGH is stimulated by arabinose, which in turn leads to increased arabinose uptake. This positive feedback loop, in combination with the non-linear intracellular regulation function, is characteristic for inducible metabolic networks and can give rise to multiple steady states. Indeed, both the arabinose system and the lactose operon, which is the best studied metabolic system, show all-or-nothing gene expression [7] [6] (illustrated in figure 5.2). This means that at certain sugar concentrations a fraction of a population expresses the operon at a high level (on state), while the rest of the cells do not express the system at all (off state). In [6] the arabinose system expression from the P_{BAD} promoter was monitored, which is also done in this work.

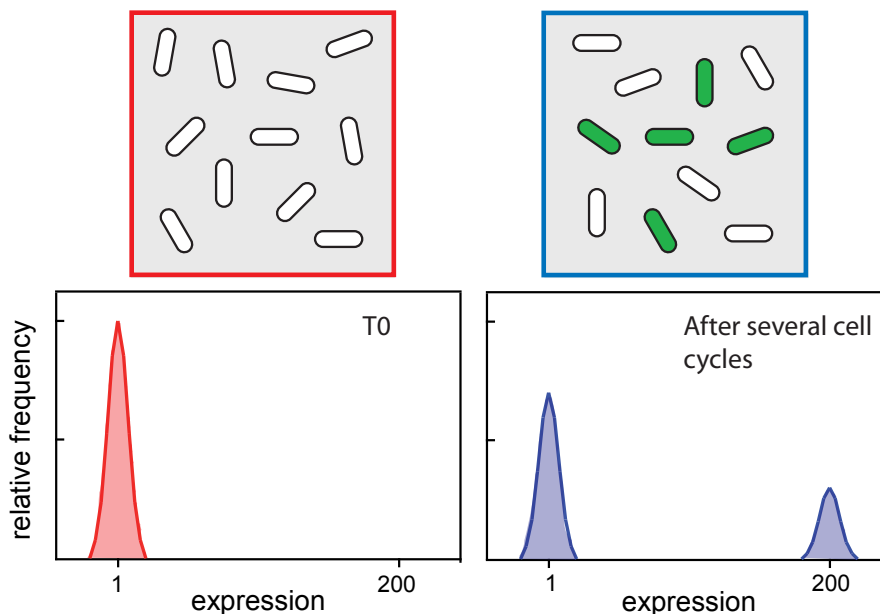


Figure 5.2: Schematic illustration of all-or-nothing gene expression (*upper row*) Gene expression of single cells, which could be observed microscopically when using fluorescent markers (*lower row*) Corresponding distributions of the gene expression level over a population of bacteria (*left*) At the time of inducer addition T_0 expression of a gene is off in all bacteria. The population distribution of the expression level is centered around a very low value due to occasional expression of the gene. (*right*) Several cell cycles after inducer addition some of the cells express the gene at a high level (green), while the rest does not express it at all (white). The distribution of the expression level over the population is thus bimodal.

The notion of how this differentiation comes about is as follows [7] [6]: Once exposed to the sugar, the bacteria accumulate it via the transporters which are present due to basal gene expression. Basal gene expression is the production of proteins at a very low rate in the absence of the inducer. Due to its stochastic nature the distribution of uptake proteins over the population is very broad. At low external sugar concentrations, cells with a large number of transporters can accumulate sufficient arabinose to cross the internal threshold concentration. Subsequently, the positive feedback leads to the production of a large number of transporters and the intracellular sugar concentration rises to a high level. In contrast, in a cell with a low number of transporters the arabinose concentration remains below the threshold as the sugar is diluted by cell growth. Thus, only a fraction of the cells becomes induced .

From this mechanism we expect to observe cell-to-cell variations in the time at which switching from off to on happens: Due to the broad distribution of uptake proteins, the rate of arabinose accumulation and thus the time until the threshold is overcome differs from cell to cell. This variation of timing has been found in a computational study of autocatalytic expression systems [16]. Experimentally, the transition between the off and on state has been characterized in detail by observing the time evolution of population distributions [14] [15], but single cell time courses have not been measured.

Care has to be taken when the benefit of all-or-nothing gene expression in natural settings is discussed: The all-or-nothing response was observed in strains incapable of arabinose degradation for the arabinose system and for induction with non- metabolizable inducers of the lac system. Induction with lactose results in transient bimodal distributions, but steady state distributions are always unimodal [14]. In addition, growth rate differences which probably exist between uninduced and induced cells can influence the distributions significantly.

Due to its tight repression, large induction fold-change and cheap inducer the P_{BAD} promoter is widely used for controlled gene expression in biotechnology [70]. However, for many applications modulation of the gene expression level within each cell is desirable. The arabinose system can be modified to yield homogeneous and regulatable gene expression by constitutively expressing the AraE transport system [71] [72].

5.2 Heterogeneous timing of single cell gene induction

Single cell gene expression dynamics following the addition of arabinose are measured by quantitative time-lapse fluorescence microscopy. These data are analyzed in order to understand the influence of stochasticity on the timing of gene expression. The theoretical models presented in this chapter were developed by G. Fritz (LMU). The results are published in [45]. Portions of [45] are reprinted with permission from Elsevier.

5.2.1 Single cell induction kinetics

To study the induction kinetics of the arabinose system, we use an *E. coli* strain in which *araBAD* and *araC* are deleted [70]. The chromosomal deletion of *araBAD* avoids the negative feedback of the internal arabinose catabolism. This feedback complicates the system, but is irrelevant for our questions, which focus on the kinetics of the induction when arabinose first becomes available externally. To monitor gene expression from the P_{BAD} promoter the bacterial strain is transformed with the reporter plasmid pBAD24-GFP, containing the *araC* gene and the gene encoding the rapidly maturing GFP variant *GFPmut3* [54] which is under the control of the P_{BAD} promoter. The *araC* gene is supplied on the

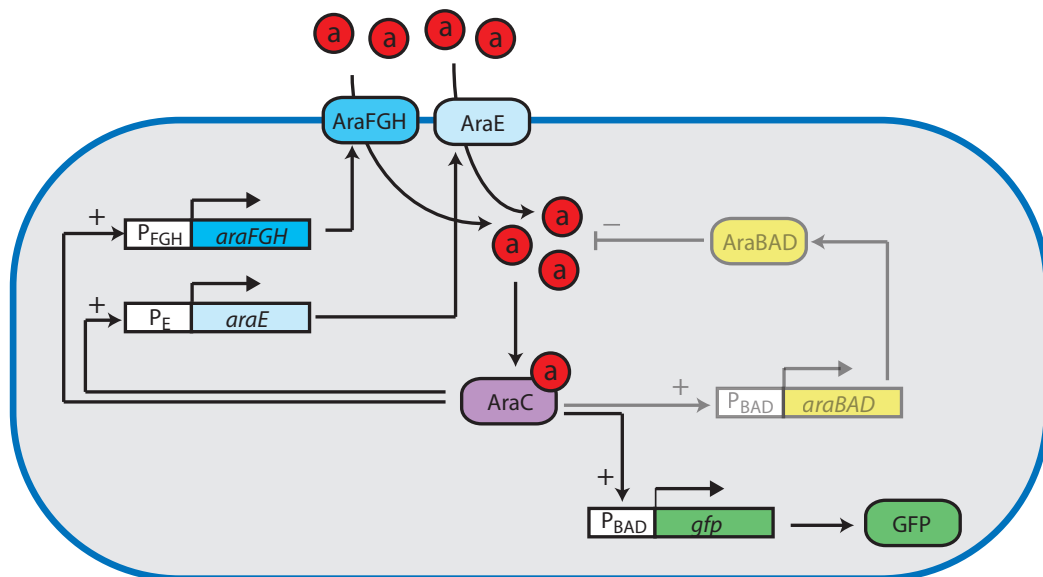


Figure 5.3: Regulatory network of the analyzed system The expression of the arabinose transporters AraE and AraFGH is stimulated by the regulator AraC when it is bound to arabinose. The permanent negative autoregulation of AraC, which keeps the regulator concentration constant is omitted in this scheme. As indicated by the light print, the degradation machinery AraBAD is deleted. As a reporter for the expression of the arabinose system we used a plasmid-borne *GFPmut3* under the control of the P_{BAD} promoter.

plasmid to guarantee full functionality of the DNA loop required for repression of P_{BAD} in the absence of arabinose [67] and to provide the proper stoichiometry of transcription factors and P_{BAD} promoters. The plasmid pBAD24 has an average copy number comparable to pUC [73], which is present in about 55 copies per cell [74]. The gene regulatory circuit of our system is illustrated in Fig. 5.3.

To perform the time-lapse fluorescence microscopy, we introduce the bacteria into a microfluidic chamber, where they attach to the Poly-L-Lysine coated chamber wall. At $t=0$ min, we induce the bacteria with 0.2% (13.3 mM), 0.05% (3.33 mM), 0.02% (1.33 mM) or 0.01% (0.66 mM) arabinose, and then record the time-evolution of GFP fluorescence in single cells. Representative fluorescence trajectories for the highest (0.2%) and the lowest (0.01%) arabinose concentration are shown in Fig. 5.4 a and b, respectively.

For all arabinose concentrations, the individual time-traces of each cell appear rather smooth and deterministic, whereas there is a significant variation in the response from cell to cell. We also observe a time lag between the addition of arabinose and the onset of fluorescence. With decreasing arabinose concentration, the typical lag time becomes longer, and its cell to cell variation becomes more pronounced.

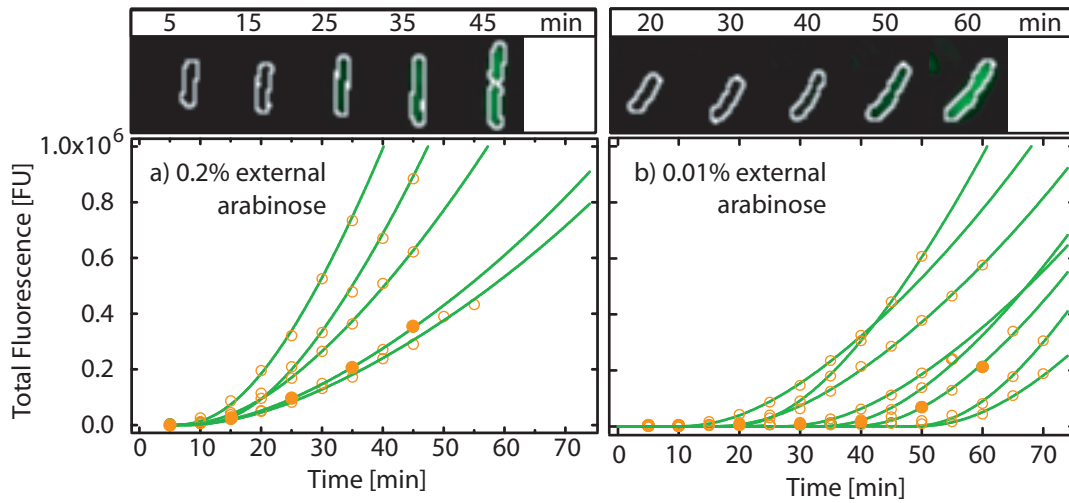


Figure 5.4: Example time traces Examples of single cell induction kinetics of the arabinose utilization network. Cells were induced at $t=0$ min with 0.2% arabinose (a) and 0.01% arabinose (b) (*empty orange circles*). The traces were analyzed up to the first cell division, which results in different numbers of data points in the traces. Fits of the deterministic gene expression function in Eq. 5.1 to the data are shown as green lines. The image panels in (a) and (b) correspond to the fluorescence traces marked with filled orange circles, respectively. The total fluorescence was determined as the sum of pixel values within the white outlines. These were created via thresholding of the respective bright field images.

5.2.2 Deterministic gene expression function

With the sudden increase of the external arabinose concentration at $t=0$ min, a cascade of biochemical processes is triggered, culminating in the fluorescent output signal measured in our experiment. In order to narrow down the origin of the stochasticity in the observed lag time, we need to analyze the individual steps in this cascade. For this analysis it is useful to separate the system into two distinct modules, an uptake module and a GFP expression module, as depicted in Fig. 5.5 a. The uptake module not only comprises arabinose import (represented here by an effective uptake protein Upt that subsumes transport by AraE and AraFGH, see Appendix A for details), but also includes the positive feedback of arabinose on the uptake protein. The expression module turns the production of the output signal on, when internal arabinose reaches the threshold level [69]. The delay time τ_D that is required to reach this threshold is solely determined by the uptake module. In other words, we assume that the transcription rate is switched from zero to a large value at the delay time τ_D . The feasibility of this assumption is indicated by the following analysis, which connects the data to the underlying biochemical processes.

The smooth shape of the time series suggests that the dynamics of individual cells follows a rather deterministic fate, while the differences between the cells stem from cell-to-cell variability of the reaction rates. Therefore, we set up a deterministic rate equation model that follows the scheme depicted in Fig. 5.5 b. We assume that the transcription

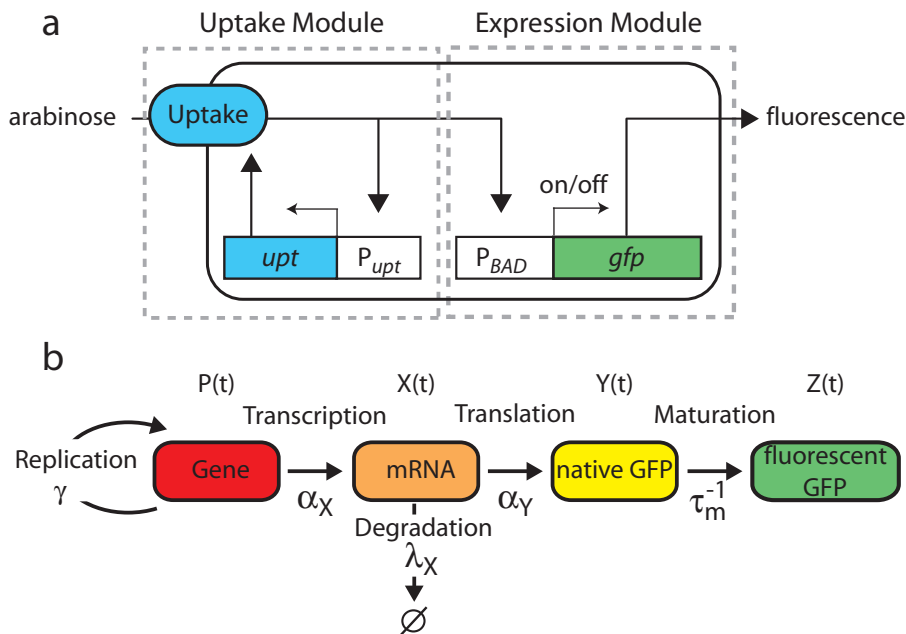


Figure 5.5: Functional system dissection and processes underlying the analytical gene expression function (a) The arabinose system is dissected into an arabinose uptake module (left) and a gene expression module (right). (b) The expression module comprises the depicted processes, which are modeled by rate equations.

rate from the promoter P_{BAD} is zero until the internal arabinose threshold for activation of P_{BAD} is reached at a time $t = \tau_D$. Then, the promoter activity jumps to its maximal value α_x . The corresponding rate-equations for the total abundance of plasmids (P), *gfp* mRNA (X), immature GFP protein (Y), and mature GFP protein (Z) per cell, are

$$\begin{aligned}\partial_t P &= \gamma P \\ \partial_t X &= \alpha_x P - \lambda_x X \\ \partial_t Y &= \alpha_y X - \tau_m^{-1} Y \\ \partial_t Z &= \tau_m^{-1} Y.\end{aligned}$$

with the cell-doubling rate γ and the rate for transcription α_x , translation α_y , maturation τ_m^{-1} , and mRNA degradation λ_x . Note that the model does not include dilution due to cell growth, since we measured the total fluorescence per cell in our experiments. Therefore the number of plasmids (number of gene copies) increases exponentially in time, keeping the number of genes per volume constant.

The time-evolution of the total number of fluorescent GFP molecules in a cell, $Z(t)$, can be computed by integration of the model equations and is given by

$$\begin{aligned}Z(\tau) &= \alpha_p \left(\frac{(\gamma + \lambda_x)e^{-\tau/\tau_m}}{(\gamma + \tau_m^{-1})(\lambda_x - \tau_m^{-1})} \right. \\ &\quad \left. + \frac{\tau_m^{-1}e^{-\lambda_x \tau}}{\lambda_x(\tau_m^{-1} - \lambda_x)} + \frac{\tau_m^{-1}e^{\gamma \tau}}{\gamma(\gamma + \tau_m^{-1})} \right) - Z_0,\end{aligned}\tag{5.1}$$

where $\tau = t - \tau_D$ is the time after transcription is switched on, $\alpha_p \equiv P_0 \alpha_x \alpha_y / (\gamma + \lambda_x)$ is a lumped constant giving the protein synthesis rate in fluorescence units per minute [FU/min], and Z_0 is a constant determined by the initial conditions ($P_0 = 55 \text{ cell}^{-1}$, $X_0 = Y_0 = Z_0 = 0 \text{ cell}^{-1}$).

τ_m	6.5 min	GFP Maturation time
λ_x	0.116 min^{-1}	mRNA degradation rate (corresponding half-life: 6 min)
γ	0.0139 min^{-1}	growth rate (corresponding doubling time: 50 min)

Table 5.1: Fixed model parameters When fitting the analytical gene expression function (equation 5.1) to the single cell traces these parameters are fixed to the given, population averaged values as their cell-to-cell variations can be assumed to be small.

5.2.3 Distribution of GFP expression rate and intrinsic delay time

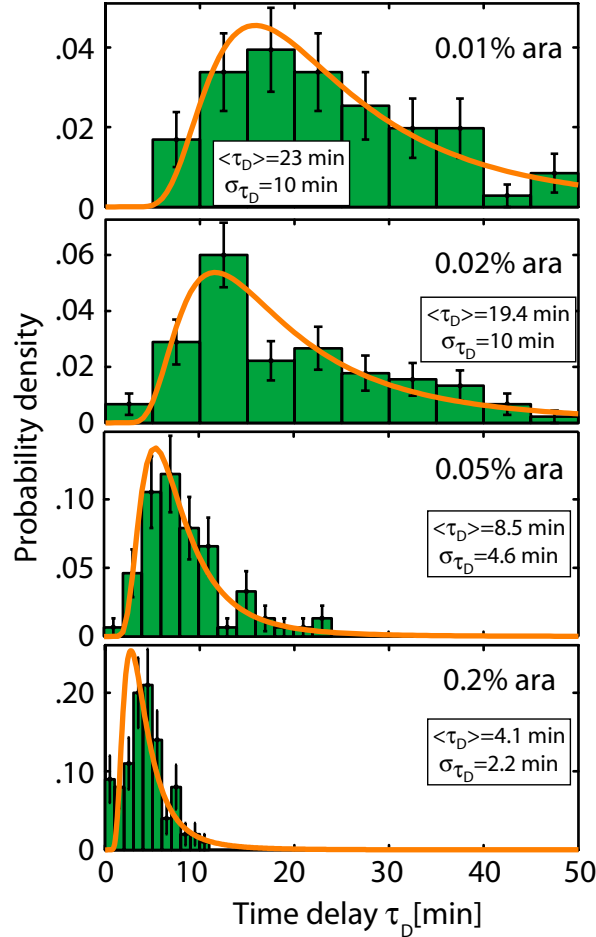


Figure 5.6: Histograms of the time delay for varying external arabinose concentrations, which were determined by fitting the analytical gene expression function to the fluorescence time series. The arabinose concentration along with the mean and standard deviation of the delay time are denoted in each panel. The experimental results (green) are well fitted by the analytical delay time distribution (equation 5.3, orange). The number of evaluated cells was: 101 (0.2%), 76 (0.05%), 90 (0.02%), 71 (0.01%).

A critical review of the literature indicates that the cell-to-cell variability of the mRNA half life can be assumed to be small. Analysis of the growth rates in our experiments and the separate analysis of the GFP maturation time on the single cell level (Chapter 4) show that this also holds for the growth rate and GFP maturation rate. In contrast, the protein synthesis rate is expected to vary significantly (see [45] for the detailed discussion). We thus fix the growth rate, the mRNA degradation rate and the GFP maturation time to their population averaged values when fitting the gene expression function to single cell fluorescence traces (see table 5.1). Thus, we can determine the protein synthesis rate

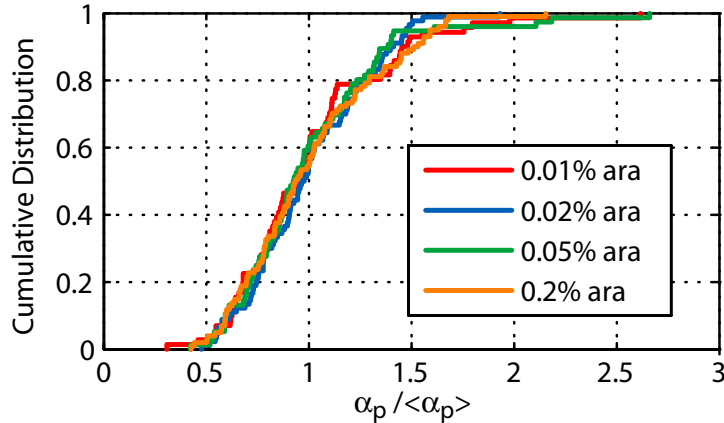


Figure 5.7: Cumulative distributions of the protein synthesis rate α_p coincide for all external arabinose concentrations when rescaled to their mean values $\langle\alpha_p\rangle$. Rescaling excludes sample-to-sample variations of the mean. Importantly, $\langle\alpha_p\rangle$ is not correlated with the inducer concentration.

α_p and the time delay τ_D for each cell, which can be assumed to capture most of the cell-to-cell variation within the expression module.

We fitted the time series of cells induced with various levels of arabinose (0.2%, 0.05%, 0.02%, and 0.01%). A few representative fitted curves for the highest and lowest concentration are plotted in Fig. 5.4 as solid lines. The resulting histograms for the delay time are shown in Fig. 5.6. For the lowest arabinose level (0.01%, *upper panel*) we find that the delay times are distributed between 5 and 50 min with a mean and standard deviation of $\langle\tau_D\rangle = 23$ min and $\sigma_{\tau_D} = 10$ min, respectively. In this case approximately 10% of the bacteria do not show any fluorescence within our time window of 70 min. With increasing arabinose concentration both the mean and the standard deviation of the delay time distribution decrease gradually, until at the highest arabinose level (0.2%, *lower panel*) a distribution with $\langle\tau_D\rangle = 4.1$ min and $\sigma_{\tau_D} = 2.2$ min is reached.

In contrast, we find that the distribution of protein synthesis rates α_p does not vary with the inducer concentration, which can be seen from the cumulative histograms which coincide for all concentrations 5.7. Pairwise Kolmogorov-Smirnov tests indicate that the samples are drawn from the same underlying distribution. Furthermore, a detailed correlation analysis shows that the delay time and synthesis rate do not depend on each other [45].

In summary, these results suggest that the uptake and the expression module are indeed functionally separate. It is likely that the level of crosstalk between the modules increases with decreasing arabinose concentration, indicated by a slight increase of the fluorescence level prior to the strong increase. Note that our experimental approach with time-lapse fluorescence microscopy was crucial for these results, which would have been impossible to obtain with flow cytometry.

5.2.4 Stochastic model for the uptake module

Using stochastic mathematical descriptions, we are now going to address the following questions:

- How do the broad distributions of delay times and their shapes arise?
- How does the scaling of the delay time with the arabinose concentration arise?

A Monte-Carlo simulation of arabinose uptake and arabinose transporter expression serves us to illustrate how cell-to-cell variations of the delay time arise. For details of the simulation see [45]. Figs. 5.8 b and c show the simulated time-evolution of the level of uptake proteins and the level of internal arabinose upon induction with 0.01% external arabinose for a few representative simulation runs. These trajectories illustrate the mechanism leading to a broad distribution of delay times within our model: Internal arabinose initially accumulates approximately linearly in time, and the accumulation accelerates only after reaching the effective arabinose threshold of $a_0 \approx 50 \mu M$ for activation of the *araBAD* and *upt* promoters, which is indicated by the black horizontal line in Fig. 5.8 c. The time delay, τ_D , caused by the uptake module is the time required for the internal arabinose concentration to reach this threshold level. The rate of arabinose import, given by the slope in Fig. 5.8 c, is proportional to the number of uptake proteins n in Fig. 5.8 b. If arabinose import is fast compared to the timescale of changes in the protein abundance, the delay time is given by the simple relation $\tau_D = a_0/(v_0 n)$, where the arabinose uptake rate per uptake protein, v_0 , depends on the external arabinose concentration. Thus, the distribution of uptake proteins in Fig. 5.8 a directly determines the distribution of import rates, which in turn are inversely proportional to the delay times, resulting in the distribution of delay times shown in Fig. 5.8 d.

In order to relate the experimentally observed shape of the distribution to the prediction of the stochastic model, we will now derive an analytical expression for the delay time distribution. Before the addition of the inducer arabinose, expression of the uptake proteins is a completely random, unregulated process. Following the work of Berg [44] and under the assumptions stated in [45], we find a steady-state distribution $P(n)$ for the number of uptake proteins n of the form

$$P(n) \approx \left(\frac{1}{1+b}\right)^\mu \left(\frac{b}{1+b}\right)^n \binom{\mu+n-1}{n}, \quad (5.2)$$

which is sometimes referred to as a ‘negative binomial’. Here, the ratio $b = \nu_p/\lambda_m$ of the translation rate and the mRNA degradation rate corresponds to the typical number of proteins produced from a single mRNA and is also known as the ‘burst size’ [75]. The ratio $\mu = \nu_m^0/\lambda_p$ of the basal transcription rate and the protein dilution rate can be interpreted as a dimensionless ‘burst frequency’ (the number of bursts within the lifetime of a protein). Both parameters determine the mean $\langle n \rangle = \mu b$ and the variance

$\delta n^2 = \langle n \rangle (1 + b)$ of $P(n)$. Fig. 5.8 a shows the steady-state distribution $P(n)$ obtained from our stochastic simulations of the uptake module (*grey histogram*) together with the analytical expression (*red line*) in Eq. 5.2 for the same rate constants. The excellent agreement suggests that the assumptions leading to Eq. 5.2 are all satisfied in the range of analyzed arabinose concentrations. As it can be shown that extrinsic noise only leads to rescaling of the parameter values [45] the following results, which are only based on intrinsic noise, are valid for the general case with intrinsic and extrinsic noise.

To obtain an approximation for the delay time distribution, we assume that arabinose uptake is rapid compared to the typical timescale of changes in the protein abundance. In this adiabatic limit, the delay time is inversely proportional to the current protein abundance in each cell, i.e. $\tau_D = \tau_0/n$, where $\tau_0 \equiv a_0/v_0$ is the time for a single uptake protein to accumulate arabinose to the threshold level a_0 . With this relation, the steady-state uptake protein distribution (Eq. 5.2) leads to a delay time distribution of the form

$$Q(\tau_D) \approx \frac{\tau_0}{\tau_D^2} \left(\frac{1}{1+b} \right)^\mu \left(\frac{b}{1+b} \right)^{\tau_0/\tau_D} \frac{\Gamma(\tau_0/\tau_D + \mu)}{\Gamma(\tau_0/\tau_D + 1)\Gamma(\mu)}, \quad (5.3)$$

where $\Gamma(x)$ is the Gamma function. In Fig. 5.8 d we compare this analytical prediction (*orange line*) to the stochastic simulation (*green bars*). The small deviation stems from the fact that the number of uptake proteins is not constant over the period of the time delay¹. The mean and variance of the delay time distribution can be approximated by

$$\begin{aligned} \langle \tau_D \rangle &\approx \frac{\tau_0}{\langle n \rangle} \left(1 + \frac{\delta n^2}{\langle n \rangle^2} \right) \approx \frac{\tau_0}{\mu b} \left(1 + \frac{1}{\mu} \right), \\ \delta \tau_D^2 &\approx \frac{\tau_0^2}{\langle n \rangle^2} \frac{\delta n^2}{\langle n \rangle^2} \approx \left(\frac{\tau_0}{\mu b} \right)^2 \frac{1}{\mu}, \end{aligned} \quad (5.4)$$

(see [45]). From these expressions it is clear that the model has two key parameters, which together determine the mean and width of the delay time distribution: the time required to reach the internal arabinose threshold by a single protein burst, τ_0/b , and the burst frequency μ .

Fits of the model to the experimental data by varying the two key parameters are in good agreement with the experimental distributions (Figure 5.6). The discrepancy at 0.2% is probably caused by the experimental procedure: The induction process takes place at room temperature, which probably causes a slight increase of the response times. Note that the two-parameter fit guarantees that the mean and standard deviation of the experimental and theoretical distribution will match. However, the fact that the *shapes*

¹Indeed, if the protein dynamics is much faster than the characteristic time of arabinose uptake ($\lambda_p^{-1} \ll \tau_D$), every cell experiences simply the average abundance of uptake protein $\langle n \rangle$ and the delay time distribution approaches a sharply peaked function around $\tau_D = \tau_0 \langle n \rangle^{-1}$. In our case, $\lambda_p^{-1} \approx 70$ min is much larger than the average delay times, so that the assumption of a constant n is sufficiently accurate.

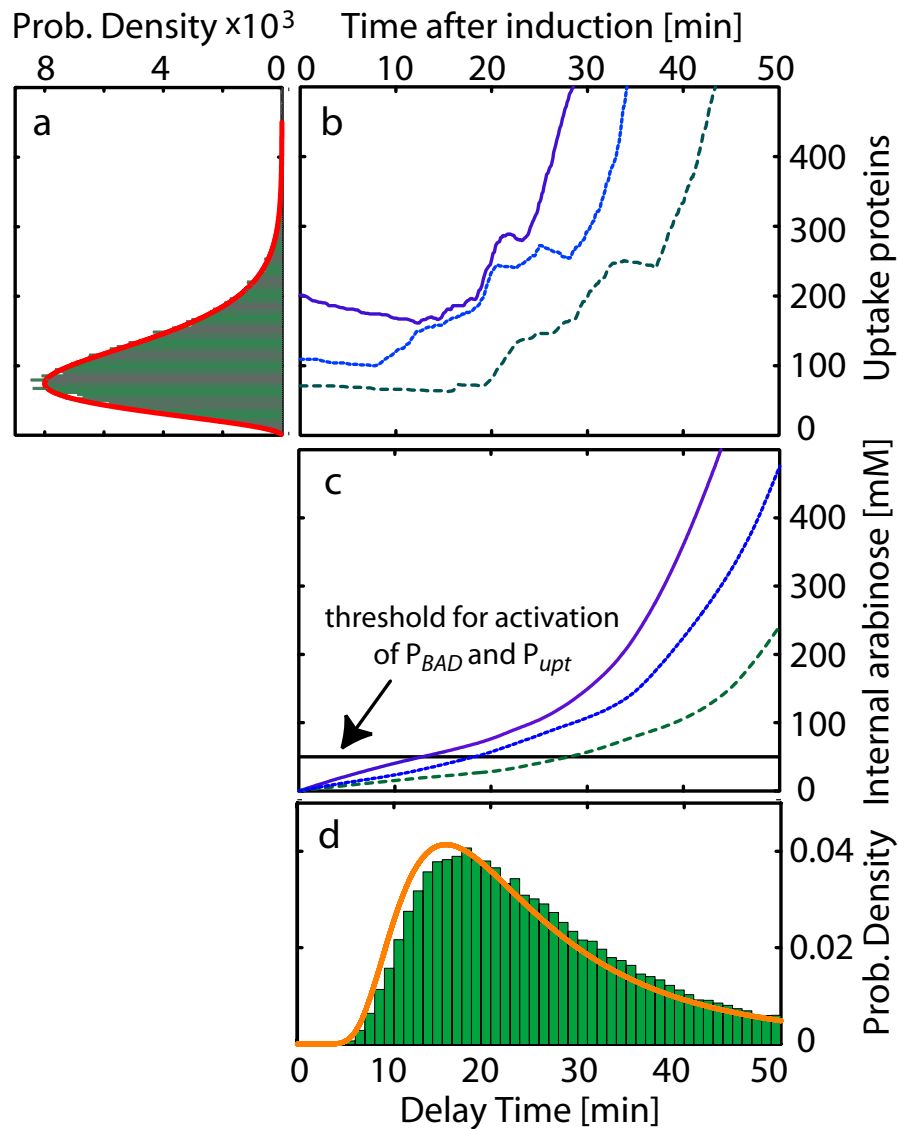


Figure 5.8: Simulation of the stochastic arabinose uptake process at 0.01% external arabinose. The simulated distribution of delay times (a, bars) is well approximated by the analytical prediction (a, red line, equation 5.2). Representative time-courses of arabinose uptake proteins (b) and internal arabinose (c) are shown. Once the internal threshold (c) is reached the positive feedback is activated, leading to a strong increase in transporter number and arabinose. The delay time distribution (d, bars) is obtained by measuring the time to reach the threshold. Despite a small deviation it is well described by equation 5.3 (orange line). The analytical predictions in (a) and (d) are shown for $\mu = 3.8$, $b = 30$, and $\tau_0 = 2100$ min.

of the distributions show excellent agreement is a nontrivial result, suggesting that the discussed delay mechanism can indeed explain our observations.

Finally, we address the consistency of the parameter values. The estimated parameters are functions of the external arabinose concentration: The timescale τ_0/b of arabinose accumulation in Figure 5.9 a decreases monotonically as a function of external arabinose and saturates for large sugar abundances, whereas the burst frequency μ in Figure 5.9 b is constant for all arabinose levels. This observation is consistent with the idea that the underlying protein distribution, characterized by μ and b , is independent of the externally provided sugar concentration, and that the differences in timing can be explained by shifts in the effective arabinose uptake velocity per uptake protein, v_0 : By assuming simple Michaelis-Menten saturation kinetics for v_0 , one expects that τ_0 scales inversely with the external arabinose concentration $[a_{ex}]$, i.e. $\tau_0 = a_0/v_{max}(1 + K_m/[a_{ex}])$, where v_{max} denotes the maximal uptake velocity per uptake protein and K_m the Michaelis-Menten constant. This behavior is indeed found in Fig. 5.9 a (*inset*) and with the resulting values for v_{max} , K_m and a typical value of $b = 30$ for the burst factor [75], all parameters are compatible with the experimentally constrained ranges [45].

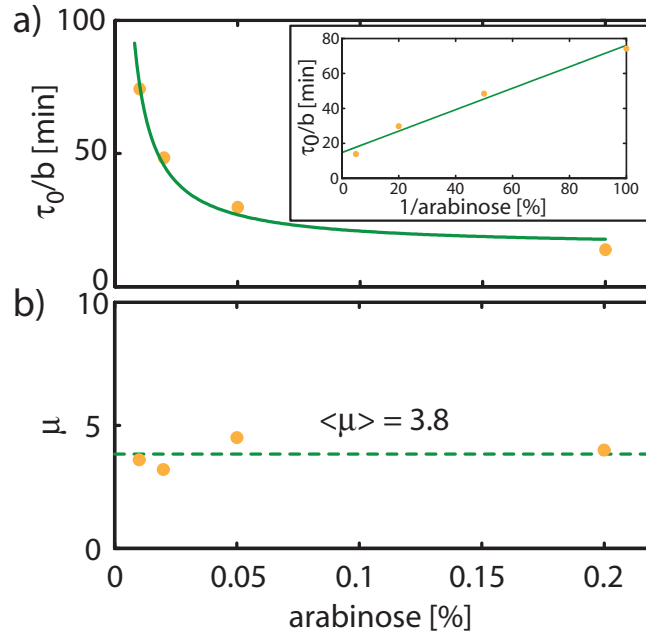


Figure 5.9: Time scale of arabinose accumulation and burst frequency as a function of external arabinose, were obtained from fits of the delay time distributions in Fig. 5.6. The timescale of arabinose accumulation τ_0/b (a) decreases monotonically with the external arabinose concentration. The Lineweaver-Burk plot (*inset*) shows the scaling with the inverse arabinose concentration. In contrast, the burst frequency μ (b) is constant for all arabinose levels.

5.2.5 Single copy reporter

It seems possible that the use of the multi-copy reporter plasmid has an impact on our results. For example, stochastic effects could be obscured by the high expression level and the averaging over many plasmids. Furthermore, the additional copies of the AraC gene and the AraC binding sites might interfere with the regulation of the system. To test whether the time delay characteristics are influenced by the use of the multi-copy reporter, the *GFPmut3* gene under control of the P_{BAD} promoter was integrated in the chromosome. An *E. coli* strain which is genetically similar to the previously used strain, but has an intact copy of the *araC* gene was used. Single cell expression kinetics of this strain (*E. coli* LKB194) were recorded at different arabinose concentrations. Prior to induction the bacteria already have a significant fluorescence signal. This cellular autofluorescence, which was negligible compared to the ≈ 50 times larger fluorescence signal generated by the plasmid, needs to be corrected for. To this end, an individual

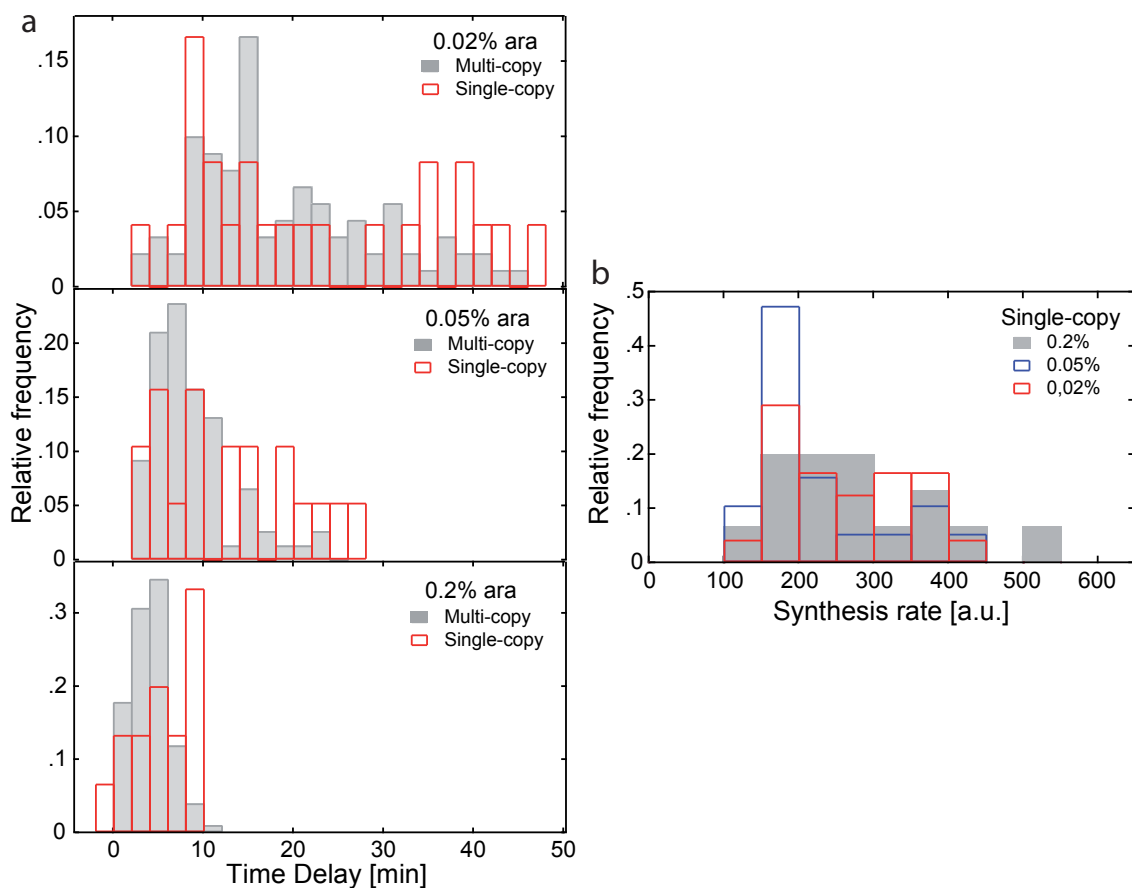


Figure 5.10: Time delay distributions measured with a single copy reporter (red), compared to the distributions measured with a multi-copy reporter at different inducer concentrations. To facilitate comparison by eye the data are differently binned than in figure 5.6.

offset value, usually 90-95% of the first acquired value, was subtracted from each single cell trace. The consistent determination of the offset for all cells is challenging due to the variations of the autofluorescence values between cells or for one cell prior to induction. The extracted delay time is influenced by the correction, resulting in an increased error of the determined parameters. In the case of the small delay times found for induction with 0.2% arabinose, we find negative delay times due to this error.

The time delays are distributed similarly with single and multi copy reporter (Figure 5.10 a, indicating that, given the experimental accuracy, the results do not depend on the copy number of the reporter. The distributions of the synthesis rate coincide for all concentrations (Figure 5.10 b). As in the multi-copy case, the relative width of the synthesis rate distribution is around 0.35. This level of variation was also found for the protein expression rate from the phage lambda P_R promoter and is probably caused by cell-to-cell variations in global cellular components, like RNA polymerase. Thus, differences in the plasmid copy number between cells do not seem to increase the overall noise significantly. Note that due to poor surface adhesion of this strain only a small number of traces could be acquired (15 (0.2%), 19 (0.05%), 24 (0.02%)).

5.2.6 Heterogeneous timing in a strain capable of arabinose degradation

We have observed heterogeneous timing in a strain which is incapable of arabinose degradation. Thus, the question remains whether this effect exists in native strains and is thus generally relevant. To address this issue, we recorded single cell expression kinetics in *E.coli* MG1655, in which the arabinose operon is native on the chromosome. To monitor gene expression the strain was transformed with the previously used reporter plasmid. MG1655 is the lab strain which is closest to the native *E.coli*, which means that it has only few mutations on its chromosome.

When adding arabinose at $t=0$ min and subsequently leaving the system undisturbed, as in the previous experiments, the fluorescence of single cells increases, but reaches a plateau after 30 to 40 min. Most likely, this is caused by the depletion of arabinose, which is taken up and degraded by the cells. When repeating the experiments under constant flow of medium with the respective arabinose concentration the effect vanished.

Again, we observe heterogeneous timing with increasing delay times at decreasing arabinose concentrations (see figure 5.11). For a given arabinose concentration the delay times are significantly longer than in the strain without arabinose degradation. For example, at 0.05% we find delay times of 8.5 ± 4.6 min without degradation and 23 ± 7.3 min. In addition, with degradation the shortest delay times are observed at a higher concentration (without degradation: 4.1 ± 2.2 min at 0.2%; with degradation 4.5 ± 2.0 min at 0.5%). Presumably, the differences are due to arabinose degradation by basally expressed degradation proteins, which slows down the initial increase of the internal arabinose concentration, resulting in longer delay times.

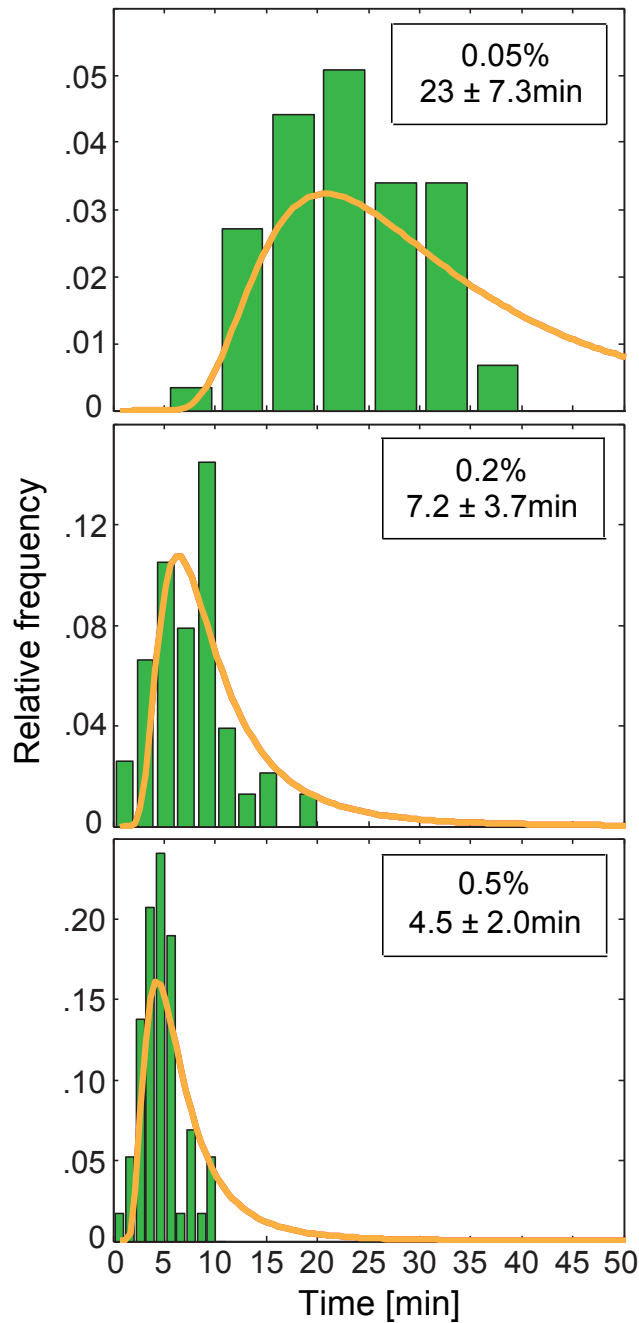


Figure 5.11: Time delay distributions measured in strain *E.coli* MG1655 following induction with 0.5 %, 0.2 % and 0.05 % arabinose. The arabinose network on the chromosome of MG1655 is intact, meaning that this strain is able to degrade arabinose. The arabinose concentration along with the mean and standard deviation of the delay time are denoted in each panel. For each experimental histogram (green) the corresponding analytical delay time distribution (Equation 5.3) is shown (orange). The number of evaluated cells was: 58 (0.5 %), 38 (0.2 %), 59 (0.05 %).

Even though the assumptions made when deriving the analytical delay time distribution (Equation 5.3) might not hold when arabinose is degraded, we use this functionality to test whether shape and scaling of the distribution is similar without and with arabinose degradation. The shape of the distributions can still be approximated when the burst frequency kept constant at $\mu = 3.8$ and a typical burst size of $b = 30$ is chosen. This constant burst frequency was determined in the experiments without degradation. As expected a Michaelis-Menten scaling of $\tau_0 = a_0/v_{max} (1 + K_m/[a_{ex}])$ is found. τ_0 again is the time to overcome the threshold of $a_0 = 50 \mu\text{M}$ by the action of a single uptake protein. The Michaelis constant is similar to the one without degradation, but v_{max} is significantly smaller. Together with the observation that the coincidence between the predicted curve and the data is worse compared to the data without degradation this indicates that the model needs to be modified to account for arabinose degradation.

5.3 Variation of the uptake protein expression

In the previous chapter we concluded that the distribution of delay times over a population is determined by a broad distribution of transport proteins at the time of induction (Figure 5.12). In this understanding, an individual cell accumulates arabinose via the n uptake proteins which are present at the time of arabinose addition until the internal binding threshold a_0 is reached. At this point, denoted as delay time τ_D , expression from the arabinose promoters starts with maximal rate. The delay time τ_D is thus given by

$$\tau_D = \frac{a_0}{v_0 \cdot n} = \frac{a_0}{v_{max} \cdot n} \left(1 + \frac{K_m}{A_{ext}}\right). \quad (5.5)$$

v_0 , the uptake velocity of each transporter shows a Michaelis-Menten type dependence on the external arabinose concentration A_{ext} , with maximal velocity v_{max} and Michaelis-Menten constant K_m . Based on this finding we expect to find similar time delay characteristics with an initially given transporter distribution, which is not arabinose dependent. This situation can be created by replacing the promoter which usually controls transporter production with another promoter. The positive feedback loop, which is a decisive characteristic of the network, is disrupted in this manner. An important prerequisite is that the number of transporters is similar to the native case. However, there are only very few promoters for which the number of expressed proteins can be predicted precisely. We chose to use the promoter P_{LAC} , which natively controls gene expression in the lactose operon: It was shown that its basal expression level is very low [59] and the expression from P_{LAC} can be increased by inducer addition. As the delay time is determined by the transporter number n and the external arabinose concentration A_{ext} , the expected differences in n will only change the scaling with A_{ext} .

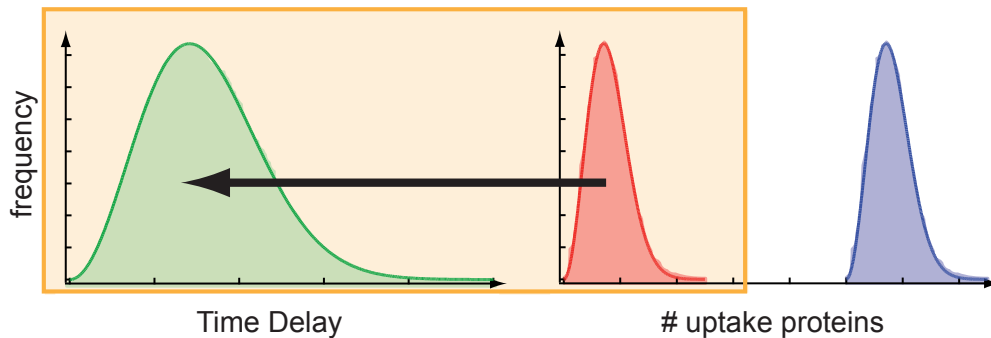


Figure 5.12: Schematic illustration of the connection between time delay and uptake protein distributions We assume that the initial distribution of uptake proteins (red) gives rise to the delay time distribution (green). Due to the positive feedback of arabinose on the uptake proteins in the native system the distribution is shifted to higher values at a later time (blue), which supposedly does not influence the delay distribution. Note that the delay distributions, but not the uptake protein distributions, have been measured.

5.3.1 Single cell induction kinetics

The induction kinetics of the strain in which the production of the low affinity high-capacity transporter AraE is controlled independent of arabinose by the lac promoter P_{LAC} were recorded for several arabinose concentrations (Figure 5.14). The basal expression level of P_{LAC} was used. For comparison, kinetics were recorded for a strain in which the native promoter p_E controls expression of AraE (Figure 5.14). The regulation networks for these strains are shown in Figure 5.13. As the transport characteristics of strains with only AraE and both AraE and AraFGH are very similar (see Appendix A) AraE was put under arabinose independent control and AraFGH was deleted. AraFGH is also deleted in the control strain with native AraE control.

We find that the arabinose concentration has to be decreased stronger with P_{LAC} controlled than with P_E controlled AraE expression to change the expression kinetics. In addition, already a qualitative comparison of the single cell traces shows that the dynamics of the two strains differ significantly. For a quantitative comparison the delay time and the protein synthesis rate are determined from each single cell trace by fitting the analytical gene expression function Eq. 5.1 derived in chapter 5.2.2. All remaining parameters are fixed to the previously discussed values. Note again that the central assumption in the derivation of this function is a step like increase of the transcription rate from zero to its maximal value at the delay time τ_D .

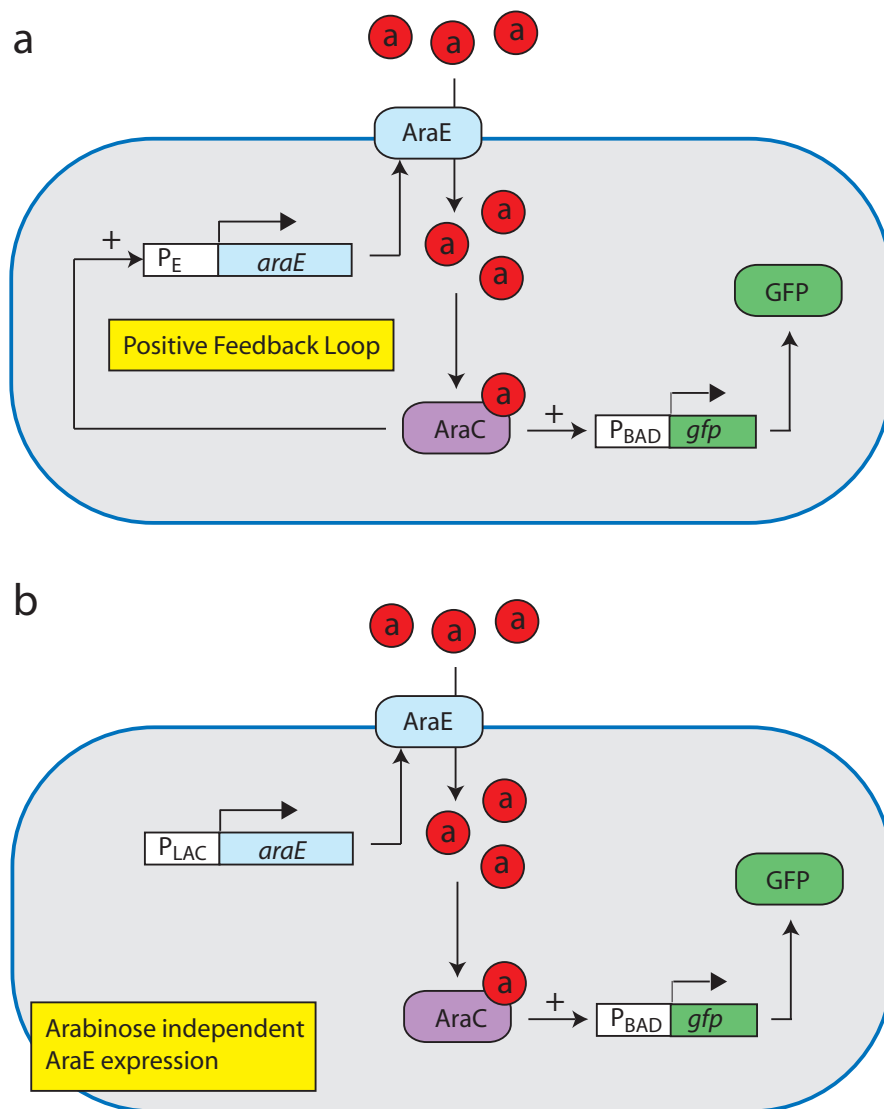


Figure 5.13: Network with native and P_{LAC} controlled AraE expression
 In the native strain (a) the positive feedback on the transporters is intact (*E. coli* JW1889-3), while under P_{LAC} control (b) AraE expression is arabinose independent (*E. coli* JW1889-5).

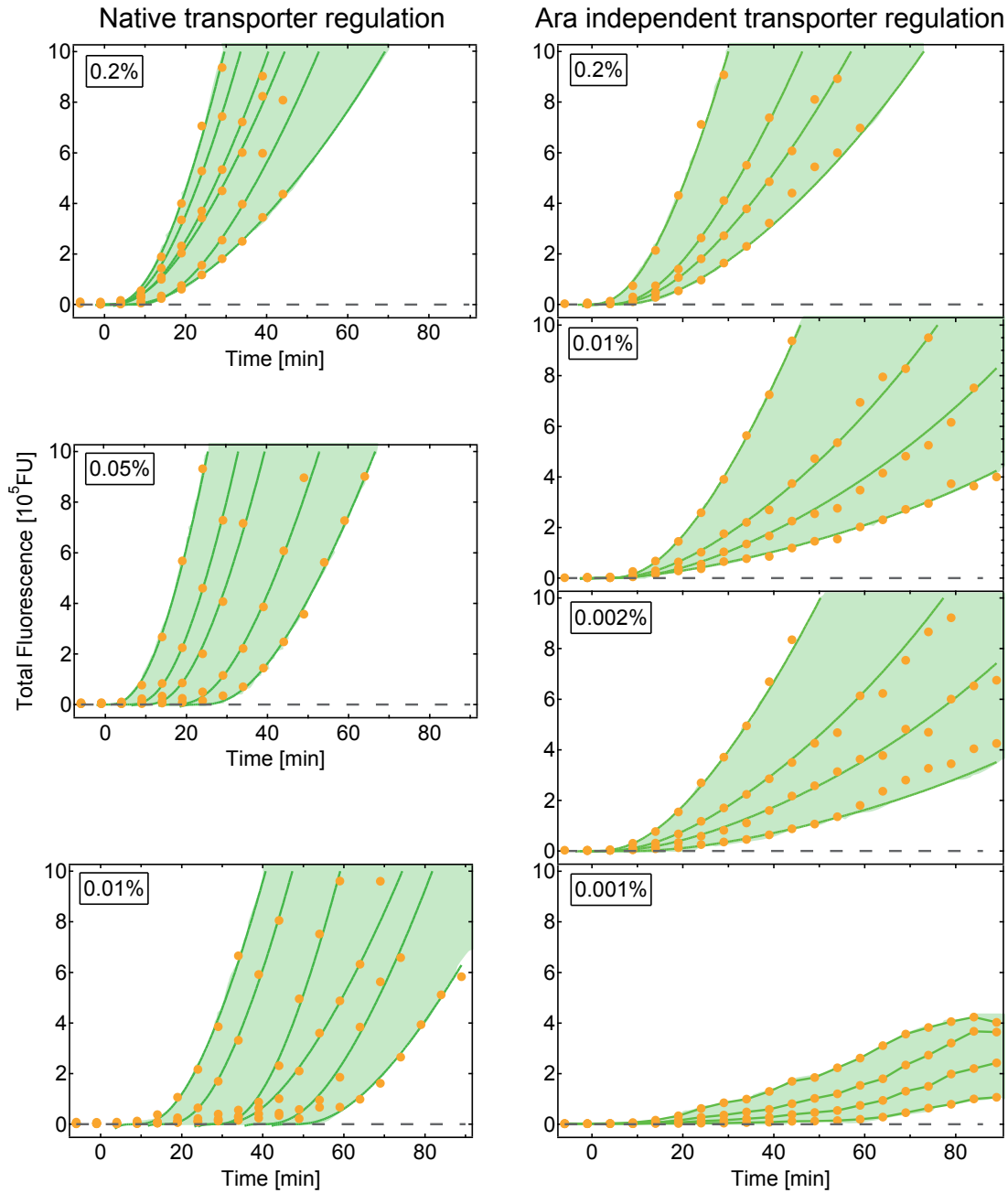


Figure 5.14: Single cell induction kinetics Representative single cell induction kinetics of cells with native (left) and arabinose independent (right) transporter production at different arabinose concentrations. The analytical gene expression function (green) is fitted to the data points (yellow dots). As the 0.001 % arabinose data cannot be fitted by this function data points are connected. To facilitate comparison the regions over which the traces are distributed are shaded. In contrast to prior evaluations, the traces were continued after division events by summing over the daughter cells. This approach now seems more feasible, as we previously found that the number of uptake proteins at the time of arabinose addition is decisive for the entire dynamics. Still, the effect of the change in the evaluation procedure on the results is only minor.

For the native control strain we find increasing delay times with decreasing arabinose concentration and similar synthesis rates for all arabinose concentrations (Figures 5.16 and 5.15), which was also observed with both transport systems (see chapter 5.2). For induction with 0.01% and 0.05% the mean values of the delay time coincide with the previously measured ones within experimental errors. Only the mean delay time for 0.2% is significantly shorter. This is probably due to a change in the experimental procedure: While we induced the bacteria prior to putting the sample on the microscope previously, inducer is now added when the sample is already on the microscope. This minimizes the time the sample remains at room temperature instead of $37^{\circ}C$. These very short times in combination with the simplifications underlying the evaluation function can result in negative values for few cells.

For P_{LAC} controlled AraE expression we find very small delay times at all concentrations and the synthesis rate decreases with decreasing arabinose concentration (Figures 5.16 and 5.15). Thus, neither the expected increase of delay time with decreasing arabinose concentration, nor the constant synthesis rates are found. The maximal synthesis rate for P_{LAC} controlled AraE expression is significantly smaller than in the native strain. The shape of the traces at 0.001% arabinose cannot be fitted with the analytical gene expression function.

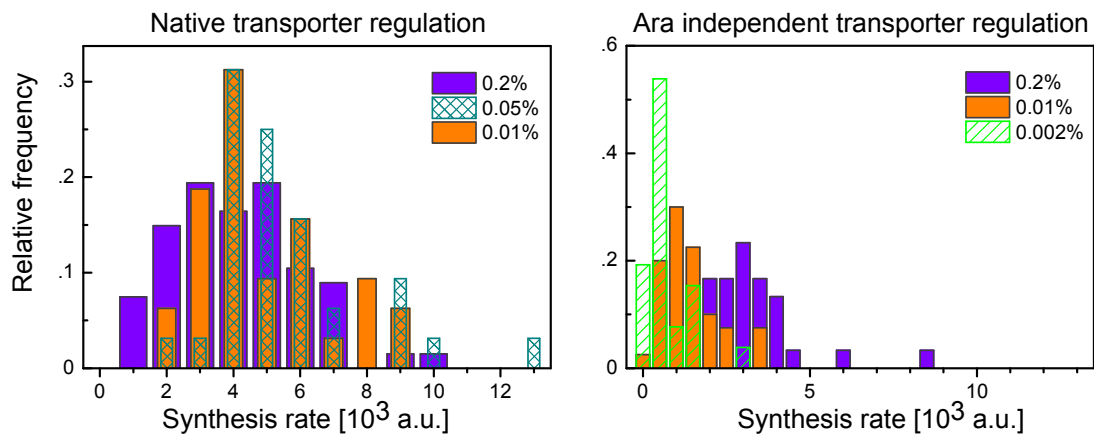


Figure 5.15: Synthesis rate distributions at different arabinose concentrations with native (left) and arabinose independent (right) transporter expression. Arabinose concentration along with mean and standard deviations of the values are denoted in each panel. For the number of cells see Figure 5.16.

Qualitatively the same behavior, fast gene expression onset in all experiments and decreasing synthesis rates with decreasing arabinose concentration, have also been observed in an *E. coli* strain in which the expression of AraE was controlled by Pcp18, another arabinose independent promoter (see traces in Appendix A). As transcription from Pcp18 proceeds at a larger rate than the basal transcription from P_{LAC} , the number of transport proteins is higher. Thus, decrease of the synthesis rate takes place at arabinose concentrations below 10^{-3} %.

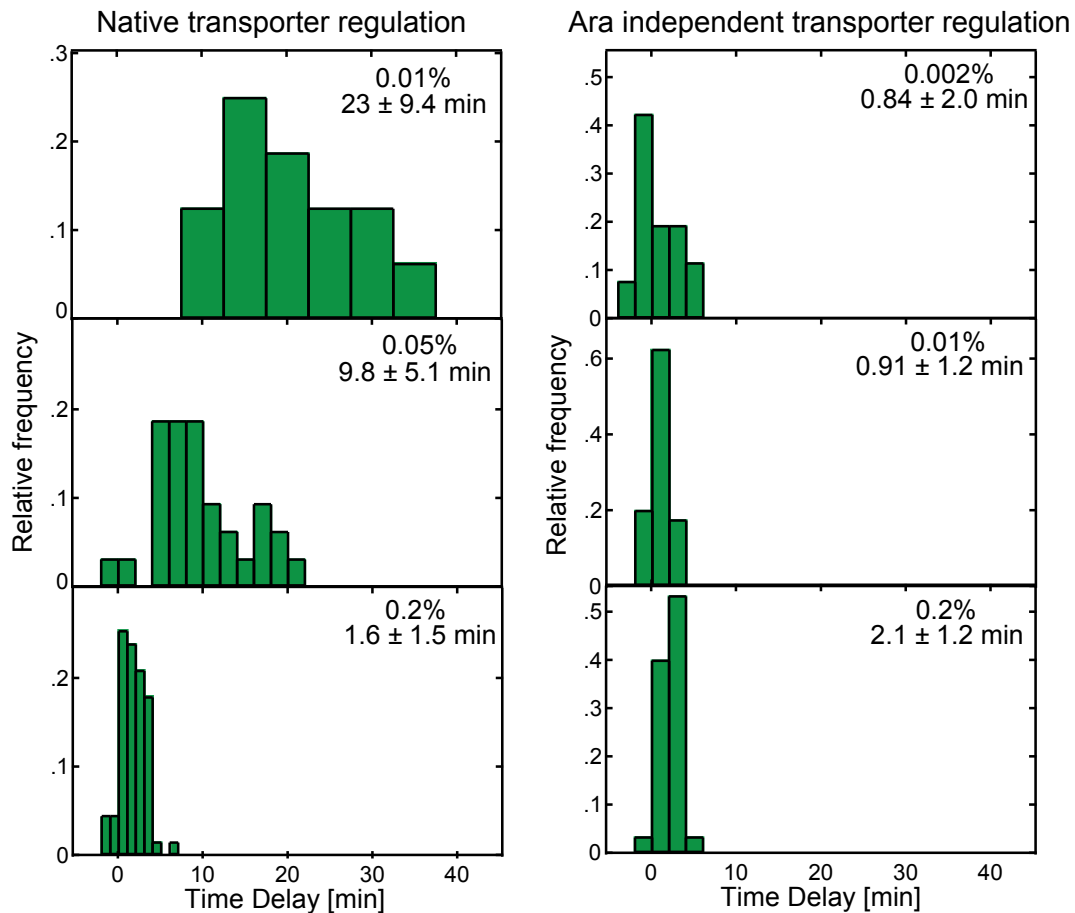


Figure 5.16: Time delay distributions at different arabinose concentrations with native (left) and arabinose independent (right) transporter expression. Arabinose concentration along with mean and standard deviations of the values are denoted in each panel. The number of evaluated cells was: native 67 (0.2 %), 32 (0.05 %), 32 (0.01 %); arabinose independent 30 (0.2 %), 40 (0.01 %), 26 (0.002 %)

5.3.2 Differences between native and arabinose independent uptake protein expression

The observation that changes in the gene expression dynamics occur for lower external arabinose concentrations with P_{LAC} controlled than with P_E controlled AraE expression suggests that native control leads to smaller numbers of transport proteins. At a given arabinose concentration this would lead to a faster onset, as this transporter number determines the initial dynamics. When comparing the single cell curves for 0.01% induction we indeed find an earlier increase with P_{LAC} controlled AraE expression. However, when the stochastic simulations, which reproduce the delay characteristics of the native case are modified accordingly (no feedback, rate of transporter production increased within the reasonable regime) we still find delayed induction and no decrease of the synthesis rate.

Moreover, the shapes of the induction kinetics differ significantly between the two strains: Following a slow and gradual increase in both cases the slope increases strongly in the native case, while there is a steady increase for arabinose independent transporter expression. The strong increase indicates the fast switching of the transcription rate from zero to a high value. Its absence in the mutant strain indicates that the fast off-on switching is caused by the positive feedback, which leads to fast increase of the number of transporters, in the native case.

As the single cell kinetics for arabinose independent transporter expression are well fit by the expression function one might assume that the transcription rate increases rather step like than gradual. However, the significantly smaller synthesis rate at 0.2% compared to the native case and the impossibility to fit the traces at 0.001% likely indicate that this approximation is worse than in the native case.

If we still consider a step like increase the small delay times in combination with the A_{ext} dependent synthesis rate indicate that the transcription rate is switched from zero to an A_{ext} dependent value shortly after induction.

Figure 5.17 illustrates the influence of the external arabinose concentration on the time evolution of the transcription rate based on the assumption of a step like increase. The time evolution is thus only characterized by the time of switching (delay time) and the final value. While in the native case the delay time varies at a constant transcription rate, the transcription rate varies and the delay time is constant.

The following consideration gives us a hint on the modifications of our model, which are obviously necessary. The dependence of the transcription rate ν on the internal arabinose concentration A is called the gene regulation function. For the promoter P_{BAD} , which controls GFP expression it is known to increase cubically with the arabinose concentration and can thus be approximated by

$$\nu = \nu^{max} * \frac{A^3}{K_A^3 + A^3}, \quad (5.6)$$

where $K_A = 50 \mu\text{M}$ [69]. As the internal arabinose concentration will likely become very

large in the native case due to the strong increase in the number of arabinose transporters the constant transcription rate found can be assumed to be ν^{max} . The variation of the transcription rates in the arabinose independent case, indicates intracellular arabinose concentrations between approximately $20\mu\text{M}$ and $100\mu\text{M}$ (see Figure 5.17 b).

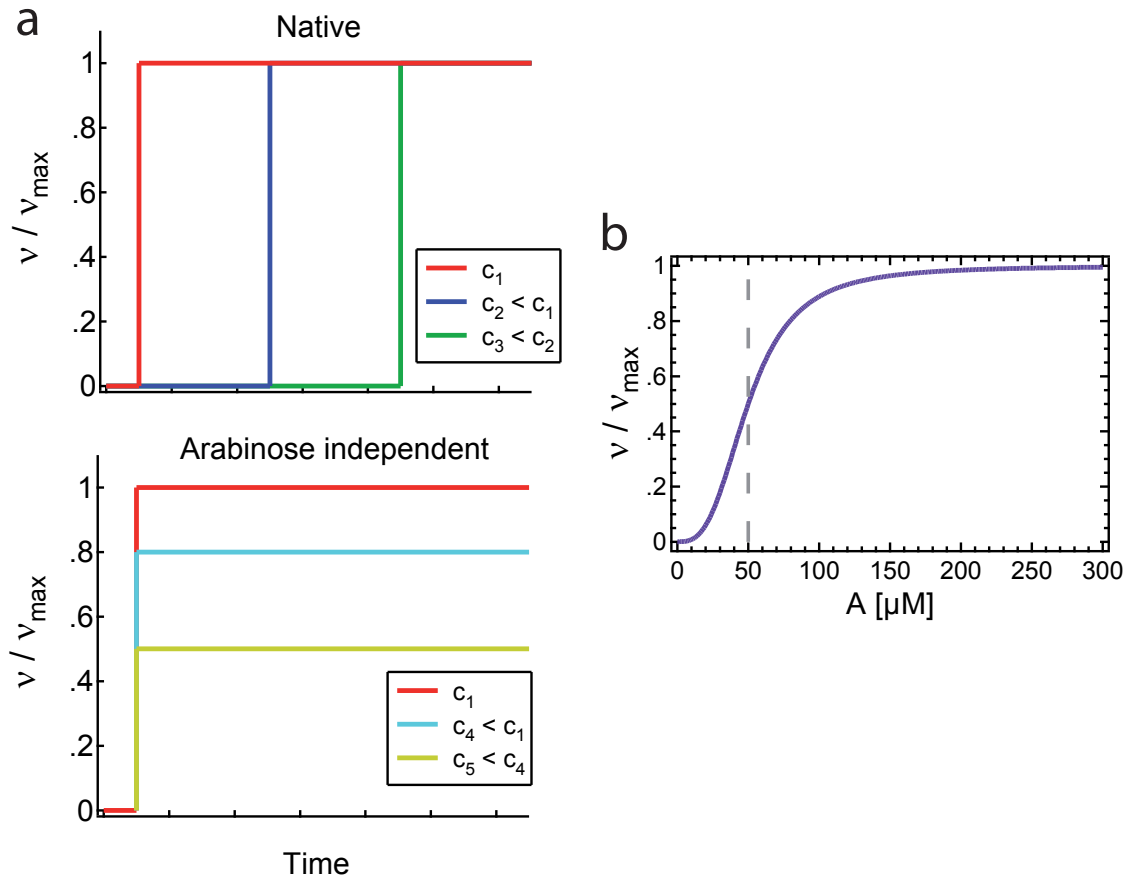


Figure 5.17: Time and concentration dependence of the transcription rate ν for native and arabinose independent transporter expression under the assumption of a step like increase (a). (b) Dependence of the transcription rate on the intracellular arabinose concentration (gene regulation function) for the promoter P_{BAD} .

5.4 A quantitative model of the arabinose system

Mathematical modeling is a powerful tool as it leads to quantitative understanding of biological networks and endows the researcher with predictive power for future experiments (see chapter 2.4). Unexpected features, like stochastic switching, can be treated in a rigorous manner and questions of evolution and optimization (cost-benefit), which are experimentally hard to grasp, can be addressed. In addition, discrepancies between a model and experimental results can indicate that the well-established picture of a process is incomplete. Based on our experimental findings and the well-known biochemical details of the arabinose system we set up a detailed model which reproduces the observed time delay characteristics observed in the arabinose system. However, we found that a mutant system, with inducer independent transporter production, does not show time delay behavior, which contradicts the predictions of this model. This indicates that the model is incomplete. In addition, it remains to be tested whether the model describes the behavior for times significantly longer than the delay period.

5.4.1 Deterministic rate equations

During the analysis of the single cell induction kinetics in the native arabinose system without degradation, mathematical modeling was employed for two purposes: First, a deterministic description of the processes involved in GFP expression yielded an analytical solution which was used to extract parameters from the data. Second, the notion that up to the delay time arabinose is accumulated via initially present arabinose transporters was

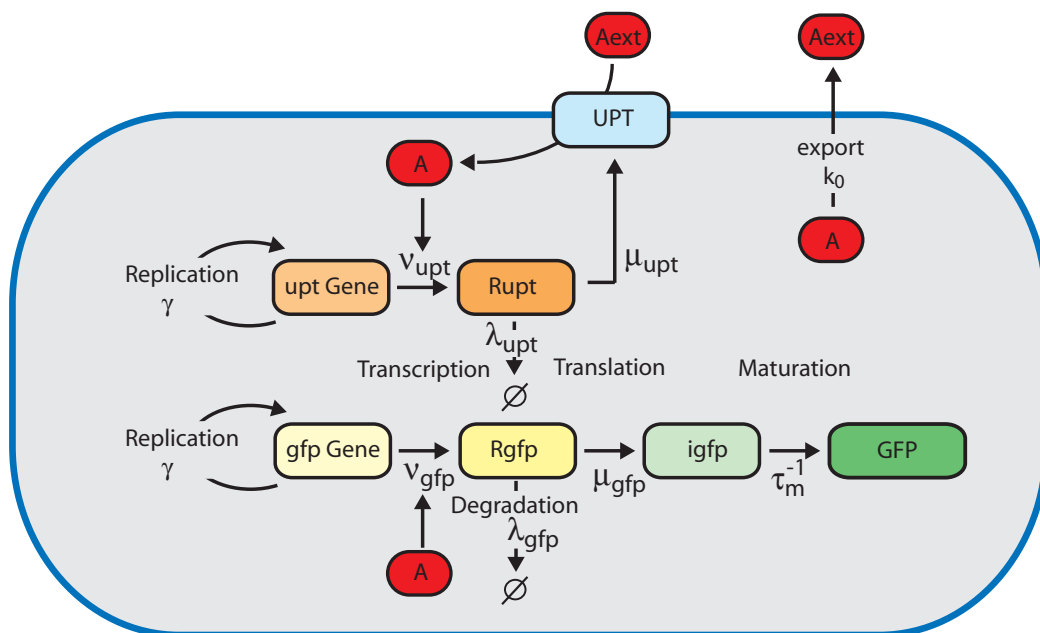


Figure 5.18: Processes, species and rates included in the model

strengthened by the derivation of an analytical description of the delay time distributions which fits the data very well. In addition, stochastic simulations of basal transporter expression and initial arabinose uptake also yielded the predicted shape of the delay time distributions. The details of this stochastic simulations can be found in [45].

To describe the entire process of gene induction and GFP expression we employ deterministic rate equations. This is feasible as we have found that the processes which are mainly subject to stochasticity are the basal production of uptake proteins and the transcription of the GFP gene. Stochasticity can thus be grasped by varying the initial number of uptake proteins and the GFP transcription rate. The processes, species and rates which are included in the model are depicted in Figure 5.18.

The first set of equations 5.7 describes parameter dependencies on the external or internal arabinose concentration. The promoter activity of P_{BAD} was found to increase cubically with the internal arabinose concentration [69] and is thus described by a Hill-function with a Hill coefficient of three. Note that $K_A = 50 \mu\text{M}$, the intracellular arabinose concentration at which transcription proceeds at half-maximal rate appears as threshold (previously denoted a_0), upon which transcription is switched on, in many instances: Due to the positive feedback of arabinose on the transport proteins the intracellular arabinose concentration usually increases very fast once a_0 is reached. Thus, the transcription rate seems to switch from zero to the maximal possible value $\nu_{\text{gfp,max}}$. The detailed promoter activity as a function of internal arabinose is not known for P_E and P_{FGH} , which regulate expression of the transport proteins. Comparison of arabinose uptake in wild type strains with *araE* and *araFGH* deletion strains revealed that the two transporters do not operate independently [76]. Instead, arabinose transport was best described by a single Michaelis-Menten function (see Appendix A for details). As they display a high similarity to P_{BAD} [68] we model transcriptional regulation of the uptake proteins by introducing a heuristic promoter P_{UPT} , that has the same characteristics as P_{BAD} , but lacks the repression in the absence of arabinose. This means that the transcription from P_{UPT} has a basal rate ν_m^0 , while the basal rate is zero for P_{BAD} . The arabinose uptake velocity V_{upt} of a single transporter shows a Michaelis-Menten type dependence on the external arabinose concentration A_{ext} . The concentration of AraC molecules changes little over time [68] and is thus omitted in the model. Binding of arabinose to AraC and of the AraC-arabinose complex to the promoter is neither included, as these processes can be assumed to be much faster than the explicitly considered reactions.

$$\begin{aligned}
 V_{\text{upt}} &= V_{\text{max}} \cdot \frac{A_{\text{ext}}}{K_{\text{upt}} + A_{\text{ext}}} \\
 \nu_{\text{upt}} &= \nu_{\text{upt,b}} + (\nu_{\text{upt,max}} - \nu_{\text{upt,b}}) \cdot \frac{A^3}{K_A^3 + A^3} \\
 \nu_{\text{gfp}} &= \nu_{\text{gfp,max}} \cdot \frac{A^3}{K_A^3 + A^3}.
 \end{aligned} \tag{5.7}$$

The second set of equations 5.8 captures the time evolution of the involved species. The internal arabinose concentration A changes by arabinose uptake and arabinose efflux which will be discussed in the next chapter 5.4.2. The mRNAs for uptake proteins (R_{upt}) and GFP (R_{gfp}) are created by transcription and chemically degraded. The translation rates are chosen according to a burst factor of 30 [75]. It is reasonable to assume that the uptake proteins (UPT) are directly functional after translation while for GFP the maturation step from immature GFP (igfp) to fluorescent GFP has to be taken into account explicitly (see Chapter 4). The number of GFP molecules has to be converted to the fluorescent signal Fl by a scaling factor S_{Fl} . The model includes replication terms for the genes and no dilution of molecules as the fluorescence signal is integrated at each time point for a growing cell.

$$\begin{aligned}
\partial_t A &= V_{\text{upt}} \cdot \text{Upt} - F_{\text{ext}} \\
\partial_t R_{\text{upt}} &= \nu_{\text{upt}} \cdot e^{\gamma t} - \lambda_{\text{upt}} \cdot R_{\text{upt}} \\
\partial_t \text{Upt} &= \mu_{\text{upt}} \cdot R_{\text{upt}} \\
\partial_t R_{\text{gfp}} &= \nu_{\text{gfp}} \cdot e^{\gamma t} - \lambda_{\text{gfp}} \cdot R_{\text{gfp}} \\
\partial_t \text{igfp} &= \mu_{\text{gfp}} \cdot R_{\text{gfp}} - \tau_m^{-1} \cdot \text{igfp} \\
\partial_t \text{GFP} &= \tau_m^{-1} \cdot \text{igfp} \\
\partial_t \text{Fl} &= S_{\text{Fl}} \cdot \partial_t \text{GFP}.
\end{aligned} \tag{5.8}$$

The solution of the model defined by Equations 5.7 and 5.8 crucially depends on the parameter values and on the initial conditions (see Table 5.2). Several of these are known from the literature or were measured directly and can thus be fixed. The GFP transcription rate, which subsumes the plasmid copy number and the initial number of uptake proteins remain free fit parameters as they vary strongly from cell to cell. For the dependence of the uptake velocity per transporter on the external arabinose concentration (see equation 5.7) a Michaelis-Menten constant on the order of $50 \mu\text{M}$ has been reported [76]. However, with such a small value there would be no variations in uptake velocity in the range of arabinose concentrations for which we observe significant differences. Thus, under the experimental conditions used here the value must be significantly larger. As we do not know this value we do not consider the dependence on the external arabinose concentration explicitly, but instead use the uptake velocity as a variable. For a reasonable set of parameters the maximal uptake velocity lies in the constrained range of 200 - 2000 arabinose molecules/transporter/min [45].

An even more realistic description of the processes considered here and their inherent stochasticity could be achieved by a detailed stochastic simulation. However, entire simulated time traces cannot be compared quantitatively to experimental ones, due to the large number of possible realizations. It is only possible to compare the distributions of quantities which can be extracted from each trace, like the time delay. However, a stochastic simulation will be used to check the plausibility of the rate model with a given

set of parameters by considering the distributions of the time delay and the initial number of uptake proteins. The simulation includes all processes captured by the rate equations and is performed analogous to [45]. Upon arabinose addition all species are allowed to accumulate as this is inherent in the experimental data evaluation.

5.4.2 Arabinose efflux

In the model used and discussed in chapter 5.2 and 5.3 only the accumulation of arabinose via import was considered. Even though there is no active arabinose degradation in the strains we use, it is likely that there is another factor: It is known that sugars are exported [77] [78] [79] and there is evidence for a relatively large arabinose export rate [80]. One of the exporters of arabinose is a major facilitator pump encoded by the *ydeA* gene [81]. At the first glance the export of sugars seems to be a waste. However, the accumulation of very high internal sugar concentrations is detrimental for the cells due to the associated osmotic pressure [82].

To address this issue experimentally, cells are induced and arabinose is taken away from the extracellular environment after a defined time. If the efflux caused a significant decrease in the intracellular arabinose concentration this would lead to a decrease of the gfp expression rate and could ultimately stop gfp expression altogether. Figure 5.19 shows single cell traces of bacteria which were exposed to 0.08% arabinose at $t = 0$ min. Arabinose was taken away at $t = 20$ min by rinsing the sample with arabinose free medium. The strain is unable to degrade arabinose and has only the AraE transport system under native control (see figure 5.13). These data clearly show that there is significant efflux, as fluorescence increase ceases approximately 15 min after arabinose removal. As this is much shorter than one cell cycle (≈ 50 min) it cannot be explained by dilution.

The data show that arabinose efflux needs to be accounted for in the model. As Novotny and Englesberg [80] found that arabinose export is a first order reaction we set

$$\partial_t A = V_{\text{Upt}} \cdot \text{Upt} - k_0 \cdot A. \quad (5.9)$$

If the number of uptake proteins is constant, which is the case prior to induction in native strains and generally in strains with p_{lac} controlled transporter expression, this leads to a steady state level of the internal arabinose concentration of

$$A_{stst} = \frac{V_{\text{Upt}} \cdot \text{Upt}}{k_0}. \quad (5.10)$$

The efflux rate k_0 determines how fast this steady state is approached. The existence of a constant intracellular arabinose concentration was suggested by the data with P_{LAC} controlled transporter expression (Chapter 5.3.2).

To show that this model reproduces the system behavior correctly we start by fitting the deterministic rate equations to the switch-off kinetics shown in Figure 5.19.

Several of the parameters can be fixed (see Table 5.2). The remaining free parameters are the uptake velocity V_{upt} , the efflux rate k_0 , the GFP transcription rate ν_{gfp} and the initial number of uptake proteins upt_0 .

As V_{upt} is determined by the extracellular arabinose concentration via Michaelis-Menten kinetics no significant cell-to-cell variations of this parameter are expected and it is required to be equal for all cells. V_{upt} is thus fixed consecutively to different values within the constrained range, while k_0 , upt_0 and ν_{gfp} are fit parameters. The GFP transcription rate adopts values on the order of 5 to 10 mRNA min^{-1} , which is in accordance with the values calculated from the mRNA levels in [68]. For each cell the value does not change significantly with the remaining parameters. V_{upt} and k_0 should fulfill two conditions, in order to be plausible: The combination of V_{upt} and k_0 should lead to delay times comparable to values found for this strain (compare Chapter 5.3.1) when simulated stochastically and the initial number of uptake proteins should conform with the distributions resulting from a realistic basal expression rate. The resulting efflux rates for $V_{\text{upt}} = 1000$ ara molecules/transporter/min was $k_0 = 4.37 \pm 0.53 \text{ min}^{-1}$. Even though the quality of the fits is slightly better for even larger efflux rates, $k_0 = 4.4 \text{ min}^{-1}$ is chosen in order to keep the V_{upt} associated with maximal expression (corresponding to delay times between 2 and

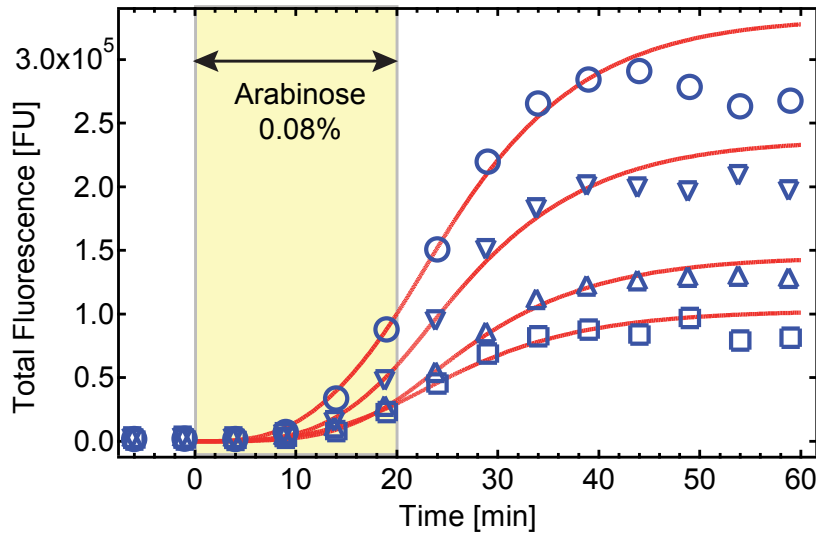


Figure 5.19: Gene expression ceases when arabinose is added and removed after a short time Experimental single cell expression traces (blue symbols) of bacteria which were exposed to a concentration of 0.08% arabinose between 0 min and 20 min. The used strain contains natively controlled AraE, but no AraFGH and is incapable of arabinose degradation. Data were fitted (red lines) with a model that contains arabinose efflux as a first order process. Note that the growth rate was 0.0092 min^{-1} in this experiment. This might explain the general trend that the off-kinetics, which is basically determined by GFP maturation, is not fitted very well.

4 minutes) in the constrained range. This efflux rate is feasible as $k_0 = 5.2 \text{ min}^{-1}$ was found in [80]. A significantly smaller value of $k_0 = 0.14 \text{ min}^{-1}$ found in [81] was determined in a strain lacking AraC, with a very small arabinose concentration and at 25°C and is thus not comparable. The mRNA levels measured in [68], the promoter fold change of 150 (for p_E and p_{FGH}) [83] and the doubling rate measured in our experiments lead to a basal expression rate of 0.03 min^{-1} . The initial numbers resulting from the fit range from ≈ 90 to 210 cell^{-1} which rather conforms with a basal expression rate of 0.05 min^{-1} . The necessity of this adjustment might result from our specific experimental conditions. In addition, the determination of the basal rate, being a very small quantity, is generally error prone.

As a final plausibility check the distribution of delay times was determined in a stochastic simulation with this parameter set. Even though the assumptions underlying the analytical derivation of the delay time distribution (Equation 5.2) are no longer fulfilled the function serves as a heuristic tool for comparison. Figure 5.20 shows that the delay distribution is still well approximated by this function with the parameters found previously. Approximately 15% of the cells in the simulation, namely those with very few transporters, did not reach the threshold level. This might indicate that, for very low initial numbers, the simulation fails to reproduce the experiment exactly.

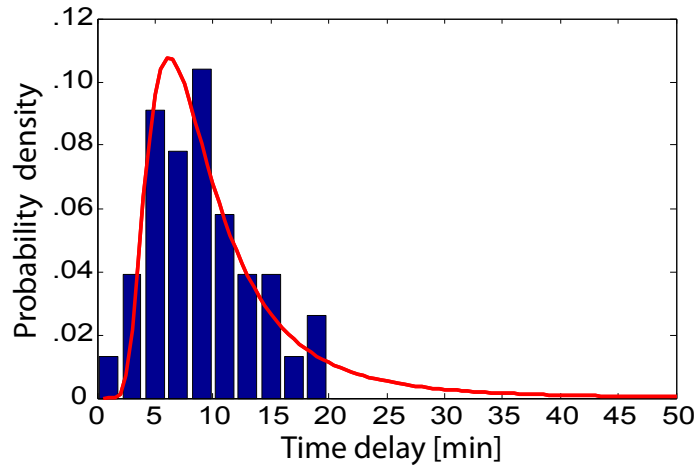


Figure 5.20: Delay time distribution resulting from a stochastic simulation (blue) with the relevant free parameters chosen as $V_{upt} = 800$ ara molecules/transporter/min, $k_0 = 4.4 \text{ min}^{-1}$ and $\nu_{upt,b} = 0.05 \text{ min}^{-1}$ following the discussion in the main text. The distribution is well approximated by the analytically derived one (red, Chapter 5.2.4) with $m = 3.8$, $b = 30$ and $\tau_0 = 900$. The delay was determined as the time at which the internal arabinose concentration crosses the threshold of $a_0 = 50 \mu\text{M}$. Approximately 15% of the cells in the simulation did not reach this threshold.

fixed parameters		
K_A	$50 \mu M$	intracellular arabinose concentration for half-maximal transcription rate [69]
γ	0.0139 min^{-1}	growth rate (corresponds to a doubling time of 50min) measured
$\nu_{upt,b}$	0.03 min^{-1}	basal transporter transcription rate results from $\nu_{upt,max}$ with a fold-change of 150 [83]
$\nu_{upt,max}$	4.16 min^{-1}	maximal transporter transcription rate derived from mRNA levels in [68]
λ_{upt}	0.347 min^{-1}	transporter mRNA degradation rate mRNA half-life 2min [68]
λ_{gfp}	0.116 min^{-1}	GFP mRNA degradation rate mRNA half-life 6min [84] [85]
μ_{upt}	10.4 min^{-1}	transporter translation rate chosen to yield a typical burst factor of 30 [75]
μ_{gfp}	3.5 min^{-1}	GFP translation rate chosen to yield a typical burst factor of 30 [75]
τ_m	6.5 min	GFP maturation time measured (chapter 4)
S_{Fl}	50	Fluorescence conversion factor arbitrary
free parameters		
V_{upt}		arabinose uptake velocity constrained to 200-2000 arabinose molecules/transporter/min[45]
$\nu_{gfp,max}$		maximal GFP transcription rate includes plasmid copy number
k_0		arabinose efflux rate see discussion in Chapter 5.4.2
initial values		
A	0	internal arabinose
R_{upt}	$\frac{\nu_{upt,b}}{\lambda_{upt}}$	uptake proteins mRNA
upt	variable	uptake proteins
R_{gfp}	0	GFP mRNA
igfp	0	immature GFP
GFP	0	fluorescent GFP
Fl	0	Fluorescence

Table 5.2: Model Parameters and initial values Prior to arabinose addition internal arabinose and all GFP species are not present. The initial value of the mRNA for the uptake proteins results from its basal expression rate, which can in contrast be assumed to be zero for GFP expression. The number of uptake proteins is the quantity which distinguishes the cells.

5.4.3 Comparison of modeling results and induction kinetics with native and arabinose independent transporter expression

Now we show that the expression kinetics of a strain in which transporter expression is arabinose independent, along with the data of the according native control strain, can be explained by the model which includes arabinose efflux (see chapter 5.3). Delay times increase with decreasing arabinose concentration and the protein synthesis rates is constant at all concentrations in the native control strain. In contrast, we find very short delay times and a decrease of the protein synthesis rate with decreasing concentration for arabinose independent transporter expression.

Direct fitting of the deterministic equations to the single cell expression kinetics is not feasible, as there are too many unconstrained parameters, which compensate each other. Instead we show that the model qualitatively reproduces the characteristics of the fluorescence traces of the strains with native and P_{LAC} controlled transporter expression. For the latter case the uptake protein transcription rate is kept constant. Representative traces for different initial numbers of transporters and different uptake velocities are shown in Figure 5.21. The key characteristics, delayed induction for decreasing uptake velocity and similar production rates (slopes) for all velocities in the native case, in contrast to fast increase for all uptake velocities with decreasing production rates for arabinose independent transporter expression are reproduced by the model.

5.4.4 Comparison of modeling results and the expression dynamics following arabinose pulses

As a first step towards more complex temporal variations of the environmental conditions bacteria which are incapable of arabinose degradation were exposed to two subsequent pulses of 0.05% arabinose. Resulting traces along with the according modelling result are shown in Figure 5.22. In this case the fluorescence signal was integrated over the entire microcolony resulting from one mother cell. Following the first arabinose addition it takes the characteristic delay time for this concentration until gene expression starts. Subsequent arabinose removal leads to the ceasing of GFP expression, similar to the observation in Figure 5.19. The response to the second arabinose addition is significantly faster than the first one. This is due to an increased number of uptake proteins, which were produced during the first pulse. This can be seen as a memory effect, which increases sensitivity to arabinose. Due to the high rate of induced uptake protein expression and the stability of the proteins, which are only degraded by dilution this memory effect is expected to persist for several cell cycles. The data are predicted very well by the model.

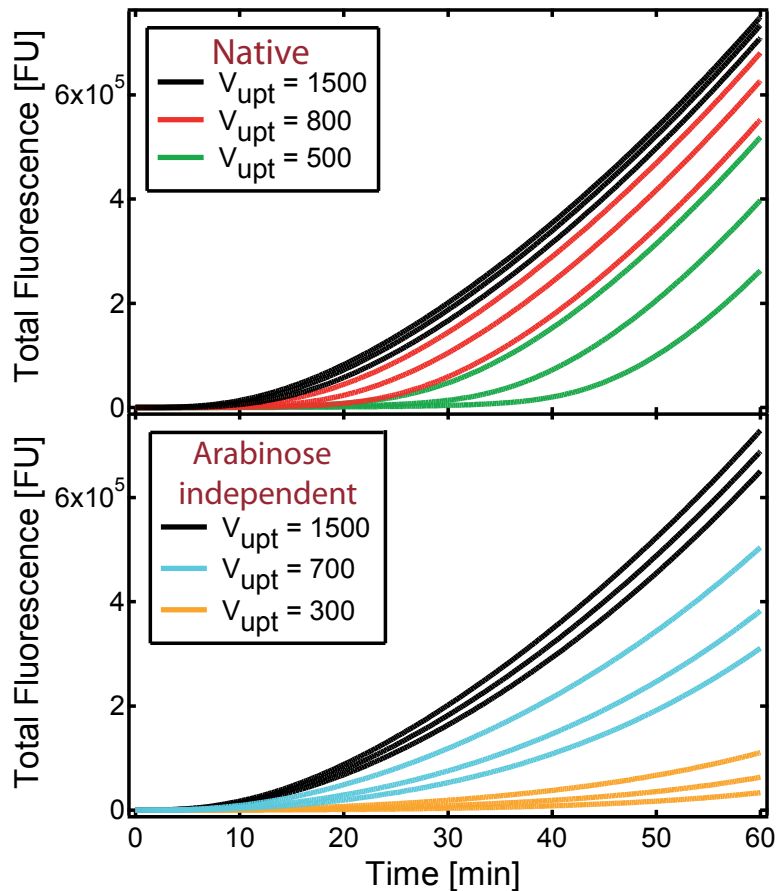


Figure 5.21: Modeling results for native and arabinose independent transporter expression The solution to equations 5.7 and 5.8 are shown for different numbers of initial uptake proteins Upt_0 at various arabinose uptake velocities V_{upt} (given in arabinose molecules/transporter/min). To reproduce arabinose independent transporter production the arabinose dependence of the transporter transcription rate in equation 5.7 is omitted. Upt_0 equals 80, 110, 150 in the native case and 180, 220, 300 in the arabinose independent case, representing the transporter distributions at the time of arabinose addition for native and P_{LAC} controlled expression. Different numbers were used as our experimental results indicate that the initial number of transporters is higher with P_{LAC} controlled than native expression. However, the numbers of uptake proteins do not influence the qualitative behavior, but only the exact scaling with V_{upt} . Variations of the transcription rate, which only represent an additional degree of variation, were omitted for this plot.

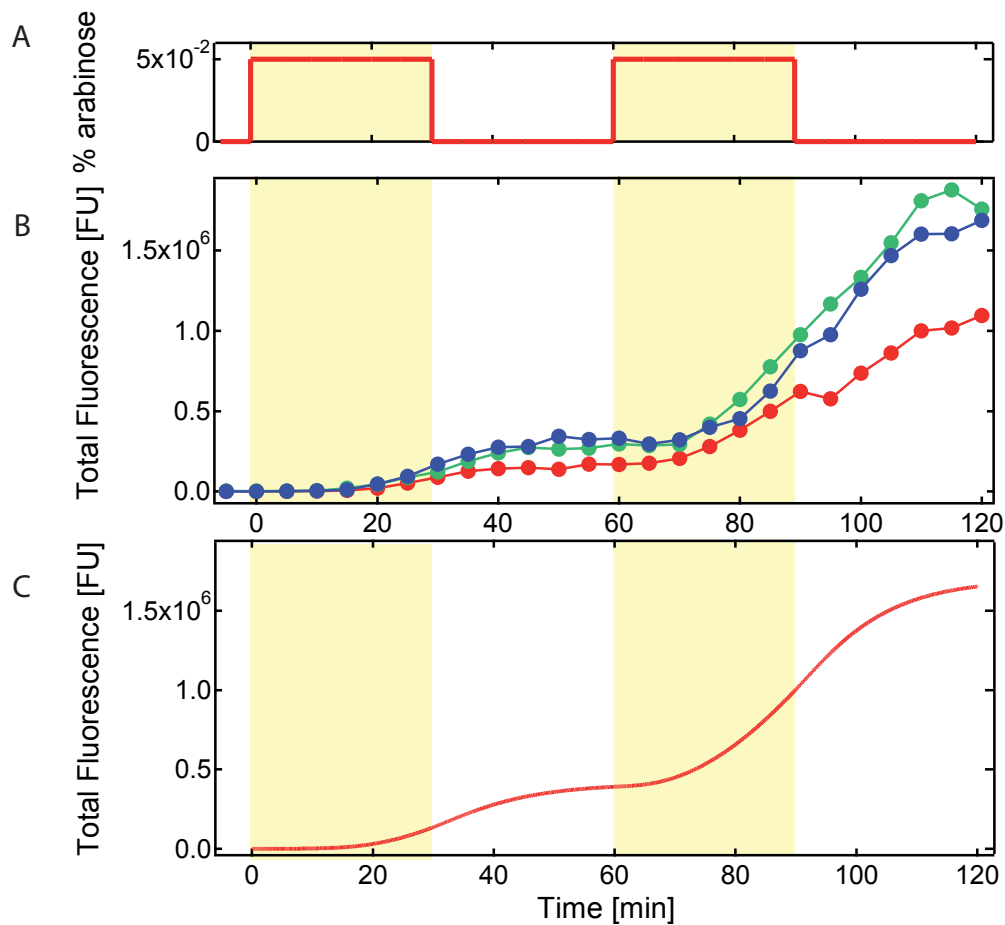


Figure 5.22: Single colony response to arabinose pulses The response of single bacterial cells, which grow into colonies over time, to subsequent pulses of 0.05 % (a) arabinose were recorded (b). The traces are qualitatively well predicted by the rate equation model. (c). The absolute fluorescence value in the model is determined by an arbitrary scaling factor. In addition, this factor is specific for each cell as it also captures differences in the protein production rate.

5.4.5 Scope of the model and future experiments

A mathematical model which describes the response dynamics of the arabinose system was developed. It is based on the known biochemistry of the arabinose system and includes an additional efflux term. With one set of parameters the model consistently explains all experimental observations, which were made for different network architectures and for different types of the arabinose stimulus. These observations are:

- The qualitative scaling of the time delay distributions with the arabinose concentration and their shape as well as the constant protein synthesis rate for native expression of the arabinose transporters.
- The decreasing protein synthesis rate without time delay for arabinose independent expression of the transporters.
- The increase and subsequent ceasing of GFP expression when arabinose is added and removed after a short time.
- The expression kinetics following two subsequent arabinose pulses.

In the first part of our study we derived an analytical gene expression function which is based on the assumption of a step like switching of the transcription rate from off to on once an intracellular arabinose threshold is reached. The full model, which in contrast includes the exact dependence of the transcription rate on the intracellular arabinose concentration, illustrates why the step like approximation is valid for the native arabinose system: Initially, arabinose is accumulated via the basally expressed transporters. Once the intracellular arabinose concentration comes into the range in which transcription is modulated the number of transporters increases strongly due to the positive feedback. Thus, the intracellular arabinose concentration increases above the range in which the transcription rate is modulated very fast, resulting in the apparent switching upon a threshold concentration.

Future experiments will address the question whether arabinose export is best described by a simple first order process. It seems likely that at least a correction is necessary, as at high intracellular arabinose concentration the arabinose transporters presumably function as exporters. This functionality is included in models of the lactose system [86] [87]. For this system, the rates and affinities for import and export are chosen equally, resulting in equal external and internal sugar concentrations. However, for the arabinose system it is known that the intracellular arabinose concentration becomes significantly higher than the external concentration [80].

We will monitor the response to arabinose addition and subsequent removal in the strain with native and arabinose independent transporter production. The concentration and the duration of the pulses will be varied. One characteristic feature predicted by the first order efflux for the case with arabinose independent transporter production is that the switch off kinetics do not depend on the duration of the arabinose pulse, as the intracellular

arabinose reaches a steady state level determined by the uptake velocity, the number of uptake proteins and the efflux rate. Furthermore, complementary experiments, such as biochemical transport studies or the determination of the number of uptake proteins in individual cells [59] might be necessary to show the exact efflux mechanism and determine the biochemical rates directly.

These future challenges for the model will help to deepen the knowledge of the processes involved in gene induction of the arabinose network, thus contributing to the understanding of the working principles of cells.

5.5 Cell density dependent gene expression

An important question regarding stochastic effects in gene expression is whether they are side effects of certain network architectures or whether they confer an advantage. Considering the cost and benefit [31] of the arabinose uptake and degradation machinery, we suggest that heterogeneous timing is beneficial when arabinose is only available for a limited time due to consumption by the bacteria (see figure 5.23 for an illustration). When a given amount of arabinose is added to differently dense populations of bacteria, this carbon source will be depleted faster at higher cell densities. The production of the arabinose degradation machinery is initially an energetical burden for each cell. This investment of energy does only pay off if the cell can consume a sufficient amount of arabinose subsequently. The probability for a cell to switch the arabinose operon on at a given point in time depends on the external arabinose concentration. At a high cell density, the depletion of arabinose due to the cells which have already started to consume arabinose decreases the external concentration and thus the probability that further cells switch the arabinose operon on. In this way, cells are prevented from switching on, which are unlikely to overcome the energy invested by producing the degradation machinery due to the total depletion of arabinose.

5.5.1 Distributions of single cell fluorescence

To test this hypothesis experimentally, we prepared cultures containing *E.coli* MG1655 at different cells densities and induced these with the same amount of arabinose. *E.coli* MG1655 is able to degrade arabinose and was transformed with the reporter plasmid pBAD24-gfp, in order to have a fluorescence readout for the expression of the arabinose system. At defined times after induction, samples were taken from the cultures and the distribution of single cell fluorescence values was determined with a flow cytometer. In our experiments, arabinose is not the only carbon source: To guarantee growth at a basal rate, glycerol or succinate are supplied as carbon source, which is usually done when studying gene expression of inducible networks. Furthermore, an additional burden is imposed on the bacteria by the expression of GFP, which is present on a plasmid with ≈ 50 copies per cell. Remarkably, growth rates measured in liquid cultures without and with arabinose were the same within the experimental accuracy.

We expect that in a culture with higher cell density a larger fraction of the bacteria is in the non fluorescent state than in a culture with lower cell density (see figure 5.23 d). In contrast, we observe the opposite, namely that the fraction of fluorescent cells is larger at higher cell densities (Figure 5.24). Remarkably, this property already becomes obvious 45 min after induction.

We performed several tests in order to rule out experimental artefacts and to find the cause of the observed effect. Due to the initially used experimental protocol, the growth rate differed between the samples with different cell densities, but a control in which

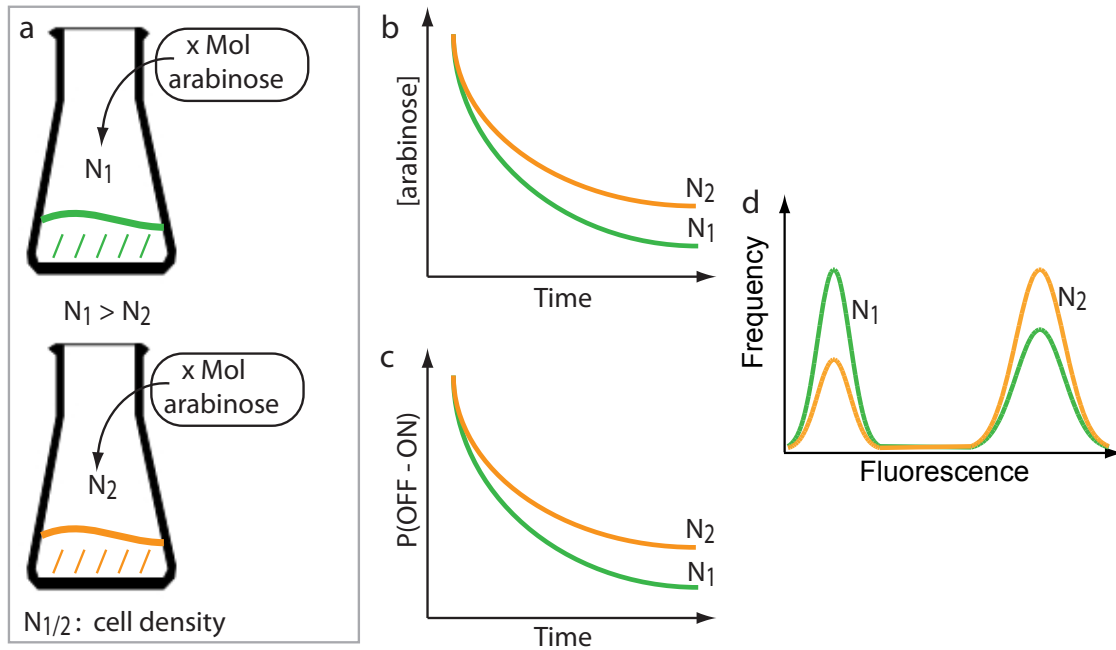


Figure 5.23: Proposed beneficial effect of heterogeneous timing at high cell densities The same amount (x Mol) of arabinose is added to two bacterial cultures with different densities $N_1 > N_2$ (a). Soon, a fraction of the population (those with the largest numbers of uptake proteins) switches the arabinose system on and consumes arabinose. Thus, the larger absolute number of consuming bacteria in the culture with higher density results in a faster decrease of the arabinose concentration in the medium (b). The probability to switch the system on decreases concomitantly as it depends on the arabinose concentration (c). As a result, the fraction of bacteria which are on, detected in the experiments as bacteria with a high fluorescence level, is expected to be larger in the culture with lower cell density at a time T_1 after induction (d).

the growth rates did not differ yielded qualitatively identical results. The effect did not either vanish, when glycerol, the carbon source used to guarantee growth at a basal rate was exchanged by succinate. The observed behavior reminds of quorum sensing (QS) mechanisms, where cell density controls gene expression (See introduction in chapter 6). QS is mediated by small molecules, so called autoinducers, which each cell produces and which can pass the cell membrane. The autoinducer concentration thus depends on the cell density. Once a sufficient concentration is reached the autoinducers stimulate or inhibit the expression of certain genes. When recording the density dependent gene expression in *E.coli* DH5alpha, which does not produce a common type of autoinducers [88] the observations were similar to the prior experiments, indicating that they are not caused by quorum sensing via these autoinducers.

To test whether some compound, which is excreted by the cells and accumulates in the medium, is the effector, we performed the following test: High density cultures² (OD600

²The density of bacterial cultures is usually determined by measuring their optical density at 600nm

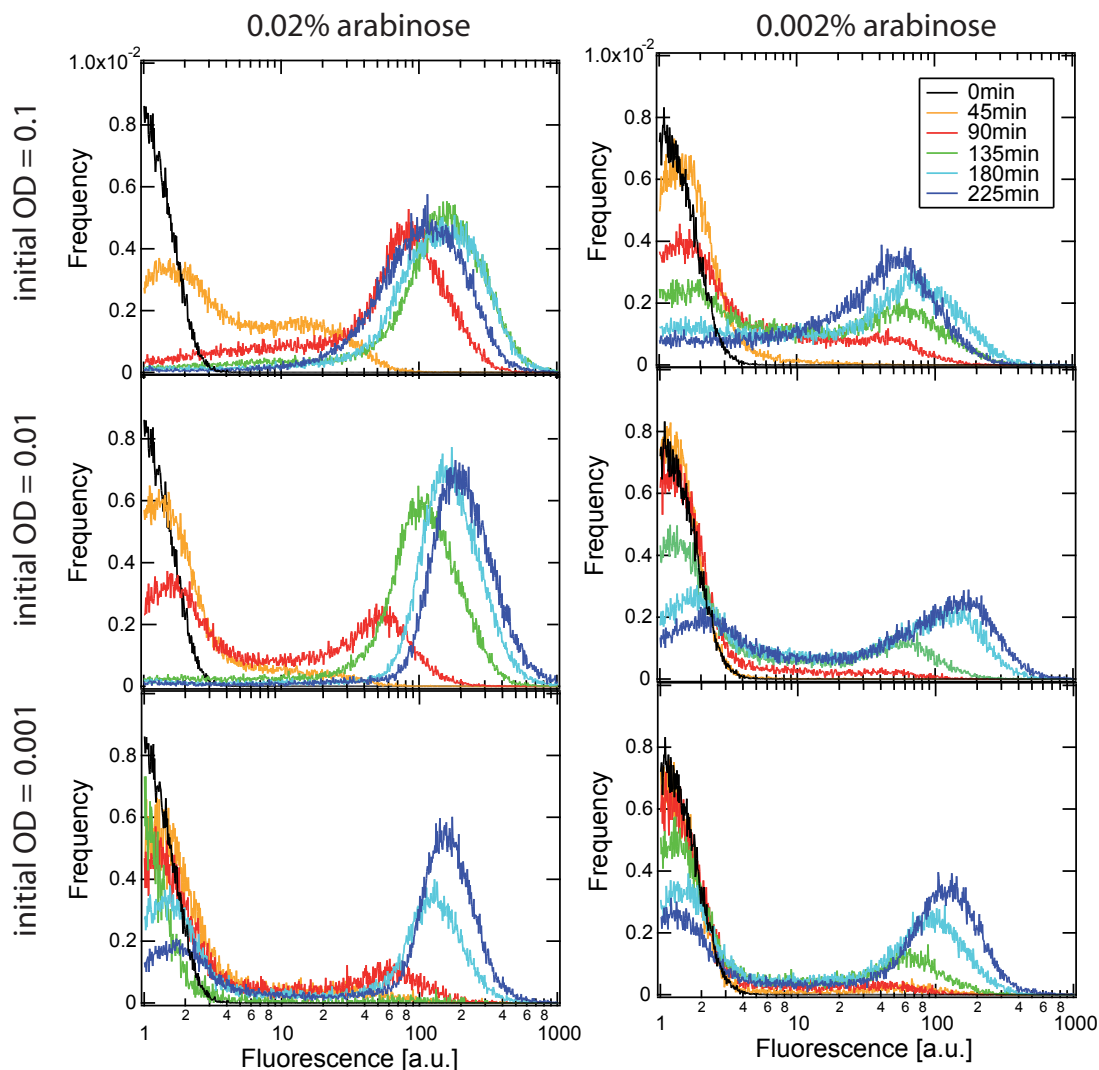


Figure 5.24: Density dependent gene expression Cultures of different initial cell densities ($OD_{600} = 0.1, 0.01, 0.001$) were induced at $t=0$ min with 0.02% (*left*) or 0.002% (*right*) arabinose. Distributions of fluorescence values were recorded with a flow cytometer at different times after induction (see legend).

$= 0.1$) were grown as in the initial experiments. 90 min and 270 min after induction the supernatant was collected by centrifugation and filter-sterilized. In a second experiment, low density cultures ($OD_{600} = 0.001$) were grown in fresh medium supplemented with 10% of the collected supernatant. Using a larger fraction of supernatant would negatively influence the growth rate due to the accumulated waste products. The population distributions of parallel cultures without supernatant or the two different collected supernatants show differences which indicate that the response is slowest in the culture with pure medium (see Figure 5.25). Thus, the density dependent effect seems to be mediated by some compound which accumulates in the medium.

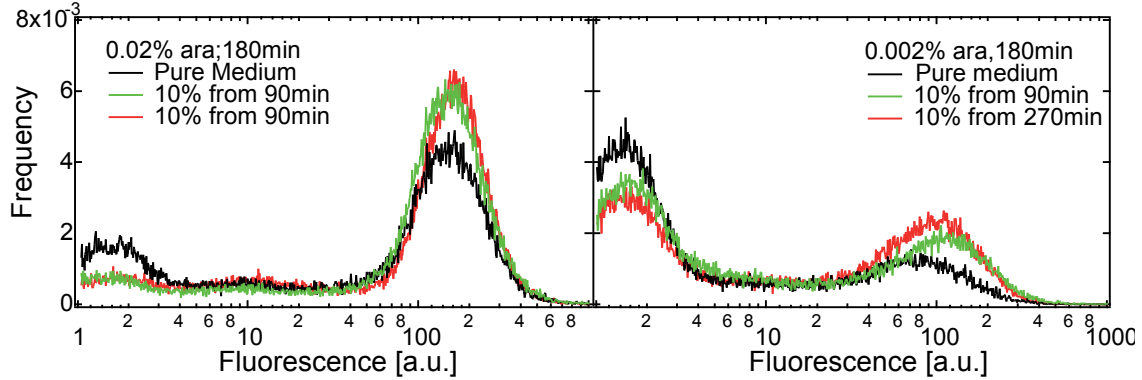


Figure 5.25: Density dependent gene expression with conditioned medium Low density cultures (initial OD600 = 0.001) were either grown in pure medium or medium supplemented with 10 % supernatant of high density cultures 90 min or 270 min after induction. The population distributions of fluorescence values 180 min after induction with 0.02 % or 0.002 % are shown.

5.5.2 Causes of density dependence

Our hypothesis was that one possible beneficial effect of the heterogeneous timing would manifest itself in a larger fraction of cells which express the arabinose system at a high level at lower cell densities. The exact opposite, larger fractions of cells which express the arabinose system at a high level at higher cell densities, was found in the experiments. We tested whether glycerol, the carbon source used to maintain basal growth, or autoinducer mediated quorum sensing gives rise to this effect, but did not find an influence. Experiments in which low density cultures were supplemented with medium from high density cultures indicate that the mediator is a small compound which accumulates in the medium. This might be a molecule such as acetate, which is a byproduct of metabolism and is excreted into the medium [89].

Even though we do not know how the effect is mediated, we can use our mathematical model to speculate about the species or rate which might be affected. For switching to occur earlier, the internal arabinose concentration A has to rise faster. As the time evolution of A is given by

$$\partial_t A = V_{upt} \cdot U_{pt} - k_0 \cdot A \quad (5.11)$$

we see that the number of transporters U_{pt} , the transport velocity V_{upt} , or the efflux rate k_0 must be varied to change the dynamics of the internal arabinose concentration. The initial number of transporters, which is responsible for arabinose uptake prior to induction, could be varied by changing the basal transporter expression rate. As arabinose promoters are known to be regulated by cAMP (cyclic-Adenosine-Monophosphate) via CRP (cyclic AMP receptor protein) [67] involvement of this mechanism seems possible. Usually, it prevents the expression from the arabinose promoters if a more favorable carbon source

than arabinose is available. Still, due to the fact that the production of cAMP and CRP are both intricately regulated, the level of one of these could be affected. The uptake velocity of AraE, which is a proton symporter, could be influenced when released compounds change the pH value of the medium. Finally, as we do not know the exact mechanism of arabinose efflux it is hard to judge whether and how this parameter might be influenced by different cell densities.

Another possibility is that the effect is mediated indirectly, via differences in the growth rate: Even though growth rate was measured and found to be identical there might still be variations: Determination of the growth rate at low culture densities has a relatively large error due to instrumental resolution. In addition, it seems possible that growth rate variations take place on time scales below the resolution of our measurement. Recent studies indicate that the initial cell density has an impact on the growth rate [90], but this subject has not been analyzed in detail. As it is known that there are global regulatory effects, which are due to the growth rate it is likely that one of the quantities discussed above might be influenced [91].

Finally, a benefit of the fast reaction at high cell densities is conceivable when considering that in natural habitats usually several species of bacteria coexist. If the sensing mechanism only indicates a high cell density, irrespective of the cell type, the faster uptake of sugar at higher cell densities is the best way to maximize the proliferation of the genotype of an individual cells.

5.6 Conclusions

In this chapter we showed that the arabinose system, which displays stochastic all-or-nothing gene expression also shows heterogeneity in the timing of gene expression onset. The delay time before onset is mainly determined by the number of arabinose uptake proteins a cell has at the time of induction: Arabinose is accumulated by these transporters until the intracellular threshold for induction is reached.

Mathematical modeling was crucial in order to understand how the experimental observations arise from the underlying biochemical network. Particularly, discrepancies between experimental observations and the predictions of an initial model, which represented the well-established view of the arabinose system, could be resolved by including arabinose efflux: Experimentally, we found this process to be so fast that it significantly influences the dynamics.

In contrast to the native network, for which the transcription rate switches from zero to the same value for all inducer concentrations, the transcription rate varies with the external arabinose concentration when the positive feedback of arabinose on the transport proteins is deleted. It has been reported previously [71] [72], that the binary response can be converted to a graded one by this modification. For this network architecture, we find that transcription starts directly upon induction, while delayed induction is expected from previous simulations [16]. To see how the discrepancy in timing arises we first need to note that the transcription rate is only modulated within a narrow range of intracellular arabinose concentrations. Below this range, the transcription rate has a low basal rate, while above this range transcription constantly proceeds at the maximally possible value. In the native system, the intracellular arabinose concentration increases fast and strongly, meaning the range in which modulation happens is crossed rapidly. Thus, we observe switching from low to maximal transcription. In contrast, without positive feedback a steady state concentration on the order of the Michaelis-Menten constant is rapidly established due to the combination of arabinose export with arabinose influx through the constant number of transporters.

The heterogeneity in timing was characterized for a range of arabinose concentrations which eventually lead to induction of all cells of a population. For the highest used arabinose concentrations induction is so fast that the number of transporters is constant between inducer addition and gene expression onset. For smaller arabinose concentrations the average number of transporters in the period between inducer addition and gene expression onset determines the delay time. For even lower concentrations arabinose uptake via the average number of uptake proteins does not lead to induction. Instead, induction is only possible directly upon a burst of transporter expression. This regime and the molecular details of transient derepression which causes the bursts have been characterized in detail for the lactose system [59].

Mechanisms by which bacteria adapt to changes in the environment have been classified into those which are based on sensing and others which use population diversity generated

by stochastic switching [92]. In this sense, the behavior of the arabinose system can be considered as a mixed strategy, which uses elements of both classes. A broad distribution of arabinose transport proteins is created by stochastic gene expression. Once arabinose becomes available, this results in temporal variability in the sensing response. In this way the bacteria can profit from a fast response which is associated with sensing as well as from population heterogeneity. In addition, the number of molecules which are necessary to render the response possible is very low, meaning that only little energy needs to be invested constantly.

Future research on the arabinose system will address the question how and under which conditions bacteria benefit from the observed heterogeneity. From a cost-benefit argument we assumed that heterogeneous timing is beneficial when the sugar is depleted due to consumption by the cells, which should lead to differences in the distribution of cells between off and on state at different cell densities. However, our experiments indicate that there is an adversary regulatory effect, so that the exact opposite of our expectation was observed. Thus, we will analyze both a possible beneficial effect at temporally limited arabinose concentrations and the density dependent gene regulation in future experiments. Theoretical work [12] also indicates that stochastic gene expression is optimal under certain kinds of environmental fluctuations. For a synthetic model system this has also been shown experimentally [13]. As we can use our experimental setup to monitor bacteria while the arabinose concentration is varied temporally, we will analyze whether the fitness, i.e. the growth rate, is larger for specific types of fluctuations, e.g. pulses of a certain duration.

Furthermore, we will check whether heterogeneous timing is also found in other inducible systems. Actually, this is to be expected, as many of them share the positive feedback architecture. If so, it is worthwhile to revisit the phenomenon of diauxic growth on the single cell level: Already in the 1950's Monod observed that bacterial cultures which were supplied with two sugars did not consume them at once, but one after the other [93]. In these experiments usually one of the sugars was glucose, the best carbon source. The glucose degradation machinery is constantly present. The other sugars had inducible degradation machineries. Thus, the question remains how single cells respond to two sugars with inducible systems: Does each cell only switch on one of the systems, which is suggested by theoretical work [94]? Is there a strong hierarchy which always leads to the preferred consumption of one sugar or are there conditions under which a fraction of the cells specialize on one sugar, while the remainder chooses the other one? The latter could be explained by the combination of the early onset of one of the systems due to heterogeneous timing and the subsequent inhibition of the other system. If there is indeed a stochastic decision between two systems, this mutual switch will be a valuable tool for the construction of artificial gene circuits.

6 Dynamics of AHL mediated quorum sensing

In nature, many bacterial species grow attached to surfaces. This situation is significantly different from the one in standard laboratory experiments, where bacteria grow in liquid medium which is constantly shaken. The spatial localization is bound to affect the dynamics of quorum sensing (QS) regulated gene expression: This very common form of cell-cell communication is mediated by signaling molecules which are excreted by the cells. The diffusive transport of these molecules between spatially localized bacteria influences the timing of gene expression. Furthermore, it is likely that spatial concentration gradients arise by reaction-diffusion mechanisms due to the constant production of signaling molecules by the surface bound cells.

Biofilms can either be symbiotic, for instance on plant roots, but they can also cause severe diseases, for example when the lung surface is colonized by pathogenic bacteria. Biofilm formation is a QS regulated phenomenon of surface attached bacteria. In a biofilm [95] [96], layers of surface bound bacteria are surrounded by a matrix of extracellular polymeric substances (EPS) which holds the cells together and protects them. In addition, chemicals might accumulate in the biofilm. Cells in a biofilm usually behave significantly different from dispersed cells in liquid medium. In particular, a biofilm has traits of multicellular organisms [97]. The bacteria exhibit different gene expression patterns [98], which can be seen as a division of work. As the bacteria in the film are genetically identical the differentiation is either stochastic or due to different conditions in different regions of the film.

The previously studied arabinose system, as well as the analyzed QS system use a positive feedback architecture to switch the gene expression rate from a very low to a high value upon an external stimulus. While arabinose requires active transport proteins to cross the cell membrane, the QS signaling molecules enter and exit the cell by diffusion.

Our experimental setup allows for the analysis of gene expression dynamics in surface attached bacterial cultures. We use this framework to study the very early stage of biofilm formation in which single cells grow into microcolonies. Essential QS dependent changes might already occur at this stage. Particularly, we analyze whether QS regulated gene expression is influenced by spatial concentration heterogeneities and whether cell-to-cell variability occurs.

6.1 Introduction to quorum sensing

Quorum sensing (QS) [99] is a form of cell-cell communication which has been found in a large number of bacterial species. Besides being a wide-spread regulation mechanism it is of particular importance as it plays a fundamental role in many severe infectious diseases.

QS is mediated by small molecules, so called autoinducers, which are constantly produced in low amounts by the bacteria. These molecules enter or exit cells by diffusion through the cell membrane. Due to the constant production and the increase of the cell number by growth, signaling molecules accumulate. Once the autoinducer concentration is sufficiently large they bind to regulator molecules. In turn, these activated regulators change the expression of QS regulated genes, usually from low to high expression rates (Figure 6.1). In most QS networks, the production of the autoinducer synthase is also stimulated, resulting in even more autoinducers. Thus, positive feedback regulation is employed (see figure 6.2 for the illustration of a typical QS system).

Functionality and purpose of QS in natural ecosystems are still debated: As the term quorum sensing indicates it was originally assumed that QS serves as an indicator of a critical density of bacteria above which certain collective behaviors, such as the production of a protective extracellular matrix, are feasible [100]. However, when bacteria grow sparsely and inhomogeneously distributed on surfaces, which is often the case under natural conditions, the autoinducer concentration rather depends on the spatial distribution of the cells. In addition, their concentration is strongly influenced by the consistence of the surrounding environment, which influences autoinducer diffusion. Thus, it is likely that quorum sensing is used to probe a combination of cell density, spatial cell distribution and environmental mass transfer properties [101]. This hypothesis, also denoted as efficiency sensing, is supported by experiments, which use artificial QS sender and receiver cells arranged in a well defined spatial pattern [102]. Further evidence comes from an experiment in which it was shown that maximal distance between sparse bacteria which can still communicate ("calling distance") varies between different parts of a plant root [103].

Our goal is to address the predictions of the efficiency sensing hypothesis in a systematic way using a natural QS system. To this end, we measure QS regulated gene expression dynamics in a small channel, which allows changing the autoinducer concentrations via a flow system. In the future, we will employ microstructured sample environments with compartments of different sizes to probe the influence of spatial limitations. In addition, patterns of adhesive coating on the surface will reveal the effect of the distance between the bacteria.

6.2 Gene expression dynamics

In the following we resolve the quorum sensing mediated induction dynamics of growing microcolonies of the biofilm forming bacterium *Pseudomonas putida* IsoF. This strain

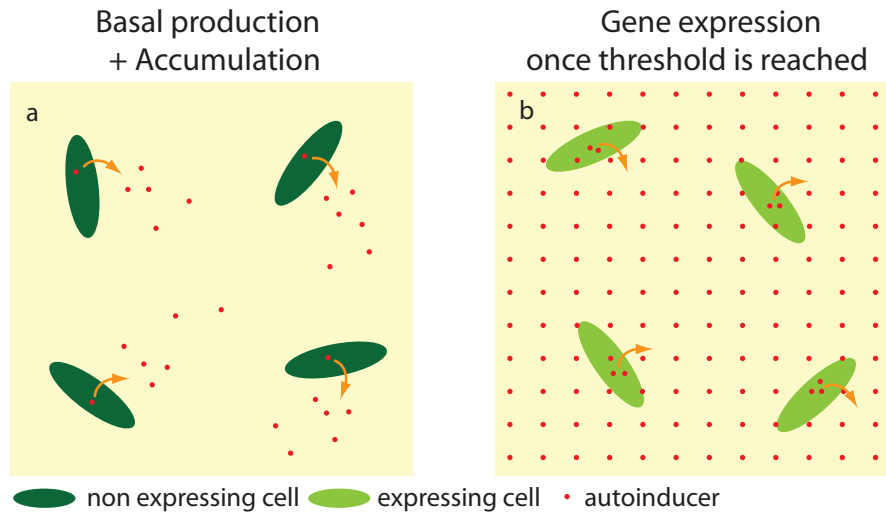


Figure 6.1: Schematic illustration of quorum sensing (a) Autoinducers are produced at a basal rate and accumulate in the medium. (b) Once the induction threshold is reached expression of QS regulated genes is stimulated. Due to the positive feedback, the production rate of autoinducers increases as well.

contains only one quorum sensing system, called Ppu system. We analyze the response dynamics when autoinducers are added to the culture. To find out whether these dynamics are influenced by spatial heterogeneities occurring due to the autoinducer accumulation we use two different experimental conditions: First, the environment is left undisturbed, so that autoinducers can accumulate. Second, the concentration is kept constant by a steady flow of medium.

6.2.1 Experimental system

The *Pseudomonas putida* PPU system consists of the regulatory activator PpuR, the autoinducer synthase PpuI and the structural gene PpuA, which is involved in biofilm formation [104] [105]. Figure 6.2 schematically depicts the quorum sensing and reporter system of the *P. putida* strain used in this study. The autoinducer of the PPU system is a N-acyl-homoserine lactone (AHL). When the AHL concentration is sufficiently large, the autoinducer AHL binds to the activator PpuR. The PpuR+AHL complex activates expression of the AHL synthase PpuI, thus constituting a positive feedback loop which results in increased AHL production. AHL molecules diffuse through the bacterial membrane, thereby mediating the signal to neighboring cells. A fluorescent reporter (GFP_{asv} [50]) under the control of P_{lasB} on a RP4 type plasmid is used as reporter for the AHL concentration [103]. Similar to the regulation in the PPU system, AHL binds to the regulator LasR when the AHL concentration is sufficiently large. The LasR+AHL complex activates gene expression from P_{lasB} .

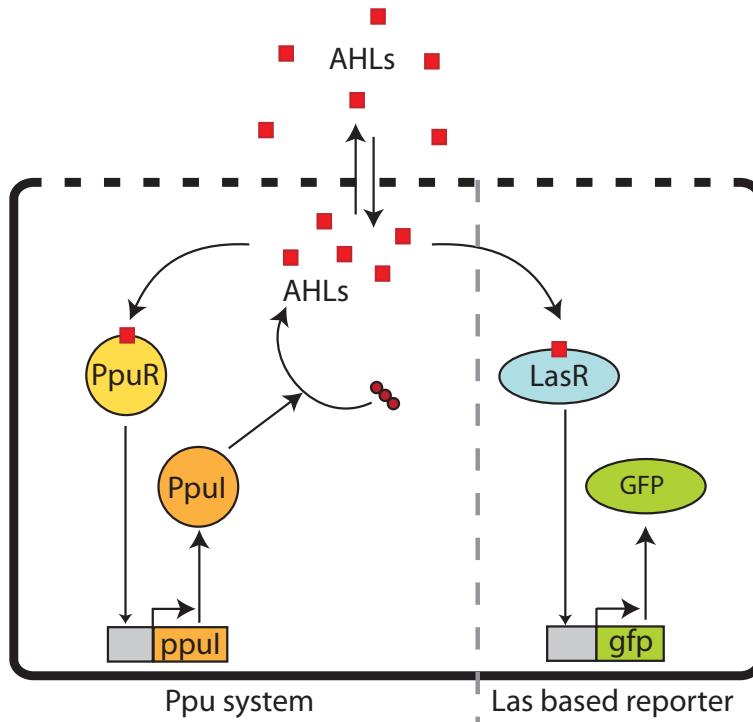


Figure 6.2: Regulatory network of the *PPU* System and Las-based reporter
 The *PPU* system consists of the regulator PpuR, the AHL synthase PpuI and the structural gene PpuA (not depicted), along with their genes and respective promoters [104] [105]. When bound to AHL, PpuR stimulates the expression of PpuI. GFP under control of the promoter P_{LasB} from the *Pseudomonas aeruginosa* Las system is used as a reporter for QS controlled gene expression [103]. When bound to AHL, LasR stimulates the P_{LasB} controlled GFP-expression. The reporter construct, which also includes the gene for the regulator LasR, is plasmid-borne.

6.2.2 Fluorescence decrease

Following overnight growth, the bacteria are highly fluorescent. We monitored the fluorescence decrease of individual cells in fresh medium (Figure 6.3) since bacteria are required to be in a well-defined, non-fluorescent off-state before the switch-on of the quorum sensing system can be measured. Cultures were diluted to different percentages of the cell density of the overnight culture, incubated for one hour and bacteria were subsequently transferred to microfluidic channels. The time course of the population average of the mean fluorescence of single cells was measured. The initial fluorescence level was higher in the previously less diluted cultures (Figure 6.3 b). Therefore, we conclude that the QS-system was still active after the dilution into liquid medium from the overnight culture. When normalized by the initial fluorescence value the curves of all samples collapse onto one universal curve which decreases exponentially with a rate of $1.16 \pm 0.02 h^{-1}$ (Figure 6.3 c). The identical rate of decrease indicates that the AHL production was reduced to basal level in all cases by the additional dilution effectuated by transferring the cells to the microfluidic

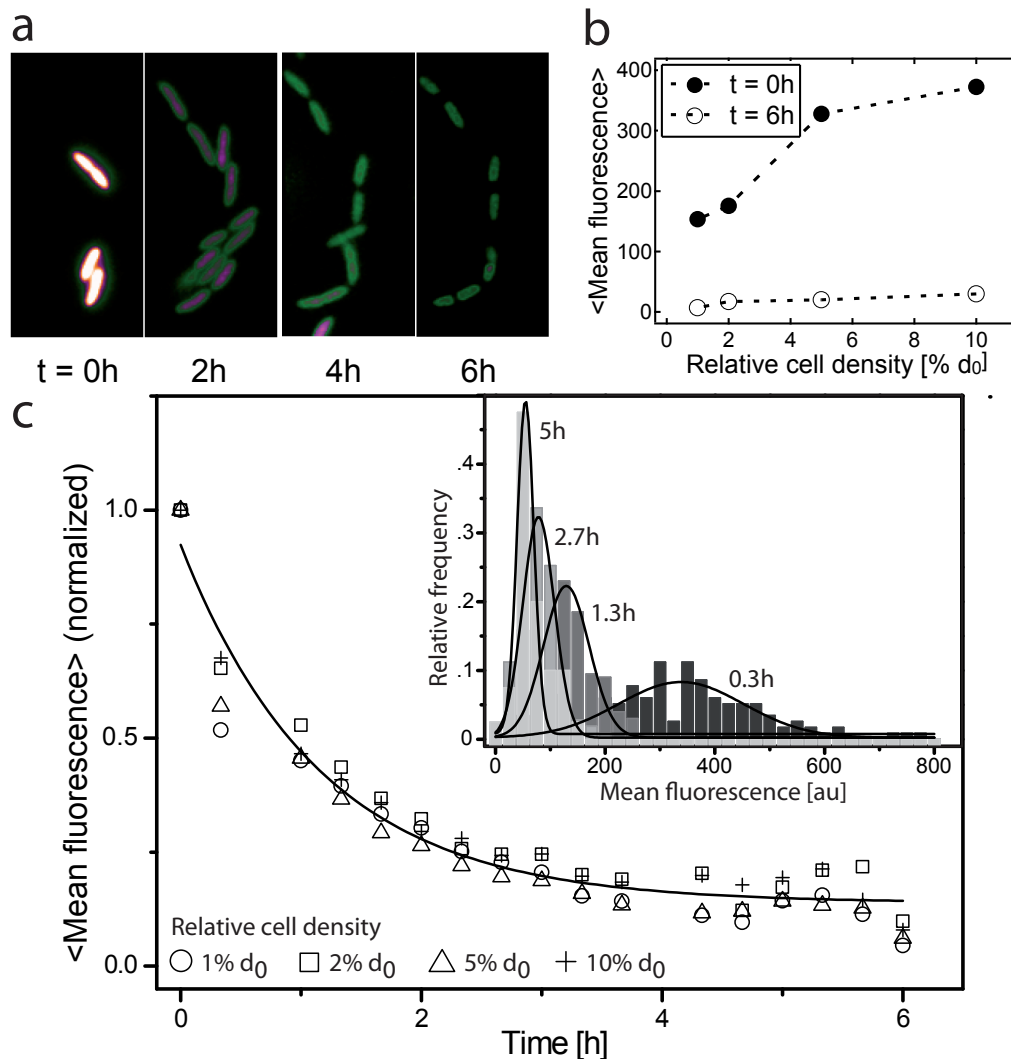


Figure 6.3: Single cell fluorescence decreases exponentially upon dilution to a low cell density in fresh medium. Bacteria were inoculated from an overnight culture, incubated for one hour, seeded into a microfluidic channel and rinsed. The first time at which the bacteria were imaged is defined as $t = 0h$. (a) A representative image series shows the fluorescence decrease of growing cells (length of one cell is approximately $3\mu m$). Some daughter cells disappear from the images as they detach from the surface after cell division. (b) Dependence of the $\langle \text{Mean fluorescence} \rangle$, which is the population averaged mean fluorescence of single cells on the initial cell density. The initial cell density was adjusted to different percentages of d_0 , the density of the overnight culture, by dilution. (c) Time evolution of the $\langle \text{Mean fluorescence} \rangle$. The results from the cultures which were adjusted to different densities collapse onto a single exponentially decaying curve when they are normalized to the initial fluorescence level. (*Inset*) The distributions of single cell fluorescence values over the population at different time points are well fitted by Gaussian functions. Their width decreases with time.

chamber. The GFP level is reduced by two processes: Dilution by cell division (doubling time 51min) and biochemical degradation of GFP_{asv} (half-life 110min). Together, these two processes are expected to give rise to decay rate of $\frac{1}{51 \text{ min}} + \frac{1}{110 \text{ min}} = 1.2 \text{ h}^{-1}$ which is in excellent agreement with the experimental value. The single cell fluorescence values fall onto a Gaussian distribution at all times which indicates homogeneous QS gene expression in the well-mixed batch pre-culture. Over time, the width of the distributions decreases, while the relative width remains constant (Figure 6.3 c). Following these results we prepare the off-state for the induction experiments by three cycles of dilution and growth (see Appendix B).

6.2.3 Induction with external AHL

In order to artificially induce QS regulation, we use 3-OxoC10 AHL (N-(3-Oxodecanoyl)-L-homoserine lactone), the cognate AHL of the *PPU* system [105]. In our experimental setup, in which the bacteria are immobilized on the Poly-L-Lysin coated surface of microfluidic channels we can either apply a constant medium flow, or leave the system undisturbed. Constant flow is provided by a pumping system, which can be connected to the channels. When investigating the undisturbed situation without flow, mineral oil is used to prevent evaporation of the medium. In this case, the AHLs which are produced by the bacteria can accumulate in the medium and contribute to induction. In contrast, under medium flow, the AHL concentration is constantly kept at the externally provided value. Leaving the environment undisturbed probably resembles the conditions in natural habitats more closely than continuous exchange of medium by a constant flow. However, due to the limited size of the flow chamber the provided nutrients might soon be depleted. Within our observation time of approximately 8 h bacteria grow continuously at similar rates both with and without flow, indicating that nutrient supply is still sufficient. The flow rate of 2 ml/h is so small that it does not give rise to shear stress on the cells.

For each external AHL concentration flow and non-flow condition are measured in a parallel set-up of two channels on a single microfluidic slide. The attached single cells grow into microcolonies over time (Figure 6.4 a). For each colony a trace of the mean fluorescence values (fluorescence per pixel) is extracted from the image series (See figure 6.4 b-d). The colony outline is determined by thresholding of bright-field images. As it is a priori not known whether there is cell-to-cell variability in QS regulated gene expression we analyze entire colonies. Thus, a lower magnification can be used and better statistics can be obtained. Evaluation of single cells is not possible, but visible inspection allows to decide whether variations are so pronounced that experiments shall be repeated with higher resolution. We only consider colonies from cells which are present in the beginning of the experiment and divide at least two times.

For the one-time addition (non-flow situation) of 0nM, 10nM and 100nM AHL the resulting single colony traces, along with their mean value, are shown in figure 6.4 b-d. At 100nM there is an approximately linear fluorescence increase starting at AHL addition. At

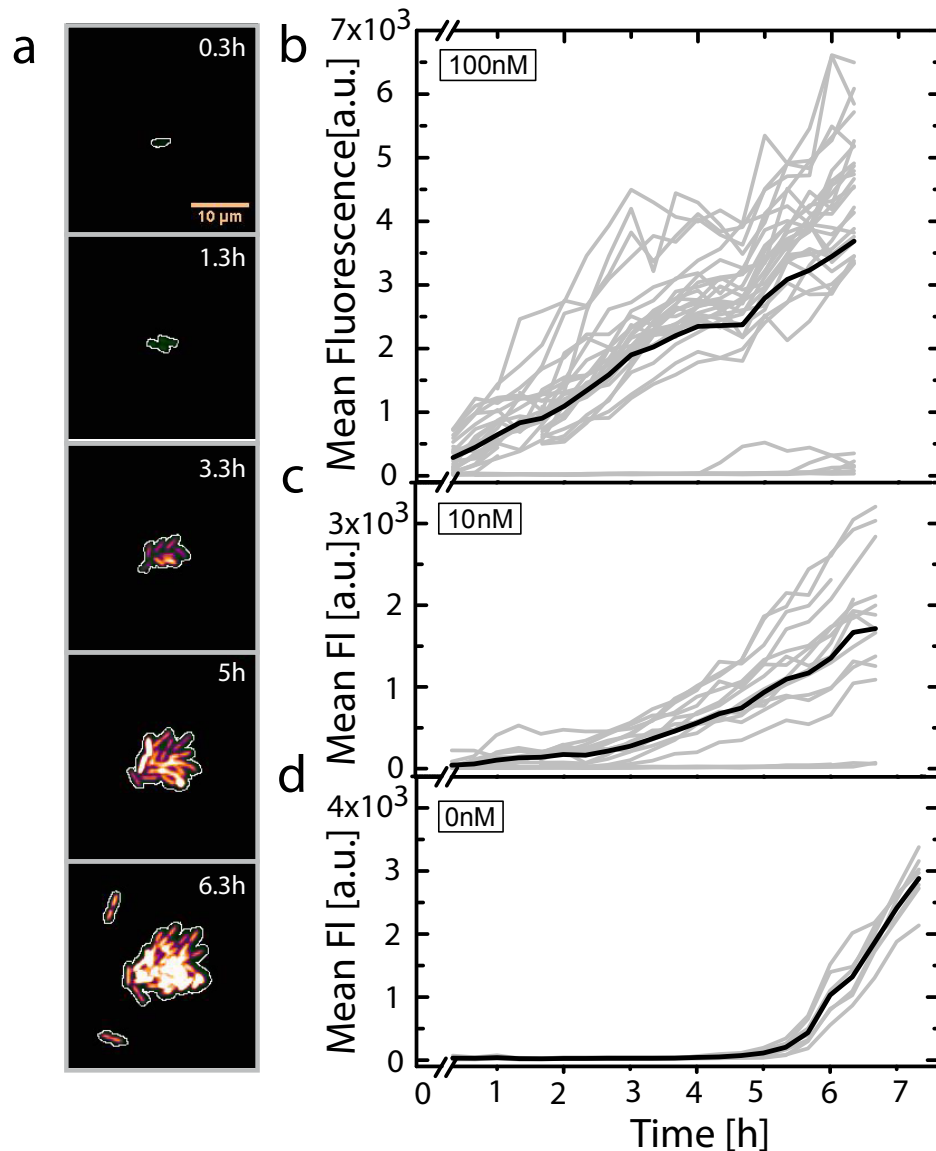


Figure 6.4: Single colony expression kinetics for induction with different external AHL concentrations (0nM, 10nM and 100nM) without flow (a) Image time series of a representative colony at 100nM AHL. The outline of the colony, determined from the bright-field image, is marked in white. Scale bar: 10 μm (b)-(d) Time-traces of the mean fluorescence of individual colonies are shown in grey, the black line represents their average.

10nM the fluorescence increase starts slowly and accelerates with time. At 0nM there is no detectable increase in fluorescence up to approximately 5h, when a sudden and strong increase starts. At all concentrations there are significant colony-to-colony variations in timing and absolute values. It is observed that for all three AHL concentrations the cells eventually detach from the surface. Since a quantitative and continuous representation of single colony fluorescence fails at this point, evaluation of experiments ends after 6 to 7

hours.

Figure 6.5 b shows the mean value of the expression profiles (*symbols*) resulting when the respective AHL concentrations are applied under constant flow (2 ml/h). For 100 nM AHL and 10 nM AHL the dynamics are similar to the non flow situation. Fluorescence increase is initially slightly slower, but speeds up over time. When 0nM AHL are applied via flow no fluorescence can be detected during the time of the experiment. It is remarkable, that in contrast to the experiments without flow bacteria do not detach from the surface with flow at any AHL concentration. Instead, bacteria grow into a dense layer.

6.2.4 Comparison of the data and a rate equation model

In [105] a rate equation model was developed which explains the expression of autoinducers of the *PPU* system in batch cultures. AHL is produced at a small basal rate all the time. Once the AHL concentration exceeds a threshold concentration K_{ppu} , the AHL production rate increases strongly due to a positive feedback loop. The time evolution of the concentration of AHL in the medium can be described by

$$\partial_t AHL = \left(\alpha + \beta \cdot \frac{AHL^n}{K_{ppu}^n + AHL^n} \right) \cdot N - \gamma \cdot AHL \quad (6.1)$$

Due to bacterial growth, the concentration of bacteria evolves as $N(t) = N_0 \cdot e^{r \cdot t}$, where r is the growth rate. The basal and induced expression rates, α and β , as well as the exponent n of the Hill function and the AHL degradation rate γ were determined by fitting experimental data sets. The parameter values are summarized in table 6.1.

To test whether this model is also able to explain our experimental observations we need an additional equation describing the GFP production which is controlled by the P_{las} promoter. From the negligible fluorescence values at 0 nM AHL with flow we conclude that a basal production of GFP can be neglected. As the Las system is similar to the *PPU* system and also activated via a positive feedback mechanism, we employ a Hill-function n with the same coefficient which was found for the *PPU* system. However, the induction

basal AHL production rate	α	$2.3 \cdot 10^{-10} \text{ nmolcell}^{-1} \text{ h}^{-1}$
induced AHL production rate	β	$2.3 \cdot 10^{-9} \text{ nmolcell}^{-1} \text{ h}^{-1}$
Hill exponent	n	2.5
Ppu Induction threshold	K_{ppu}	70 nM
AHL degradation rate	γ	0.005545 h^{-1}
Las Induction threshold	K_{las}	30 nM
GFP degradation rate	k	0.378 h^{-1}
Initial concentration of bacteria	N_0	$4.3 \cdot 10^9 \text{ l}^{-1}$
Growth rate	r	0.44 h^{-1}

Table 6.1: Parameter values for AHL production and GFP expression. α , β , n , K_{ppu} , γ from [105], K_{las} adjusted from [103], k from [50], N_0 , r measured.

threshold K_{las} has a different value, which is reported in the literature [103]. As we do not measure numbers of GFP molecules, but the fluorescence signal, the transcription rate β_{GFP} has an arbitrary value, which includes the conversion of the number of GFP molecules to a fluorescence signal. The used GFP variant, GFP_{asv} , is not stable, so degradation at a rate k is included. GFP_{asv} is similar to GFP_{mut3} for which we determined the maturation time. As it is small, compared to the time-scales we investigate here, GFP maturation is omitted. The time evolution of the fluorescence per cell in arbitrary units, Fl , is thus described by

$$\partial_t Fl = \beta_{GFP} \cdot \frac{AHL^n}{K_{las}^n + AHL^n} - k \cdot AHL \quad (6.2)$$

N_0 , the initial concentration of bacteria, is estimated from the number of bacteria attached to the surface and the flow chamber volume. The growth rate r is also determined from the experiments. Both quantities are equal for all experiments within the resolution of the estimates (see table 6.1 for parameter values). To compare model and data we use these values and adjust the scaling between the number of GFP molecules and the fluorescence signal appropriately. When we shift K_{Las} from 20nM [103] to 30nM, which is reasonable, as the exact value is strain dependent [106] the data without flow are quite well approximated. However, for the data with flow there are large discrepancies, particularly at 10 nM and 100 nM AHL.

The data recorded at 100 nM increase rather linearly, both without and with flow. The model, however, predicts a sigmoidal increase. Presently, there is no explanation for this

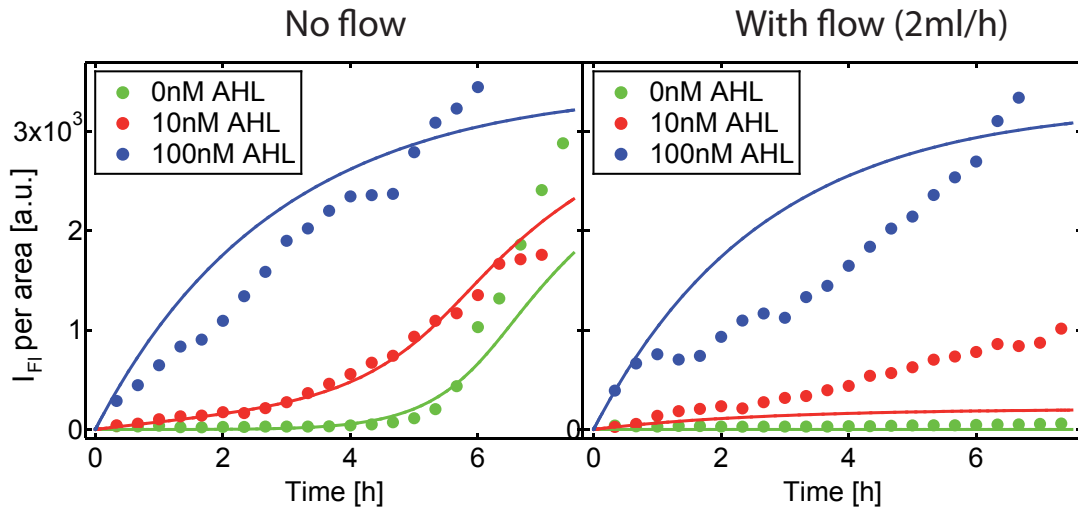


Figure 6.5: Experimental and predicted gene expression dynamics Experimental population averages of the mean fluorescence of single colonies acquired without and with flow at different external AHL concentrations are depicted by markers. Corresponding predictions by the rate equation model (Equation 6.1 and 6.2) are shown as solid lines.

discrepancy.

A challenge which remains to be solved is the determination of the initial bacterial concentration N_0 and the growth rate r : These parameters strongly influence the theoretically determined dynamics, but could only be estimated roughly.

The curve measured at 10 nM with flow lies above the model prediction and shows an upward kink similar to the observation for 10 nM AHL without flow. This observation indicates that the AHL concentration is higher within the cells than in the medium flow. Possibly, the increase of the concentration arises by one of the following three mechanisms: First, the intracellular AHL concentration might be larger due to the combination of AHL production and diffusion of AHL through the cell membrane. Second, the AHL concentration might be larger in the vicinity of the cells and thus also within the cells due to the combination of AHL production and excretion by the cells and the diffusion of AHL in the medium. Third, accumulation could be caused by the production of extracellular matrix proteins which inhibit diffusion of the AHLs. To distinguish between these possibilities models will be established which consider two compartments with different AHL concentration: An internal one, which includes the AHL producing bacteria and an external one, which comprises the cell free surrounding. The rate of AHL exchanged between the two compartments determines the degree of accumulation. If AHL accumulates within the cells, the inner compartment is each individual cell, so that the cell-membrane is the exchange barrier. If AHLs accumulate in the vicinity, the inner compartment is a small volume which includes one colony.

6.2.5 Colony-to-Colony and Cell-to-Cell variability

It is known that within biofilms genetically identical cells differentiate to fulfill specific functions. These specialized cells are not distributed equally within the biofilm, but preferentially occupy certain regions, e.g. close to the edge [98]. As we are interested in the timing of biofilm formation it is interesting whether variability already arises at the early stage we are investigating.

As noted above, we find significant variations between the fluorescence time courses of individual colonies (Figure 6.4 b-d). The colony to colony variability is similar without and with flow. We assume that these variations are to a large extent caused by differences in the size and composition of the initially seeded cells and by extrinsic noise. The colony to colony variation is once more illustrated by the distribution of single colony fluorescence values (Fig. 6.6 b). Variability is much more pronounced at 10 nM AHL and 100 nM AHL, indicating that it is increased by the positive feedback architecture of the regulatory network. It remains unclear why at 10 nM AHL and 100 nM AHL a fraction of the cells do not become fluorescent at all.

There is also heterogeneity between cells within each colony. Within colonies the fluorescence decreases in most cases from one center towards the rim (Fig 6.4 a, 6.6 a). This indicates a larger AHL concentration near the center of the colonies. To address these

variations in detail, experiments with higher optical resolution will be performed, so that fluorescence values of single cells within colonies can be determined.

A reaction-diffusion model will be used to finally understand the entire spatio-temporal gene expression pattern. This model will also show to which extent the heterogeneous gene expression is due to spatial heterogeneities of the AHL concentration and to which extent it is caused by stochasticity in gene expression.

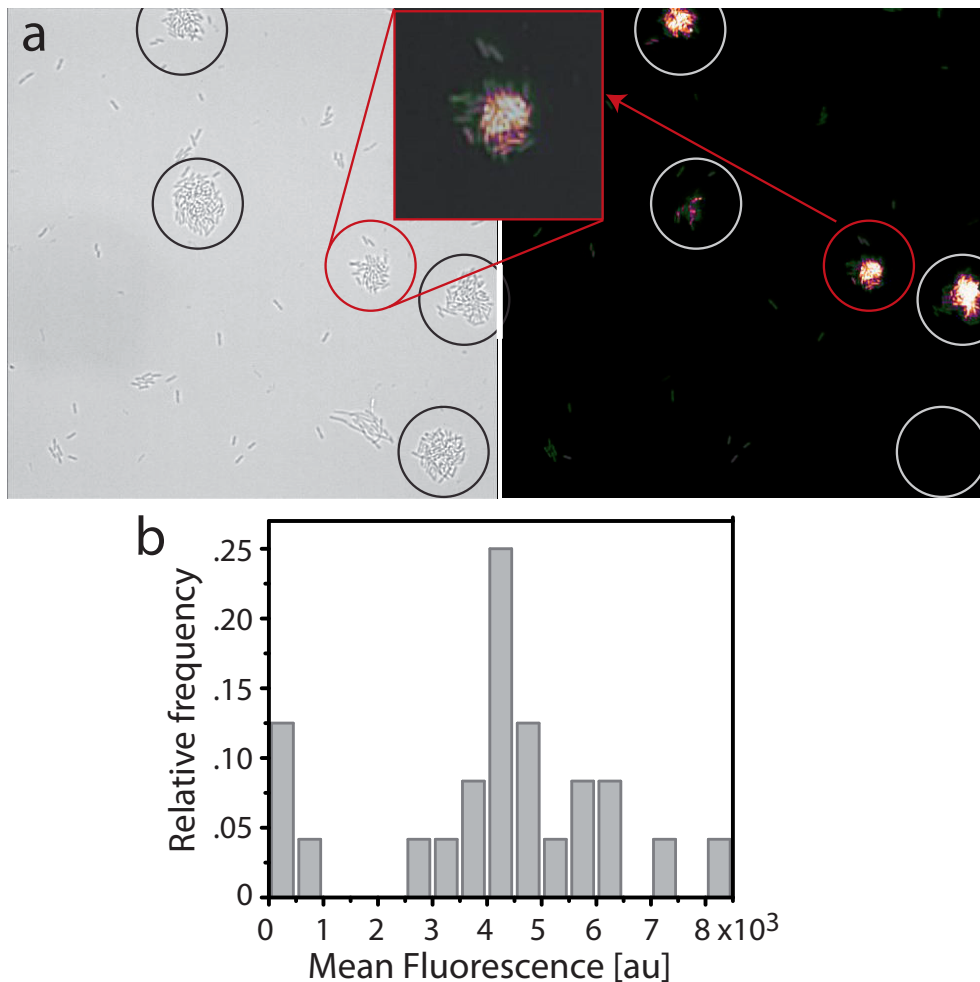


Figure 6.6: Intra and inter colony heterogeneity (a) Bright field and fluorescence image of a sample after 7h of growth with 100nM external AHLs under constant flow. Colonies of similar size show different fluorescence levels. The overlay picture of a single colony shows that the brightness of the cells decreases from the middle to the edges of a colony. (b) Distribution of the mean fluorescence of colonies at 100nM AHL after 7h of growth under constant flow.

6.3 Discussion

We analyzed the dynamics of quorum sensing regulated gene expression, while leaving the environment either undisturbed or keeping concentrations constant by a continuous flow of medium. Kinetics recorded in the undisturbed environment are very similar to predictions of a rate equation model which was developed to describe gene expression in mixed liquid cultures. In contrast, particularly the data set for one concentration close to the induction threshold recorded under constant flow deviates strongly from predictions by the same model. This effect indicates that the autoinducer concentration is higher in the cells than in the medium flow. Thus we assume that the autoinducer accumulate either within or in the vicinity of the cells.

Strong variations in gene expression are found, both between colonies and between cells within colonies. Inter colony variations are probably caused by stochasticity in gene expression and differences in the cells which initiate the colonies. In contrast, the decrease of single cell fluorescence from the center towards the rim of colonies seems to originate from spatially heterogeneous autoinducer distributions. It is conceivable that the concentration of accumulated autoinducers is largest near the center of the colony.

In order to evaluate the cause of the apparently higher intracellular autoinducer concentration a two-compartment model will be developed. Furthermore, experiments with a higher optical resolution will allow quantifying the fluorescence levels of single cells within colonies. A stochastic reaction-diffusion model will account for both the origins of heterogeneity and the accumulation of AHL molecules.

Finally, we will experimentally address the influence of spatial limitations and the distance between cells on QS regulated gene expression. To this end, the sample environment will be refined: By using microfabricated chambers, differently sized compartments will be generated. From the gene expression dynamics of the same number of bacteria in different volumes we can draw conclusions on the impact of diffusion or accumulation in the vicinity of the bacteria. In addition, we will generate surfaces with regular, adhesive patterns, so that the cells are spaced by well-defined distances. Thus, it will be possible to measure the calling distance [103], meaning a maximal distance over which cells or cell aggregates can communicate.

The combination of a detailed quantitative knowledge of the early stage of biofilm formation with long time studies of the evolution of biofilms will help to understand the mechanisms controlling biofilm development. Particularly, we hope to understand the initiation of the process by which the cells differentiate into different subpopulations.

7 Outlook

In this work the gene expression dynamics of regulatory networks which respond to external stimuli by switching on the expression of certain genes were studied. In order to understand the fundamental relations between the network architecture and the resulting dynamics idealized conditions were used: On the one hand, the response to simple step functions of the inducer concentration was analyzed. On the other hand, when studying the arabinose system, the genes for arabinose degradation were deleted to avoid changes in the sugar concentration. Thus, a temporal stochastic effect was revealed and a model could be established that consistently explains gene expression dynamics produced by the native and a modified network.

In the future, we will continue our research on gene expression dynamics under natural conditions. In typical habitats, sugars are usually not constantly available, but rather appear in bursts. Furthermore, the sugar concentration is depleted by the bacteria. The necessity to change the inducer concentration over time has already been considered in the design of our experimental setup. This functionality was exploited in an experiment where we successfully recorded the response to inducer pulses (see chapter 5.4.4). Additional

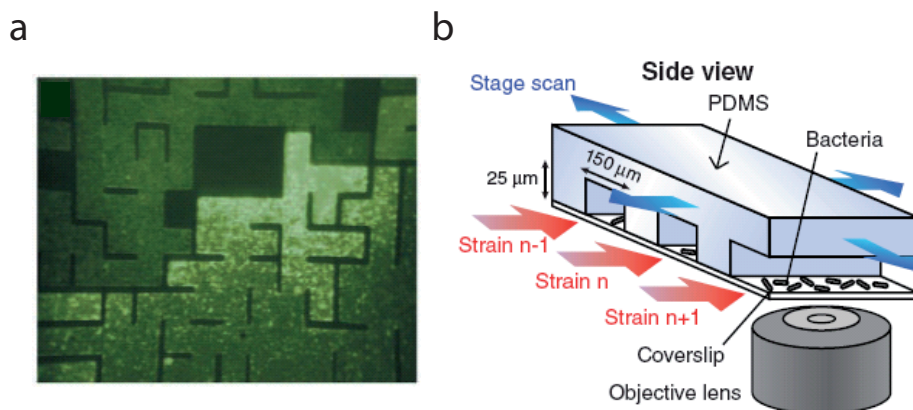


Figure 7.1: Microfluidic devices (a) GFP expressing bacterial cells (green) in a microfluidic maze. The dark lines are chamber boundaries. The device was used to probe the influence of topology on bacterial social interactions (From [107]). (b) Illustration of a microfluidic device in which up to 96 bacterial cultures can be monitored in parallel. Images of each channel are acquired using a scanning stage which automatically moves the objective across the device (From [108]). Reprinted with permission from AAAS.)

experiments of this type will provide insights into memory effects in inducible systems.

Our experimental approach could be extended by using custom made microfluidic devices [63]. These are usually manufactured from a transparent polymer that is shaped by using a structured silicon waver as mold and bonded to a glass slide which enables microscopic imaging. Using an appropriate mold it is possible to generate geometrically defined environments (see figure 7.1 a) such as compartments of different sizes. These can be used, for example, for more detailed studies on the influence of spatial limitations on quorum sensing regulation [109]. Furthermore, data acquisition can be highly parallelized in microfluidic devices, since one device can contain up to 100 identical microchannels. Data can be acquired automatically using a microscope equipped with a scanning stage which moves the device across the device. The channels are independent, so that different mutants and environmental conditions can be studied at the same time. Using this methodology, the noise characteristics of more than 1000 genes of *E.coli* were quantified recently [108] (see figure 7.1 b).

The regulatory networks we find in bacteria have been shaped by evolution. It is thus assumed that the combination of the network architecture and the parameter values is the optimal solution of a cost-benefit problem, resulting in a maximized growth rate. For a metabolic system such as the lactose or arabinose network, the cost is a decrease of the growth rate due to the production of the sugar uptake and degradation machinery. The benefit is the growth rate increase due to the energy which is released by the degradation of the specific sugar. Obviously, the benefit depends on the availability of the sugar, whereas there is at least a small, permanent cost for the maintenance of a low number of sugar transporters. The optimality of the protein output level of the lactose operon has been addressed by measuring cost and benefit at different lactose concentrations [11]. It turns out that there is an optimal protein expression level for each concentration. Experiments in which the cells are constantly exposed to this concentration for many generations reveal that the network evolves towards this optimum.

Presumably, not only the protein expression level, but also the response dynamics and the cell-to-cell variability have undergone optimization. To understand their shaping temporal variations of inducer concentrations need to be considered when cost and benefit are determined. For example, it has been shown that responses with significant differences between the cells are beneficial under temporally varying conditions [12] [13]. Thus, we assume that the heterogeneous timing we found in the arabinose system might be beneficial when the arabinose concentration varies. We will first address this hypothesis by theoretically comparing the influence of different types of fluctuations in the concentration, such as different durations of arabinose pulses, on the growth rate. Subsequently, we will compare population growth for conditions which are predicted to yield low and high benefit using our experimental setup in which we can vary the arabinose concentration temporally.

Finally, the entire functionality of a network, including expression levels, dynamics and noise characteristics, will be considered as a regulation strategy and addressed by

evolutionary game theory. In this context, not only the short term cost and benefit are considered, but also the evolutionary stability of a strategy. For example, a strategy with low cost and high benefit under normal conditions is probably not the optimal one when it is associated with a large probability that many cells die in a rare event. In general, it is conceivable that strategies with a high level of cell-to-cell variability are favored: These situations can be considered as mixed strategies, which have been shown to be optimal in many cases [32].

This combination of theoretical and experimental investigations will yield deeper insights into gene network shaping by evolutionary forces. In addition, natural principles for the optimal design of gene networks under different environmental conditions can be transferred when designing synthetic circuits for applications such as biofuel or drug production.

A Arabinose uptake

Arabinose uptake via AraE and AraFGH

In [76] the functionality of the arabinose transport systems AraE and AraFGH was analyzed. It turned out that the dependence of the uptake velocity V_{upt} on the external arabinose concentration A_{ext} of strains with both AraE and AraFGH follows a simple Michaelis-Menten function:

$$V_{upt} = V_{max} \cdot \frac{A_{ext}}{K_{upt} + A_{ext}} \quad (\text{A.1})$$

Here, K_{upt} is the Michaelis-Menten constant and V_{max} the maximal uptake velocity (given per dry cell mass). This finding is unexpected, as strains with either of the systems also exhibit a Michaelis-Menten type dependence. Thus, one would expect that the mutant with both systems is characterized by the sum of the single systems. The nature of this interaction remains unknown. The parameters for comparable strains, which are able of arabinose degradation, with only AraE, only AraFGH or both systems are given in table ??.

The uptake characteristics of the strain with AraE and AraFGH and of a strain with only AraE are very similar (see illustration of the parameter values in figure A.1). Thus AraE is used, when it is desired to use only one transport system in order to change its expression level. The constraint of the uptake velocity per protein to 200-2000 arabinose molecules/transporter/min was determined from the values of V_{upt} and is valid for strains with both AraE and AraFGH and only AraE.

transport system(s)	strain	K_{upt} [μM]	V_{upt} [$nmolmin^{-1}mg^{-1}$]
AraE + AraFGH	RS1wt	82.3	13.9
AraE	KD2	168	17.7
AraFGH	RS1E-thi	4.1	7.6

Table A.1: Transport characteristics of strains containing AraE or AraFGH or both systems. From [76].

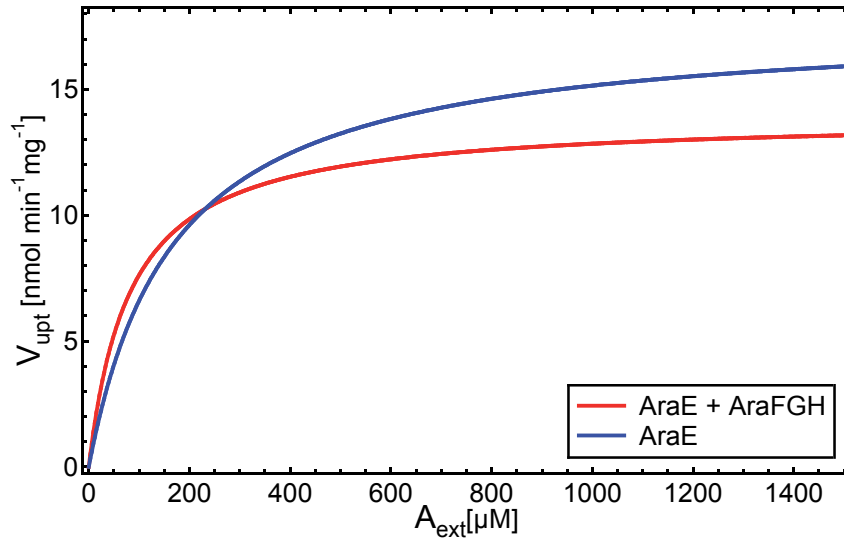


Figure A.1: Arabinose uptake characteristics of AraE and AraFGH

Gene expression dynamics in a strain with constitutive uptake protein expression

Figure A.2 shows gene expression dynamics of a strain in which the promoter Pcp18 controls AraE expression (strain Ara#3). AraFGH is deleted in this strain. Gene expression starts rapidly at all inducer concentrations and the protein synthesis rate decreases with decreasing arabinose concentration.

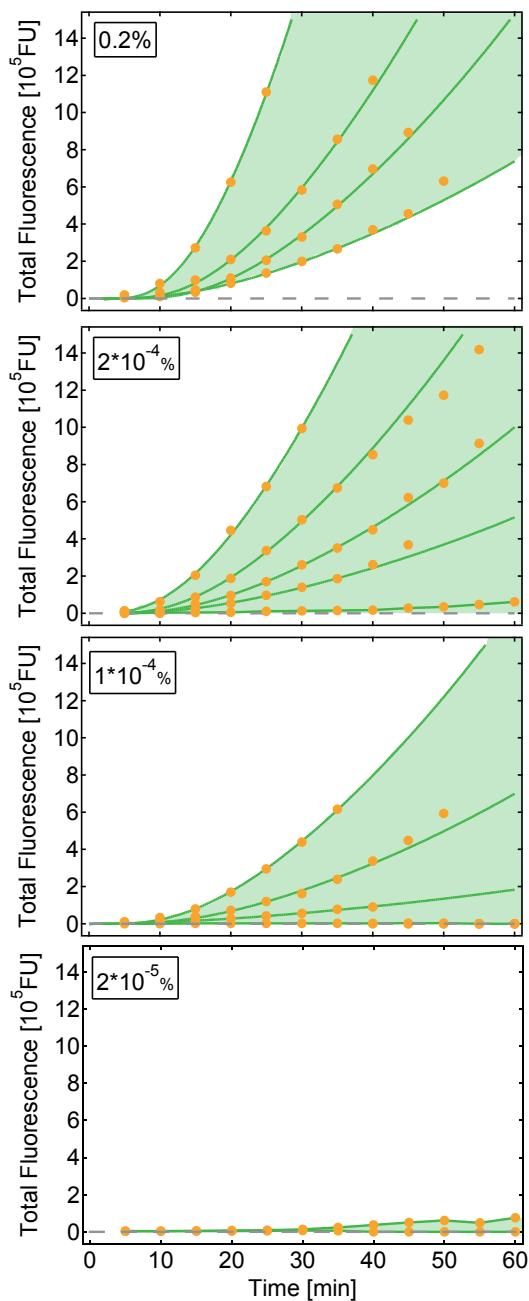


Figure A.2: Representative single cell induction kinetics at different arabinose concentrations. Transporter production is controlled by the constitutive promoter Pcp18. The analytical gene expression function (green) is fitted to the data points (yellow dots). As the $2 \cdot 10^{-5}$ % arabinose data cannot be fitted by this function data points are connected. To facilitate comparison the regions over which the traces are distributed are shaded.

B Experimental details

Measurement of the GFP maturation time

In strain *E.coli* LMG194 containing the reporter plasmid pBAD24/GFP, translation was blocked by the addition of 200 $\mu\text{g/ml}$ chloramphenicol, 30 min after the induction of *gfp*-expression with 0.2% arabinose. Fluorescence images were acquired every 3 to 5 min before and after inhibition. As this measurement was more sensitive the illumination was reduced and the EM Gain of the camera was used. Photobleaching could thus be neglected. Cellular fluorescence was determined by summing all pixel values above the background level for each bacterium. This method is qualitatively equal to the use of cell outlines, but can only be applied if the range of fluorescence values is limited and bacteria do not grow strongly. The resulting maturation time courses were fitted by an exponential function.

Analysis of the arabinose system

Reporter plasmid The reporter plasmid pBAD24/GFP ([45], see figure B.1) was created in the lab of Prof. K.Jung (LMU microbiology) based on plasmid pBAD24 [70]. The plasmid contains the gene for *GFPmut3* [54] under control of the promoter p_{BAD} and the gene encoding AraC under native control. The origin of replication pBR322ori controls the copy number while the *bla* gene confers resistance to ampicillin.

Time lapse experiments Bacteria were inoculated in M63 minimal medium from single colonies grown on LB agar plates and grown overnight (37°C, shaking at 300rpm). M63 minimal medium containing 0.2% or 0.5% glycerol as carbon source was used in all experiments. Overnight cultures were diluted 1:50 in prewarmed M63 medium and grown for 2 h (37°C, shaking at 300rpm; data in Chapter 5.2) or diluted 1:400 and grown for 4h (remaining data). 30 to 50 μl of an *E.coli* culture (OD \approx 0.1-0.2) are incubated for \approx 5min in a flow channel at 37°C. Some strains adhere to chambers pre-coated with PLL, while for other strains channels are coated with PLL on the day of the experiment (50 μl PLL, Biochrom AG, Berlin, 0.1mg/ml, incubate for 1 to 3 hours, rinse with water and medium). Following incubation, samples are rinsed three times with pre-warmed medium to wash away non adherent cells and improve adhesion of the cells remaining at the surface. Subsequently, the sample is either directly induced with medium containing the desired arabinose concentration (rinse 3 times) or transferred to the microscope and induced on

stage after the acquisition of initial frames. In the case of arabinose pulses, the sample is rinsed with arabinose-free medium ($\approx 500\mu\text{l}$) on stage. Bright-field and fluorescence images were acquired every 5 min with illumination times of 0.1 to 0.5 s.

Density dependent gene expression E coli MG1655/pBAD24-gfp was grown overnight in M63 medium. The overnight culture was diluted 1:100 into 15ml of fresh medium and grown for 4h. This culture was centrifuged and the pellet resuspended in fresh M63. For the inoculation of the OD 0.1 two or three rounds of centrifugation were necessary to obtain a sufficiently high density. The OD 0.01 and 0.001 cultures were inoculated from the same stock, which had been centrifuged one time. As several rounds of centrifugation resulted in different growth rates for differently dense cultures the samples were only centrifuged once in later experiments. The centrifuged stocks were diluted to the desired ODs into 25ml of prewarmed, fresh M63 medium, containing the desired concentration of arabinose at $t = 0\text{min}$. M63 medium always contained 0.5% glycerol as carbon source and ampicillin. Cultures were grown at 37°C and shaking at 300rpm throughout the experiment.

Every 45min samples were taken. At these times bulk fluorescence and OD were measured in a fluorescence plate reader. 10% Glycerol was added to the samples for flow cytometer analysis which were frozen in liquid nitrogen and subsequently stored at -80°C . For flow cytometer measurements the samples were thawed on ice, centrifuged and resuspended in PBS buffer.

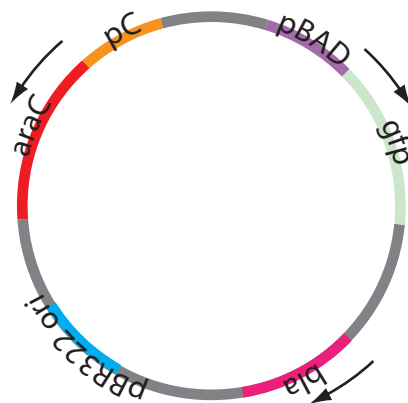


Figure B.1: Reporter plasmid pBAD24/GFP

Analysis of the *PPU* system

Reporter plasmid A RP4 type reporter plasmid (see figure B.2), containing the gene encoding *GFP_{asv}* [50] under control of the promoter *pl_{asB}* was created in the lab of Prof. L. Eberl (Zürich). The plasmid also contains the gene encoding the transcriptional activator LasR and a gene which provides resistance to kanamycin.

Preparation of OFF state Bacteria were inoculated from single colonies grown on LB-agar plates and grown overnight (30°C, shaking at 300rpm) in LB-medium. LB-medium was inoculated 1:100 with the overnight culture. This culture was incubated 3,5h (30°C, 300rpm). This step was repeated once. Then the culture was diluted again 1:100 in fresh pre-warmed LB-medium and incubated at 30°C (shaking with 300rpm) until OD600 = 0.4. The culture was distributed in a 1:1 mixture with glycerol (50%, sterile) in cryovials, vortexed and stored immediately at -80°C. To revive the cells, a cryovial was centrifuged at 3500rpm for 5 minutes, resuspended in FAB-medium [110] and incubated for 1h at 30°C with agitation (300rpm). FAB-medium contained 1mM sodium citrate as C-source.

Preparation for microscopy To prepare the cells for microscopy they were applied to channels of a poly-L-lysine-coated microfluidic chamber (μ -slide VI, Ibidi, Martinsried, Germany). The slide was then incubated at 30°C for 15 minutes and rinsed with FAB medium subsequently. When indicated, 10nM or 100nM N-(3-oxodecanoyl)-L-homoserine lactone (Sigma-Aldrich Chemie GmbH, Munich, Germany) was added. For each AHL concentration non-flow and flow conditions were measured in parallel in two microfluidic channels. Under flow, the channel was rinsed with 2ml/h of the FAB-medium containing the chosen AHL-concentration by a syringe pump (model infusion, TSEsystems, Bad Homburg, Germany). In the absence of flow, 150 μ l of FAB-medium containing the particular AHL-concentration were applied to the channel. To prevent evaporation, the reservoirs were sealed with mineral oil.

Fluorescence decrease Cultures were inoculated with different amounts (dilution factors 1:10, 1:20, 1:50, 1:100) from an overnight culture. After one hour of incubation the cells were seeded into microfluidic channels and rinsed with fresh medium. The mean fluores-

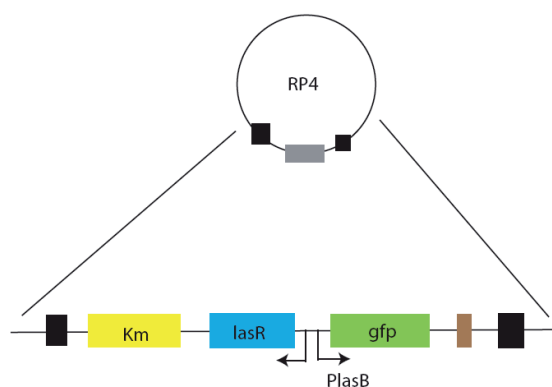


Figure B.2: Las reporter plasmid

cence of single cells was measured every 20min for 6h, by summing the intensity of all pixels belonging to a cell and dividing it by the number of pixels.

Time-lapse microscopy Bright-field and fluorescence images of several fields in one sample were acquired every 20 minutes, acquisition times of 0.5 seconds. The temperature in the sample environment was maintained at 30°C. Fluorescence decrease was analyzed at 100x magnification, while a 40x objective was used for all other experiments. The outline of single cells or single colonies was determined by thresholding the respective bright-field images. In the background-corrected fluorescence image, the sum over all pixel values within the outline (total fluorescence) was determined and divided by the number of pixels within the outline. This mean fluorescence value corresponds to the concentration of GFP molecules, given in arbitrary fluorescence units. For the analysis of fluorescence decrease single cells were analysed, which means that cells were regarded as separate objects after division. For fluorescence onset experiments the mean fluorescence of entire single colonies, originating from one "mother" cell, was determined.

Data fitting and simulations

Igor Pro 4.0 (WaveMetrics, Lake Oswego, OR) and Matlab (REF) were used to fit mathematical models to time series data. BioSys, a simulation tool developed by P. Hillenbrand (Group of Prof. U.Gerland, LMU) was used for stochastic simulations.

Bacterial strains

Escherichia coli strains

- LMG194 no arabinose degradation [70]
 LKB194 chromosomal integration of P_{BAD} -gfp, no arabinose degradation
 (based on *E. coli* MC4100 (E. coli Genetic Resources at Yale
 CGSC, The Coli Genetic Stock Center)
 created in the lab of Prof. K. Jung, LMU)
 MG1655 native arabinose system, thus capable of arabinose degradation
 laboratory strain with minimal genetic manipulation
 JW1889-3 no AraFGH, native control of AraE
 JW1889-5 no AraFGH, AraE expression controlled by p_{lac} (from *E. coli* MG1655)
 (JW1899-3,-5 are based on JW1889-1 (E. coli Genetic Resources
 at Yale CGSC, The Coli Genetic Stock Center)
 created in the lab of Prof. K. Jung, LMU)
 Ara#3 AraE expression controlled by Pcp18 promoter, no arabinose degradation
 (obtained from Prof.T.Hwas lab, UCSD)

Pseudomonas putida strain

- IsoF [104]

M63 Medium

For 1000 ml H_2O

KH_2PO_4	13,6 g
$(NH_4)_2SO_4$	2 g
$FeSO_4 \times 7 H_2O$	0.5 mg (1.8 μ l from 1 M stock)

adjust pH to 7.0 with KOH

autoclave

add:

1 ml $MgSO_4 \times H_2O$ from 1 M stock (sterile)

1 $\frac{\mu g}{ml}$ Thiamin from stock (1 $\frac{mg}{ml}$) (sterile)

0.2% Caseinhydrolysate from 10% stock (sterile)

0.2% Glycerol (sterile) as carbon source

Bibliography

- [1] D. Ro, E. M. Paradise, M. Ouellet, K. J. Fisher, K. L. Newman, J. M. Ndungu, K. A. Ho, R. A. Eachus, T. S. Ham, J. Kirby, M. C. Y. Chang, S. T. Withers, Y. Shiba, R. Sarpong, and J. D. Keasling. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, 440(7086):940–943, 2006.
- [2] D. Endy. Foundations for engineering biology. *Nature*, 438(7067):449–453, 2005.
- [3] H. Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, 2002.
- [4] U. Alon. Biological networks: the tinkerer as an engineer. *Science*, 301(5641):1866–1867, 2003.
- [5] A. Raj and A. van Oudenaarden. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–226, 2008.
- [6] D. A. Siegele and J. C. Hu. Gene expression from plasmids containing the arabid promoter at subsaturating inducer concentrations represents mixed populations. *Proc Natl Acad Sci U S A*, 94(15):8168–8172, 1997.
- [7] A. Novick and M. Weiner. Enzyme induction as an all-or-none phenomenon. *Proc Natl Acad Sci U S A*, 43(7):553–566, 1957.
- [8] C. J. Davidson and M. G. Surette. Individuality in bacteria. *Annu Rev Genet*, 42:253–268, 2008.
- [9] R. Y. Tsien. The green fluorescent protein. *Annu Rev Biochem*, 67:509–544, 1998.
- [10] J. C. W. Locke and M. B. Elowitz. Using movies to analyse gene circuit dynamics in single cells. *Nat Rev Microbiol*, 7(5):383–392, 2009.
- [11] E. Dekel and U. Alon. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436(7050):588–592, 2005.
- [12] M. Thattai and A. van Oudenaarden. Stochastic gene expression in fluctuating environments. *Genetics*, 167(1):523–530, 2004.
- [13] M. Acar, J. T. Mettetal, and A. van Oudenaarden. Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet*, 40(4):471–475, 2008.

-
- [14] E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman, and A. Van Oudenaarden. Multistability in the lactose utilization network of escherichia coli. *Nature*, 427(6976):737–740, 2004.
- [15] 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(19):7304–7309, 2006.
- [16] T. A. Carrier and J. D. Keasling. Investigating autocatalytic gene expression systems through mechanistic modeling. *J Theor Biol*, 201(1):25–36, 1999.
- [17] D. M. Wolf and A. P. Arkin. Motifs, modules and games in bacteria. *Curr Opin Microbiol*, 6(2):125–134, 2003.
- [18] U. Alon. Network motifs: theory and experimental approaches. *Nat Rev Genet*, 8(6):450–461, 2007.
- [19] J. E. Ferrell. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol*, 14(2):140–148, 2002.
- [20] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet*, 6(6):451–464, 2005.
- [21] A. Eldar and M. B. Elowitz. Functional roles for noise in genetic circuits. *Nature*, 467(7312):167–173, 2010.
- [22] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.
- [23] P. Guptasarma. Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of escherichia coli? *Bioessays*, 17(11):987–997, 1995.
- [24] H. H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A*, 94(3):814–819, 1997.
- [25] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox. Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(6):1025–1036, 2005.
- [26] 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(1):69–73, 2002.
- [27] 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(20):12795–12800, 2002.

- [28] M. Leisner, K. Stingl, E. Frey, and B. Maier. Stochastic switching to competence. *Curr Opin Microbiol*, 11(6):553–559, 2008.
- [29] N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler. Bacterial persistence as a phenotypic switch. *Science*, 305(5690):1622–1625, 2004.
- [30] I. Lestas, G. Vinnicombe, and J. Paulsson. Fundamental limits on the suppression of molecular fluctuations. *Nature*, 467(7312):174–178, 2010.
- [31] T. Kalisky, E. Dekel, and U. Alon. Cost-benefit theory and optimal design of gene regulation functions. *Phys Biol*, 4(4):229–245, 2007.
- [32] M. Nowak. *Evolutionary Dynamics: Exploring the equations of life*. Harvard University Press, 2006.
- [33] E. Kussell, R. Kishony, N. Q. Balaban, and S. Leibler. Bacterial persistence: a model of survival in changing environments. *Genetics*, 169(4):1807–1814, 2005.
- [34] T. Cagatay, M. Turcotte, M. B. Elowitz, J. Garcia-Ojalvo, and G. M. Süel. Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell*, 139(3):512–522, 2009.
- [35] J. Hasty, D. McMillen, F. Isaacs, and J. J. Collins. Computational studies of gene regulatory networks: in numero molecular biology. *Nat Rev Genet*, 2(4):268–279, 2001.
- [36] J. M. G. Vilar, C. C. Guet, and S. Leibler. Modeling network dynamics: the lac operon, a case study. *J Cell Biol*, 161(3):471–476, 2003.
- [37] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz. Gene regulation at the single-cell level. *Science*, 307(5717):1962–1965, 2005.
- [38] Y. Setty, A. E. Mayo, M. G. Surette, and U. Alon. Detailed map of a cis-regulatory input function. *Proc Natl Acad Sci U S A*, 100(13):7702–7707, 2003.
- [39] S. Kaplan, Anat B., A. Zaslaver, E. Dekel, and U. Alon. Diverse two-dimensional input functions control bacterial sugar genes. *Mol Cell*, 29(6):786–792, 2008.
- [40] R. Fernandez-Lopez, I. Del Campo, R. Ruiz, V. Lanza, L. Vielva, and F. de la Cruz. Numbers on the edges: a simplified and scalable method for quantifying the gene regulation function. *Bioessays*, 32(4):346–355, 2010.
- [41] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips. Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev*, 15(2):116–124, 2005.

-
- [42] J. Hasty, D. McMillen, and J. J. Collins. Engineered gene circuits. *Nature*, 420(6912):224–230, 2002.
- [43] J. Paulsson. Models of stochastic gene expression. *Physics of Life reviews*, 2:157–175, 2005.
- [44] O. G. Berg. A model for the statistical fluctuations of protein numbers in a microbial population. *J Theor Biol*, 71(4):587–603, 1978.
- [45] J. A. 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(4):2103–2115, 2008.
- [46] C. V. Rao, D. M. Wolf, and A. P Arkin. Control, exploitation and tolerance of intracellular noise. *Nature*, 420(6912):231–237, 2002.
- [47] D. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of physical chemistry*, 81:2340–2361, 1977.
- [48] R. F. Bruinsma. Physics of protein-dna interaction. *Physica A*, 313:211–237, 2002.
- [49] S. A. Kay K. F Sullivan, editor. *Green Fluorescent Proteins*, volume 58 of *Methods in Cell Biology*. Elsevier, 1998.
- [50] J. B. Andersen, C. Sternberg, L. K. Poulsen, S. P. Bjorn, M. Givskov, and S. Molin. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol*, 64(6):2240–2246, 1998.
- [51] G. Miesenböck, D. A. De Angelis, and J. E. Rothman. Visualizing secretion and synaptic transmission with ph-sensitive green fluorescent proteins. *Nature*, 394(6689):192–195, 1998.
- [52] G. H. Patterson, S. M. Knobel, W. D. Sharif, S. R. Kain, and D. W. Piston. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys J*, 73(5):2782–2790, 1997.
- [53] 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(26):12501–12504, 1994.
- [54] B. P. Cormack, R. H. Valdivia, and S. Falkow. Facs-optimized mutants of the green fluorescent protein (gfp). *Gene*, 173:33–38, 1996.
- [55] 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(1):87–90, 2002.

- [56] R. M. Morgan-Kiss, C. Wadler, and J. E. Cronan. Long-term and homogeneous regulation of the escherichia coli arabid promoter by use of a lactose transporter of relaxed specificity. *Proc Natl Acad Sci U S A*, 99(11):7373–7377, 2002.
- [57] X. S. Xie, P. J. Choi, G. Li, N. K. Lee, and G. Lia. Single-molecule approach to molecular biology in living bacterial cells. *Annu Rev Biophys*, 37:417–444, 2008.
- [58] J. Elf, G. Li, and X. S. Xie. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, 316(5828):1191–1194, 2007.
- [59] P. J. Choi, L. Cai, K. Frieda, and X. S. Xie. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*, 322(5900):442–446, 2008.
- [60] 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(5767):1600–1603, 2006.
- [61] G. M. Süel, J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*, 440(7083):545–550, 2006.
- [62] G. M. Süel, R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, and M. B. Elowitz. Tunability and noise dependence in differentiation dynamics. *Science*, 315(5819):1716–1719, 2007.
- [63] M. R. Bennett and J. Hasty. Microfluidic devices for measuring gene network dynamics in single cells. *Nat Rev Genet*, 10(9):628–638, 2009.
- [64] S. Youssef, S. Gude, and J. O. Rädler. Image analysis of live-cell time lapse microscopy. *in preparation*.
- [65] N. Rosenfeld, T. J. Perkins, U. Alon, M. B. Elowitz, and P. S. Swain. A fluctuation method to quantify in vivo fluorescence data. *Biophys J*, 91(2):759–766, 2006.
- [66] 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(2):175–181, 2007.
- [67] R. Schleif. Regulation of the l-arabinose operon of escherichia coli. *Trends Genet*, 16(12):559–565, 2000.
- [68] C. M. Johnson and R. F. Schleif. In vivo induction kinetics of the arabinose promoters in escherichia coli. *J Bacteriol*, 177(12):3438–3442, 1995.
- [69] R. Schleif. Induction of the l-arabinose operon. *J Mol Biol*, 46(1):197–199, 1969.

- [70] L. M. Guzman, D. Belin, M. J. Carson, and J. Beckwith. Tight regulation, modulation, and high-level expression by vectors containing the arabinose pbad promoter. *J Bacteriol*, 177(14):4121–4130, 1995.
- [71] A. Khlebnikov, O. Risa, T. Skaug, T. A. Carrier, and J. D. Keasling. Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J Bacteriol*, 182(24):7029–7034, 2000.
- [72] A. Khlebnikov, K. A. Datsenko, T. Skaug, B. L. Wanner, and J. D. Keasling. Homogeneous expression of the p(bad) promoter in escherichia coli by constitutive expression of the low-affinity high-capacity araE transporter. *Microbiology*, 147(Pt 12):3241–3247, 2001.
- [73] J. E. Cronan. A family of arabinose-inducible escherichia coli expression vectors having pbr322 copy control. *Plasmid*, 55(2):152–157, 2006.
- [74] S. Lin-Chao, W. T. Chen, and T. T. Wong. High copy number of the puc plasmid results from a rom/rop-suppressible point mutation in rna ii. *Mol Microbiol*, 6(22):3385–3393, 1992.
- [75] M. Thattai and A. van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci U S A*, 98(15):8614–8619, 2001.
- [76] K. R. Daruwalla, A. T. Paxton, and P. J. Henderson. Energization of the transport systems for arabinose and comparison with galactose transport in escherichia coli. *Biochem J*, 200(3):611–627, 1981.
- [77] M. H. Saier. Families of transmembrane sugar transport proteins. *Mol Microbiol*, 35(4):699–710, 2000.
- [78] J. Y. Liu, P. F. Miller, J. Willard, and E. R. Olson. Functional and biochemical characterization of escherichia coli sugar efflux transporters. *J Biol Chem*, 274(33):22977–22984, 1999.
- [79] J. Y. Liu, P. F. Miller, M. Gosink, and E. R. Olson. The identification of a new family of sugar efflux pumps in escherichia coli. *Mol Microbiol*, 31(6):1845–1851, 1999.
- [80] C. P. Novotny and E. Englesberg. The l-arabinose permease system in escherichia coli b/r. *Biochim Biophys Acta*, 117(1):217–230, 1966.
- [81] S. Carole, S. Pichoff, and J.P. Bouche. Escherichia coli gene ydeA encodes a major facilitator pump which exports L-arabinose and isopropyl-beta-D-thiogalactopyranoside. *JOURNAL OF BACTERIOLOGY*, 181(16):5123–5125, 1999.

- [82] A. Danchin. Cells need safety valves. *Bioessays*, 31(7):769–773, 2009.
- [83] D. Kolodrubetz and R. Schleif. Regulation of the l-arabinose transport operons in escherichia coli. *J Mol Biol*, 151(2):215–227, 1981.
- [84] J. A. Bernstein, A. B. Khodursky, P. Lin, S. Lin-Chao, and S. N. Cohen. Global analysis of mrna decay and abundance in escherichia coli at single-gene resolution using two-color fluorescent dna microarrays. *Proc Natl Acad Sci U S A*, 99(15):9697–9702, 2002.
- [85] C. D. Smolke, T. A. Carrier, and J. D. Keasling. Coordinated, differential expression of two genes through directed mrna cleavage and stabilization by secondary structures. *Appl Environ Microbiol*, 66(12):5399–5405, 2000.
- [86] P. Wong, S. Gladney, and J. D. Keasling. Mathematical model of the lac operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotechnol Prog*, 13(2):132–143, 1997.
- [87] M. Santillán and M. C. Mackey. Influence of catabolite repression and inducer exclusion on the bistable behavior of the lac operon. *Biophys J*, 86(3):1282–1292, 2004.
- [88] M. G. Surette, M. B. Miller, and B. L. Bassler. Quorum sensing in escherichia coli, salmonella typhimurium, and vibrio harveyi: a new family of genes responsible for autoinducer production. *Proc Natl Acad Sci U S A*, 96(4):1639–1644, 1999.
- [89] J. B. Kaper P. D. Karp F. C. Neidhardt T. Nyström J. M. Slauch C. L. Squires A. Böck, R. Curtiss III and D. Ussery (ed.), editors. *EcoSalEscherichia coli and Salmonella: Cellular and Molecular Biology*. <http://www.ecosal.org>. ASM Press, Washington, DC.
- [90] P. L. Irwin, L. T. Nguyen, G. C. Paoli, and C. Chen. Evidence for a bimodal distribution of escherichia coli doubling times below a threshold initial cell concentration. *BMC Microbiol*, 10:207, 2010.
- [91] S. Klumpp, Z. Zhang, and T. Hwa. Growth rate-dependent global effects on gene expression in bacteria. *Cell*, 139(7):1366–1375, 2009.
- [92] E. Kussell and S. Leibler. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309(5743):2075–2078, 2005.
- [93] J. Monod. From enzymatic adaptation to allosteric transition. *Science*, 154:475, 1966.
- [94] M. Thattai and B. I. Shraiman. Metabolic switching in the sugar phosphotransferase system of escherichia coli. *Biophys J*, 85(2):744–754, 2003.

- [95] D.López, H. Vlamakis, and R. Kolter. Biofilms. *Cold Spring Harb Perspect Biol*, 2(7):a000398, 2010.
- [96] H. Flemming and J. Wingender. The biofilm matrix. *Nat Rev Microbiol*, 8(9):623–633, 2010.
- [97] J. S. Webb, M. Givskov, and S. Kjelleberg. Bacterial biofilms: prokaryotic adventures in multicellularity. *Curr Opin Microbiol*, 6(6):578–585, 2003.
- [98] H. Vlamakis, C. Aguilar, R. Losick, and R. Kolter. Control of cell fate by the formation of an architecturally complex bacterial community. *Genes Dev*, 22(7):945–953, 2008.
- [99] W. Ng and B. L. Bassler. Bacterial quorum-sensing network architectures. *Annu Rev Genet*, 43:197–222, 2009.
- [100] W. C. Fuqua, S. C. Winans, and E. P. Greenberg. Quorum sensing in bacteria: the luxR-luxI family of cell density-responsive transcriptional regulators. *J Bacteriol*, 176(2):269–275, 1994.
- [101] B. A. Hense, C. Kuttler, J. Müller, M. Rothballer, A. Hartmann, and J. Kreft. Does efficiency sensing unify diffusion and quorum sensing? *Nat Rev Microbiol*, 5(3):230–239, 2007.
- [102] W. Timp, U. Mirsaidov, P. Matsudaira, and G. Timp. Jamming prokaryotic cell-to-cell communications in a model biofilm. *Lab Chip*, 9(7):925–934, 2009.
- [103] S. Gantner, M. Schmid, C. Dürr, R. Schuhegger, A. Steidle, P. Hutzler, C. Langebartels, L. Eberl, A. Hartmann, and F. B. Dazzo. In situ quantitation of the spatial scale of calling distances and population density-independent n-acylhomoserine lactone-mediated communication by rhizobacteria colonized on plant roots. *FEMS Microbiol Ecol*, 56(2):188–194, 2006.
- [104] A. Steidle, M. Allesen-Holm, K. Riedel, G. Berg, M. Givskov, S. Molin, and L. Eberl. Identification and characterization of an n-acylhomoserine lactone-dependent quorum-sensing system in *Pseudomonas putida* strain isof. *Appl Environ Microbiol*, 68(12):6371–6382, 2002.
- [105] A. Fekete, C. Kuttler, M. Rothballer, B. A. Hense, D. Fischer, K. Buddrus-Schiemann, M. Lucio, J. Müller, P. Schmitt-Kopplin, and A. Hartmann. Dynamic regulation of n-acyl-homoserine lactone production and degradation in *Pseudomonas putida* isof. *FEMS Microbiol Ecol*, 72(1):22–34, 2010.

-
- [106] K. Riedel, M. Hentzer, O. Geisenberger, B. Huber, A. Steidle, H. Wu, N. Høiby, M. Givskov, S. Molin, and L. Eberl. N-acylhomoserine-lactone-mediated communication between *Pseudomonas aeruginosa* and *Burkholderia cepacia* in mixed biofilms. *Microbiology*, 147(Pt 12):3249–3262, 2001.
- [107] S. Park, P. M. Wolanin, E. A. Yuzbashyan, H. Lin, N. C. Darnton, J. B. Stock, P. Silberzan, and R. Austin. Influence of topology on bacterial social interaction. *Proc Natl Acad Sci U S A*, 100(24):13910–13915, 2003.
- [108] Y. Taniguchi, P. J. Choi, G. 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(5991):533–538, 2010.
- [109] T. Danino, O. Mondragón-Palomino, L. Tsimring, and J. Hasty. A synchronized quorum of genetic clocks. *Nature*, 463(7279):326–330, 2010.
- [110] A. Heydorn, A. T. Nielsen, M. Hentzer, C. Sternberg, M. Givskov, B. K. Ersbøll, and S. Molin. Quantification of biofilm structures by the novel computer program comstat. *Microbiology*, 146 (Pt 10):2395–2407, 2000.

Vielen Dank an...

Prof. Dr. Joachim Rädler der mir die Möglichkeit gegeben hat mich in dieses spannende Thema zu vertiefen und mich dabei betreut und gefördert hat. Besonder bedanken möchte ich mich für die Gelegenheit meine wissenschaftliche Neugier zu entfalten und selbstständig zu arbeiten ohne dabei das große Ganze aus den Augen zu verlieren.

Prof. Dr. Ulrich Gerland, dessen Begleitung dieser Arbeit von theoretischer Seite enorm wichtig war.

Prof. Dr. Kirsten Jung und **Prof. Dr. Karin Schnetz** für viele hilfreiche Diskussionen zu biologischen Fragen und die Herstellung von zahlreichen Mutanten.

Prof. Dr. Christina Kuttler, **Dr. Burkhard Hense**, **Prof. Dr. Johannes Müller**, sowie **Prof. Dr. Leo Eberl** und **Dr. Claudia Aguilar** die entscheidend zur Untersuchung der Quorum Sensing regulierten Genexpression beigetragen haben.

Georg Fritz für die tolle Zusammenarbeit, ohne die diese Arbeit nicht möglich gewesen wäre. Mit Kollegen wie Dir macht Wissenschaft besonders viel Spass.

Dr. Madeleine Leisner, die mir bei dieser Arbeit mit Rat und Tat zur Seite stand, nicht selten mit der zusätzlichen Unterstützung eines Kaffee Latte.

Andrea Meyer und **Sonja Westermeyer**, die im Rahmen ihrer Diplomarbeiten meine Arbeit sehr bereichert haben. Danke auch an die Bachelor- und Werkstudenten **Delia Brick**, **Raphael Andre**, **Daniel Schiffels** und **David Grodzki**.

Hanna Engelke, **Simon Youssef**, **Martin Hennig**, **Tobias Stögbauer** und **Daniel Hönig**, meine langjährige Bürogemeinschaft, die viele Gelegenheiten für fachliche und auch ein paar andere Diskussionen bot.

alle Mitglieder des Lehrstuhl Rädler für ein tolles Arbeitsklima. Danke an **Susi Kempter** und **Gerlinde Schwake** für die Unterstützung im Labor und an **Margarete Meixner** für die Hilfe bei zahlreichen organisatorischen und büokratischen Fragen.

Meine **Freunde**, von denen jeder mein Leben auf seine ganze eigene Art bereichert.

Meine **Familie**, besonders an **Mama** und **Papa**, die immer Zeit und ein offenes Ohr für mich haben.

Uwe und **Julius** für das beste Leben, das man sich vorstellen kann.

Curriculum vitae

Personal data

Judith Megerle, nee Leierseder
born 1982 in Regensburg

Education

1992 - 2000 Von-Müller-Gymnasium Regensburg
06/2002 Abitur
2000 - 2006 Studies in physics at Ludwig-Maximilians-Universität München,
Scholarship by the bavarian state government
2002 - 2003 ERASMUS student at Lund University (Sweden)
2005 - 2006 Diploma thesis in the group of Prof. J.O. Rädler,
Ludwig-Maximilians-Universität München
05/2006 Diploma in physics
2006 - 2011 PhD student in the group of Prof. J.O. Rädler,
Ludwig-Maximilians-Universität München
since 2007 Member of the international doctorate program nano-biotechnology

Publication

J. A. 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(4):21032115, 2008

The following papers appeared under my maiden name J. A. Leierseder

M. B. Hochrein, J. A. Leierseder, L. Golubovic, and J. O. Rädler. DNA molecules on periodically microstructured lipid membranes: localization and coil stretching. *Physical Review E* 75, 021901, 2007

M. B. Hochrein, J. A. Leierseder, L. Golubovic, and J. O. Rädler. DNA localization and stretching on periodically microstructured lipid membranes. *Physical Review Letters* 96, 038103, 2006