Quantitative study of the effects of low DnaA concentrations in Escherichia coli, using the uhp pathway as an inducible expression system

Bernard Chelli Ponce de Leon

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **Physique pour les sciences du vivant**

Arrêté ministériel : 25 mai 2016

Présentée par

« Bernard Armand CHELLI PONCE DE LEON »

Thèse dirigée par **Irina Mihalcescu, Professeur à l'Université Grenoble Alpes**, et
codirigée par **Johanes Geiselmann, Professeur à l'Université Grenoble Alpes**

préparée au sein du **Laboratoire Interdisciplinaire de Physique**
dans **l'École Doctorale de Physique**

Étude quantitative de basses concentrations de DnaA chez *Escherichia coli,* en utilisant le système d'expression uhp

Thèse soutenue publiquement le **25 avril 2017**
devant le jury composé de :

Monsieur Hidde DE JONG
Directeur de recherche, Inria, Président du Jury
Madame Bianca SCLAVI
Chargée de recherche, Laboratoire de Biologie et de Pharmacologie Appliquée, Rapporteur
Monsieur David BENSIMON
Directeur de recherche, Laboratoire de Physique Statistique, Rapporteur
Madame Meriem ELKAROUI
Directeur de recherche, University of Edinburgh, Examinateur
Monsieur Thierry VERNET
Directeur de recherche, Institut de Biologie Structurale, Examinateur

Acknowledgments

I wish to start this manuscript by thanking the people who helped me during my thesis. I tried to be exhaustive but rest assured, if I was to realize I forgot someone, he will become the one person I would never forget ever again.

I will begin by expressing my gratitude to both my PhD supervisors, Irina Mihalcescu and Hans Geiselmann. They played a major role in me choosing this subject for my thesis. They guided me and patiently imparted on me the skills necessary to do quality scientific research. They weren't discouraged when we had experimental failures and did a very in depth revision of this manuscript.

Next i would like to thank both my parents, Leticia Ponce de León García and Alain Chelli. It is because of them that I chose to follow a scientific career, they have always supported me and helped me even when times were rough. I greatly appreciated their input and their thoughts on my work, weather it was on the biological or statistical aspects and thank them for everything they have done.

I also thank Emma Goux, who has stayed by my side for many months. Her support was key at times and it would have been challenging to meet the deadlines without her skills in drawing figures.

I was fortunate enough to be surrounded by very competent individuals in my research group. The many technicians from the team: Ludowic Lancelot, Mathilde Gateau, Corinne Pinnel, Natale Scaramozzino, imparted on me the many techniques needed for the experimental part of this thesis. They also made no small number of experiments in order to have sufficient data for pertinent interpretations. In particular Ludowic Lancelot did a tremendous effort to finish one of the cell strains needed for this project (The Chelli-Lancelot strain). I much enjoyed talking to them on a daily basis on a number of topics which ranged from science to cooking to football. I also had the pleasure of learning from Cindy Gomez Balderas and Caroline Ranquet. I feel very grateful to the latter specially for she allowed me to user her cloning protocol to build my cell strains. I much enjoyed working with Manon Lazaro, who contributed with both experiments and some modeling to this work. Although they were not directly involved, the many members of the BIOP team made useful remarks and made many suggestions during the course of my thesis. I am very grateful to them and enjoyed being in the same research group as them.

I also thank the PhD students with whom I shared joy, frustration, anger and laughter. From my team Sebastien Boyer, Mathieu Hemery, Nils Giordano, as well as all the other PhD students from the lab, who are so many that I dare not single each of them out for fear of forgetting someone. We spent much time together, weather it was in the lab or outside, bowling, doing

a laser game or just enjoying a beer. It is my privilege to call myself your friend and to have experienced so much together.

I would also like to thank Antoine Delon, whose counsel allowed me to consider new possibilities after defending my thesis. It is also very important to me to thank Sylvie Zanier and Sylvie Spagnoli, who allowed me to teach for 2 years and gave me many pointers on how to properly instruct students. I also thank Stephan Lacour who gave me advice not only on my thesis, but also on how to tackle the one biology course I taught.

I thank as well the RESET project, who financed this thesis along all of its partners. In particular i would like to thank Hidde de Jong, who spearheads RESET and gave me advice on many occasions.

Finally i would like to thank all the people from the LIPhy lab, from the cleaning personnel to the secretaries to the research staff. I spent the past three and a half years along side all of them and very much enjoyed doing so.

Contents

Part I

Introduction

During the past 50 years, technological advances have greatly facilitated the investigation of biological objects by physical methods. Furthermore, these technological developments in biotechnology and bio-engineering, such as fast sequencing or the use of fluorescent reporter proteins, have increased the need for precise models to help direct research and understand the observed biological phenomena. Improved measurement tools, such as time lapse microscopy and microfluidics, allow observing not only the collective behavior of a population of cells, but also the activity of individual cells. Such data have revealed significant differences between the behavior of a single cell compared to the average behavior of the population. The combination of all these methods, tools and techniques have transformed the study of living matter by exponentially increasing the amount of precise information that can be obtained about living organisms. In other words we can now build models that allow us to quantify the different elements of our biological system and predict its behavior.

Quantification of the different properties of living organisms nevertheless remains challenging. Many cellular components are present at intermediate or low copy numbers, making it difficult to distinguish them from the background even with the use of fluorescent reporters. Population studies help to increase the sensibility of these measurements but loose all information about the individuals and the stochastic properties of these systems. Yet, precise quantification of biological parameters and elements is essential for building correct models as well as understanding how and why individual cells behave the way they do in diverse circumstances. Mathematical modeling is particularly important when low copy numbers of proteins are in play because stochastic phenomena can be counter-intuitive. One vital process in this class is the initiation of DNA replication. In the bacterium *Escherichia coli*, the central regulator of this process is the DnaA protein, which binds to a small number of specific sites on the chromosome to start DNA replication. Much of the work in this thesis focuses on the control of DNA replication in *E. coli*.

The mechanism of DNA replication is complex and highly regulated, involving many proteins. Since DNA replication is one of the most fundamental properties of living organisms, understanding the details of the underlying molecular mechanism is of great general interest. Single cell organisms, such as *E. coli* in particular, have a DNA replication mechanism that is based on counting a relatively small number of the initiation protein DnaA in the cell and therefore lacks the precision of replication control found in higher organisms. We will address this problem in the present work by a combination of biological engineering and physical measurement techniques. This will also allow us to study other biological phenomena that appear when DNA replication is artificially blocked.

Chapter 1

The cell

1.1 *Escherichia coli*: an overview

Before going deeper into the objective of this thesis, we will start by describing the system that we will be studying, *Escherichia coli*. This bacterium is probably the most studied organism on Earth and our laboratory has ample previous experience with this organism. Since it is relatively easy to modify the genome of *E. coli*, this model organism is well suited for scientific studies. Furthermore, the bacterium grows relatively fast, with doubling times as short as twenty minutes. Being a prokaryote, most regulatory systems are simpler than in eukaryotic cells. An extensive knowledge base about this organism facilitates the study of the biophysical mechanisms of any particular regulatory system.

1.1.1 General Information

Escherichia coli is a bacteria commonly found in the lower intestine of warm-blooded animals. It has a rod-shaped bacterium with a diameter of about 1 μm and a typical length of 1 to 2 μm. However, the size may vary depending on the growth conditions. A single flagellum allows the organism to actively move, for example, toward a nutrient source. The bacterium is a facultative anaerobe, meaning that it can grow in liquid medium in the absence and presence of oxygen. As a Gram-negative bacterium, *E. coli* has two membranes surrounding the cytoplasm, which holds most of the proteins and DNA. The outer membrane is negatively charged, essentially due to phosphate group of the phospholipids. Like most bacteria, there are many different types of *Escherichia coli* with slightly different genomes. Some strains are pathogenic, provoking stomach sickness or cause urinary tract infections. Others, such as the one used in this work, are benign. They do not provoke infections and live inside the host organism as part of the normal intestine flora. Depending on the growth medium, *E. coli* has a doubling time typically ranging from 20 minutes in rich medium to several hours in minimal media. These values can change depending on temperature and the amount of oxygen available to the cells during growth. Now that we have described the form of our model organism, and summarized its characteristics, let us quickly review the major component that make it tick.

1.1.2 Inner workings

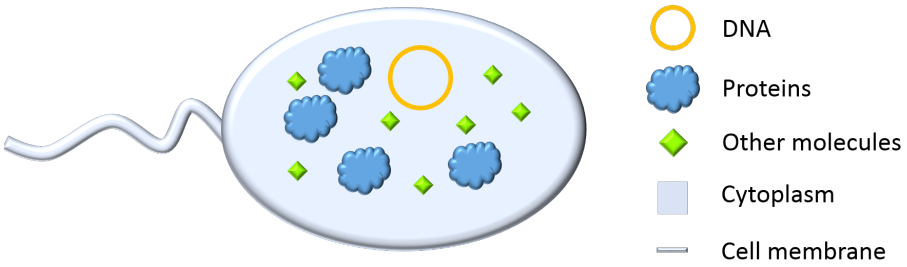

Figure 1.1: *E. coli* cell. *E.coli is a rod shaped bacterium (shape not represented in the generic representation) that possesses a negatively charged inner and outer membrane enclosing the cytoplasm. The DNA and most proteins and metabolites are located in the cytoplasm. A flagellum confers motility to the bacterium.*
.

Escherichia coli possesses two main compartments: the cell membrane and the cytoplasm. As all prokaryotes, *E. coli* does not possess any membrane-delimited compartments in the cytosol, such as the nucleus of eukaryotes, specifically designed to contain the genome of the organism. The inner and outer cell membranes are barriers that protect the cell. They allow for specific molecules to be exchanged with the outside medium. In general, specific channels or transporter proteins ensure the selectivity of the transport. This is, for example, the case for the transporter of glucose-6-phosphate (g6p), the membrane-spanning protein UhpT [1]. Other membrane-spanning proteins sense the presence of a molecule in the medium and transmit this signal to the inside of the cell via a conformational change. An example of this would the UhpC protein, which detects the presence of glucose-6-phosphate in the medium, but does not transport this molecule across the membrane [1]. Nonetheless, the majority of proteins are soluble and therefore located in the cytoplasm. In this thesis work, we will particularly study on such protein, DnaA, the initiator of DNA replication. The general structure of the *E. coli* cell is summarized in figure 1.1.

So far, we have presented a general overview of *Escherichia coli*. The interaction of the different cellular components with DNA, thereby controlling gene expression, make *E. coli* an interesting stochastic system. Since we are specifically interested in this aspect of the biological system, we will now describe in more detail the three main actors of gene expression: DNA, mRNA and proteins.

1.2 Gene Expression

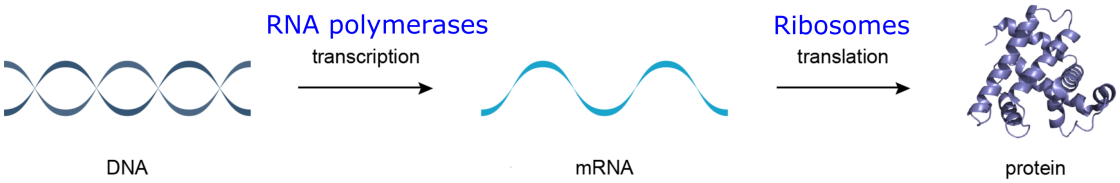

Figure 1.2: Central dogma. *DNA is transcribed by RNA polymerase into mRNA which in turn is translated by ribosomes into proteins*

Gene expression is the process of producing proteins, starting from the genetic code stored in DNA. This two-step process, also called the central dogma of biology, is summarized in Figure 1.2. The first step, reading the information in the DNA, is called transcription. RNA polymerase produces a single-stranded copy of the information contained in the DNA. The resulting messenger RNA, mRNA, is then translated into proteins by ribosomes. Proteins constitute the largest fraction of the cell mass, representing up to 52.4% of the dry weight of *E. coli*. The gene expression machinery itself, i.e., RNA polymerase and ribosomes, are made of proteins and RNA (in the case of ribosomes).

Although the process of gene expression may seem straightforward, each step of the process is highly regulated by proteins or RNA. These complex regulatory interactions introduce highly non-linear relationships between environmental signals and the response of the cell. Furthermore, since the number of particular proteins in the cell can be very small (on the order of tens of molecules), stochastic effects add to the difficulty of non-linear models. Therefore, in order to predict the behavior of the organism, we need good mathematical models and accurate parameters.

To show more explicitly how gene expression works, I will start from the beginning with a more in depth description of DNA, and then move on to the other actors of gene expression.

1.2.1 Transcription

Deoxyribonucleic acid

Deoxyribonucleic acid, or DNA, stores all the information on the different components of the cell, that is to say it contains the information necessary to express all proteins (and RNAs) in the cell. DNA is located in the cytoplasm in the form of a supercoiled double helix. Each helix is comprises a succession of 4 possible nucleic bases; adenine (A), guanine (G), cytosine (C) or thymine (T), which are joined by a sugar-phosphate backbone. Nucleotide bases pair up with a specific complementary base of the other helix, A with T and C with G, forming hydrogen bonds with the corresponding nucleic base of the other helix as seen in figure 1.3. *E. coli* contains a single, circular chromosome of 4.6 million base pairs. The complementarity of the two strands of DNA also provides the capacity to repair a damaged strand. This precaution is necessary since DNA is THE database of all functions of the cell and each strand acts as a possible backup of the other.

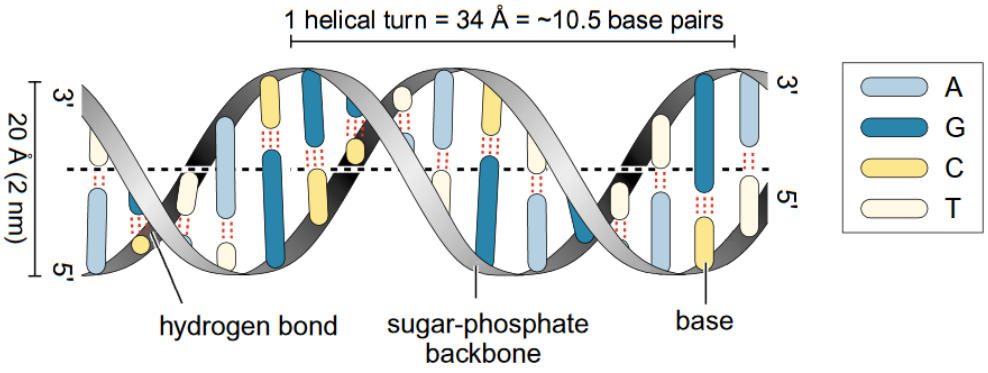

Figure 1.3: DNA structure. *DNA is formed by a double helix of nucleic acids linked up by a sugar-phosphate backbone. Nucleotides of one strand form hydrogen bonds with the complementary base of the other strand (A-T, C-G). The double helix is supercoiled and forms a single, circular chromosome in E. coli of 4.6 million base pairs.* Image taken from "Molecular Biology of the Gene", 5^{th} edition, Chapter 6.[2]

Transcription

The information contained in DNA is transcribed by RNA polymerase, a complex molecular machinery formed of five proteins; two α subunits, encoded by *rpoA*, the two large, catalytic subunits, β and β', coded by the genes *rpoB* and *rpoC*, and a small, dispensable subunit, ω, coded by the gene *rpoZ*. RNA polymerase recognizes and initiates transcription at specific sequences on the chromosome called promoters. After binding to the promoter, RNA polymerase "opens" the DNA for about one helical turn and starts transcription. The production of the single-stranded ribonucleic acid (RNA) stops at specific signals in the RNA (therefore coded by the DNA), called terminators. Transcription from a particular promoter can span several genes, called an operon. The regulation of these functional units ensure the expression of the selected part of the genetic information.

Promoters

The promoter sequences determine the rate of transcription by modulating the affinity of RNA polymerase for the DNA and/or the rate of opening of the DNA. Transcription rates can be slow, with only some tens of mRNAs transcribed per generation, or very fast, with hundreds or even thousands of mRNAs produced per generation. The maximal promoter strength is limited by the maximal rate at which RNA polymerase "escapes" from the promoter, estimated to about one transcript per ten seconds. The binding affinity of RNA polymerase for the promoter is determined to a large part by the sequence of two hexamers, located around the positions -10 and -35 with respect to the transcription start site.

Other molecules, mostly proteins, can accelerate or slow any of the steps leading to a transcribing ternary complex composed of DNA, RNA polymerase, and mRNA. These regulators modulate the promoter strength by recognizing specific sequences in the promoter region and changing the affinity of RNA polymerase for the promoter. These molecules are called transcription factors and can act as activators, if they increase the transcription rate, or as repressors,

if they reduce it instead. Some promoters are hardly transcribed at all without the aid of such molecules. Transcription factors can be global or specific. Global transcription factors affect a very wide variety of genes. A prime example are sigma factors, a necessary, additional subunit of RNA polymerase that provides the interaction with the recognition sequences at -10 and -35. *E. coli* possesses seven different types of sigma factors that are expressed in particular physiological situations. The expression of, for example the "heat-shock sigma factor, will for a holoenzyme with RNA polymerase, which will then recognize the promoters of a set of genes necessary for dealing with this specific condition. The house-keeping sigma factor of *E. coli* is expressed in all conditions and transcribes the majority of genes of this bacterium. Because this protein possesses a molecular weight of 70 kD, it is called σ^{70}. The genes studied in this thesis are all transcribed by an RNA polymerase containing the σ^{70} subunit. Specific transcription factors target a small number of genes and are usually limited to their specific pathways.

Regulation of transcription by transcription factors

An important class of transcription factors respond to external stimuli. The conditions detected by these transcription factors can be very divers, ranging from temperature sensing to the detection of the presence/absence of certain molecules in the growth medium. Promoters that are controlled by such a transcription factor are referred to as inducible promoters since transcription can be regulated by adding (or removing) the molecule, called "inducer", to the growth medium. Inducible promoters can be regulated by a repressor, such as P_{lac} or by an activator, such as P_{uhpT}.

P_{lac}, the lactose promoter in *E. coli*, is repressed by the lac repressor, coded by the gene *lacI*. In the absence of inducer (allolactose or a molecule with a similar structure), the repressor binds to operator sites overlapping P_{lac} and thereby prevents binding of RNA polymerase to the promoter. In the presence of an inducer, the repressor changes conformation, resulting in a much lower affinity for the specific DNA binding sites [3]. Since LacI and RNA polymerases compete with each other for binding to overlapping sites, the transcription rate of P_{lac} is a monotonous function of the concentration of inducer in the growth medium. In the case of P_{lac}, the most common choice of an inducer is Isopropyl β-D-1-thiogalactopyranoside (IPTG) instead of allolactose since the latter is metabolized by *E. coli*. The mechanism of regulation of the *lac* promoter is summarized in figure 1.4 A.

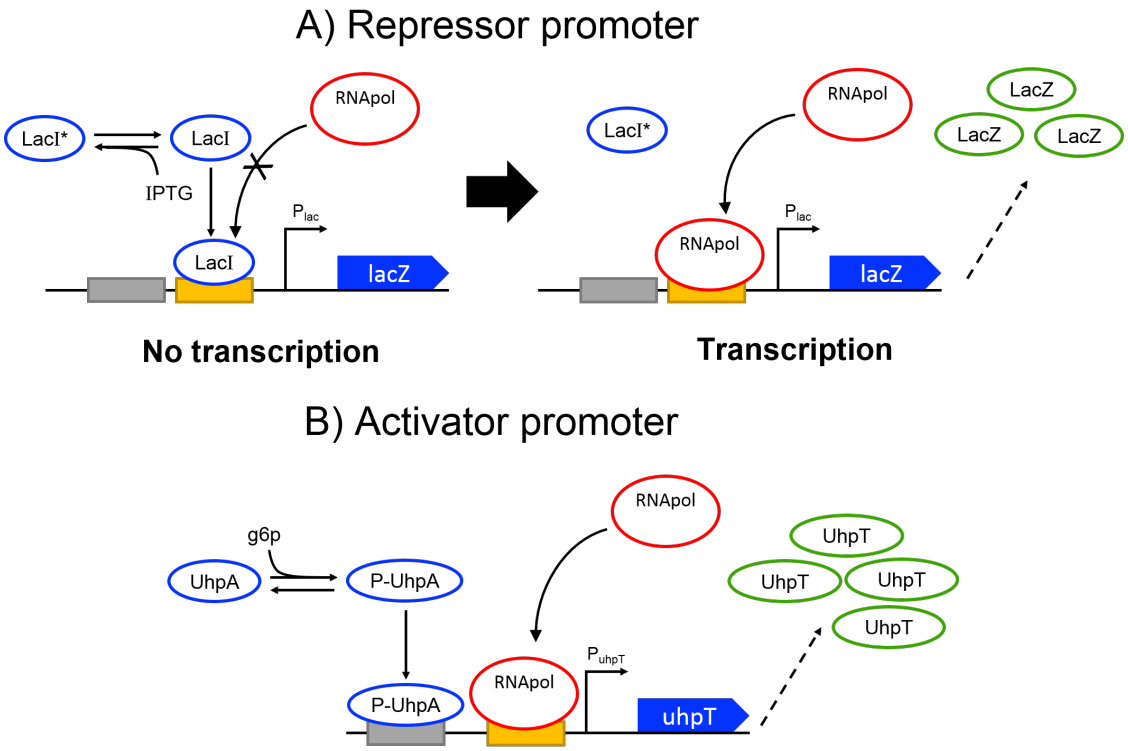

Figure 1.4: Inducible promoters. A) Promoter regulated by a repressor. *There is a competition between RNA poloymerase and lac repressor, LacI, for binding to the lac promoter. When LacI is bound to its operator (orange), it prevents binding of RNA polymerase to the promoter. In the presence of IPTG, this inducer molecule will bind to LacI, change the conformation of the protein, thereby lowering the affinity of LacI for the binding sites overlapping the promoter and liberating this latter for the binding of RNA polymerase.* **B) Promoter regulated by an activator.** *RNA polymerase binds to the uhpT promoter with low affinity. The phospohorylated activator protein UhpA∼P binds upstream and established protein-protein interactions with RNA polymerase. This cooperativity of binding increases the occupancy of the promoter by RNA polymerase, and therefore the transcription rate.*

Contrary to repressors, transcriptional activators do not reduce transcription rates, but favor the recruitment of RNA polymerase and therefore increase promoter strength. P_{uhpT} is the promoter of the *uhpT* gene, which regulates the uptake of glucose-6-phosphate (g6p) and other phosphate sugars. P_{uhpT} is activated by the phosphorylated version of the UhpA transcription factor (UhpA∼P) and is part of the *uhp* transcriptional network [1]. Extracellular g6p triggers the phosphorylation of UhpA, resulting in the activation of P_{uhpT}. g6p is therefore the inducer molecule for P_{uhpT}. In the absence of UhpA∼P, RNA polymerases barely binds to this promoter, resulting in virtually no transcription [1]. However, once UhpA∼P binds upstream of the promoter region, it interacts with RNA polymerase and thus activates transcription (Figure 1.4 B). Of course, UhpA∼P does not eternally remain phosphorylated our bound to the promoter. UhpA∼P spontaneously de-phosphorylates and is actively de-phosphorylated by another component of the *uhp* system, the protein UhpB. Therefore, the transcription rate of P_{uhpT} is dependent on the concentration of phosphorylated UhpA, which, in turn, depends on the extracellular concentration of g6p. Increasing extracellular g6p is equivalent to increasing the number of UhpA∼P proteins, which results in a higher probability for P_{uhpT} to be actively

transcribed by RNA polymerase. Of course, there is a saturating concentration of external g6p and increasing its concentration above this value does not increase the transcription rate of the promoter. In mathematical terms, we model the dependency of the transcription rate on inducer concentration by a monotonously increasing, saturating function of the concentration of g6p, usually using a Hill function (see below).

The two types of promoters described in the previous sections are almost entirely dependent on their inducers, with virtually no transcription in their absence. However, many genes are transcribed without further control by transcription factors in given physiological conditions. Their transcription is said to be constitutive. The expression of the corresponding genes is only dependent on the abundance of the global gene expression machinery, for example, the concentration of RNA polymerase. Furthermore, specific transcription factor can act as an inducer of one gene and a repressor of another. This is the case of the DnaA protein, which acts as an activator of the *nrdAB* operon [4] and as a repressor of the *guaB* gene [5].

Regulation by repression versus regulation by activation

The difference between repressors and activators is critical in many respects. Since we are interested in the long term stability of our strains, I will discuss here the difference with respect to mutations. Since transcription factors are proteins coded by about 1 kb of DNA, they constitute a large target for mutations. Inactivating an activator will render the promoter silent, while a mutation inactivating a repressor will lead to the constitutive expression of the target gene. Using our previous promoters as examples: if the activity of LacI were compromised by mutation, then P_{lac} would no longer be repressed and transcribe at the maximum rate, irrespective of the absence or presence of IPTG. On the contrary, if UhpA were mutated, the transcription rate of P_{uhpT} would remain near zero independently of the presence or absence of extracellular g6p. In both cases we loose control over the promoter, but, depending on the function of the target gene, one situation is preferable to the other.

In a situation where an activator controls a promoter transcribing a protein needed for cell growth, a mutation in the activator will lead to a loss of viability for that specific cell. In other words, in a population, the mutated cell would be eliminated. On the other hand, if the same gene were controlled by a repressor, the mutated cells would overtake the population since their growth is no longer controlled.

There are, of course, situations in which using a repressor is more advantageous that using an activator. For example, when the expressed protein impedes cell growth or its over-expression is lethal, a mutation in a repressor would be eliminated by the same argument as above. The choice of a promoter for genetic studies depends entirely on the study and the benefit versus the risks of using one over another should be thoroughly thought of before the construction of the recombinant strain.

To finish this comparison between activators and repressors, we must point out that it is possible for a promoter regulated by a repressor to become silent after mutations in the repressor. For example, a mutation that locks the repressor permanently into the state of high affinity for the DNA, irrespective of the presence of the inducer, would result in permanent repression of the target gene. Such a mutation is equivalent to a mutation that no longer binds the inducer.

However, only very specific base-pair changes will produce such a mutant, whereas a stop codon at any position of the protein coding sequence will lead to the inactivation of the repressor. In other words, the target for the second kind of mutation is a couple of nucleotides at most, whereas the target for the first kind of mutation (inactivation of a protein) is on the order of a thousand nucleotides. The mutation rate of *E. coli* is estimated to be about 10^{-10} per nucleotide per generation [6]. In practical terms, this means that we need about 10 ml of a dense culture, corresponding to 10^{10} cells, in order to obtain a specific nucleotide change. Our typical culture volumes are 1 ml of medium density cells. In other words, we do not expert to observe a mutation of the second type in our experiments.

Measuring promoter strength

We have been talking about promoter strength and transcription rates without explaining how to measure these variables that are fundamental for quantitative models. There are two methods for quantifying the activity of a promoter. We can write a mathematical model describing the reactions and, provided the reaction rates are known, calculate the activity of the promoter. The second method consists in experimentally measuring the transcription rates of the promoter. One of the best known methods of this type is the Miller assay, developed by Jeffrey H. Miller [7], and later simplified by Zhang and Bremer [8]. The assay is based on constructing transcriptional fusions of the promoter with β-galactosidase and measuring the the activity of the enzyme in a cell extract. At steady state, the amount of β-galactosidase produced is directly proportional to the promoter strength. This assay gives the activity of the promoter in Miller Units per time unit ($MU \cdot time^{-1}$), which is equivalent to roughly two molecules of protein produced per minute in an *E. coli* cell [9]. This assay can therefore be used to directly estimate the absolute number of proteins expressed by a given promoter. Note, however, that the value obtained comprises transcription *and* translation and is different from the "simple" transcription rate, which only measures the quantity of mRNA produced. The Miller Units take into account the efficiency of both, transcription and translation.

1.2.2 Translation

Translation is the second step of protein synthesis and revolves around the interaction of the newly created mRNA molecules with the ribosomes. During this step, ribosomes will attach to mRNA and synthesize the associated proteins. Let us take a look at mRNA first.

Ribonucleic Acid

RNA is a macromolecule synthesized by RNA polymerase and composed of a succession of nucleotide bases connected by a sugar-phosphate backbone. This molecule is very similar to DNA, but there are some key differences. While DNA is very stable, the additional hydroxyl on the sugar moiety renders the molecule much more instable in an aqueous environment. Furthermore, the degradation of RNA is rapid *in vivo* due to the presence of RNAse enzymes in the cell. In *E. coli*, the average half-life of mRNA is on the order of three minutes [10]. Another chemical difference between DNA and RNA is the replacement of thymine (DNA) by uracil (RNA).

There are different kinds of RNAs in *E. coli*, each with a specific function. The most abundant RNA in the cell is "stable RNA", which comprises ribosomal RNA (rRNA) and transfer RNA (tRNA). These molecules are part of the translation machinery. The most diverse type of RNA is messenger RNA (mRNA), that will be translated into protein as described above. Finally, small, untranslated RNAs (smRNA) participate in the regulation of gene expression, mostly at the translational level by forming repressive or activating base-pairing structures with mRNA. In this work, we will focus on mRNA only. If one were to equate DNA to a biological blueprint of the cell, then mRNA would be the instruction manual giving detailed information about how to build the elements of the blueprint. mRNA is the intermediary between DNA and ribosomes, which will synthesize the proteins based on the sequence of mRNA.

Ribosomes

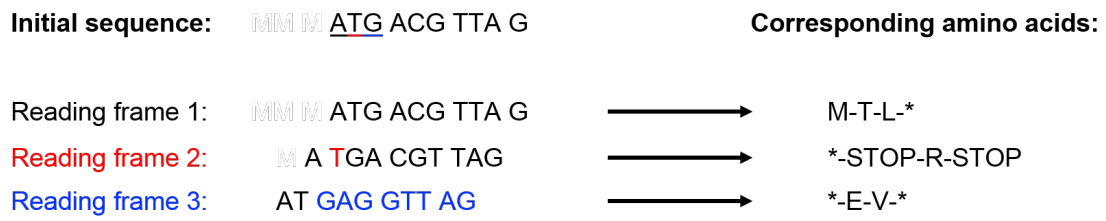

Figure 1.5: Reading frame. *Three identical sequences can give completely different proteins just by changing the reading frame.*

We can push the analogy further by thinking of ribosomes as the workers that will build the proteins from the instructions given by mRNA. They start translating the instructions encoded in the mRNA, starting at specific sequences called ribosome binding sites (RBS). As with promoters, the closer the sequence of the RBS to the consensus, the more efficiently the ribosomes will recognize the signal. In other words, the strength of an RBS is determined by the sequence and by potential interaction with regulatory RNAs that impede or facilitate the access of ribosomes to the RBS.

Ribosomes read mRNA in groups of three, called codons. Each triplet of nucleotides codes for one specific amino acid. Reading bases in groups of three heavily impacts the importance of regulatory sequences in mRNA, since the starting base can potentially alter the entire sequence as shown in figure 1.5. Certain codons play a special role for the behavior of ribosomes, such as the start and the stop codons.

The start codon tells the ribosome to start the translation process. To be recognized as a start codon, the sequence, usually AUG, has to be located just downstream (within some ten nucleotides) of the ribosome binding site.

There are three codons that stop translation. The sequence between the start and stop codons is called an open reading frame and corresponds to an uninterrupted translation sequence. By the constraint of reading codons as triplets, the length of an open reading frame is necessarily a multiple of three. Since there are only twenty different amino acids, but 64 different codons, the code is redundant: several codon sequences will lead to the same amino-acid. In other words, the mapping of codons to amino acids is unique; the inverse is not true.

Mutations Since DNA represents the blueprint of all cellular proteins, mutations have a lasting impact only if they occur in the DNA. An error in transcription, i.e., a mutation in the mRNA, will only affect the few proteins translated from this particular mRNA. This, coupled with the very short halflife of RNA, clearly establishes DNA as the most important target for mutations.

Because of the importance of the reading frame, adding or deleting bases within a protein-coding sequence has a very strong impact on the resulting protein. The protein sequence starting from an altered reading frame becomes completely different from the original. This is very important to remember when modifying the genome of a cell, since one can easily change the reading frame by inserting or removing portions of genetic material.

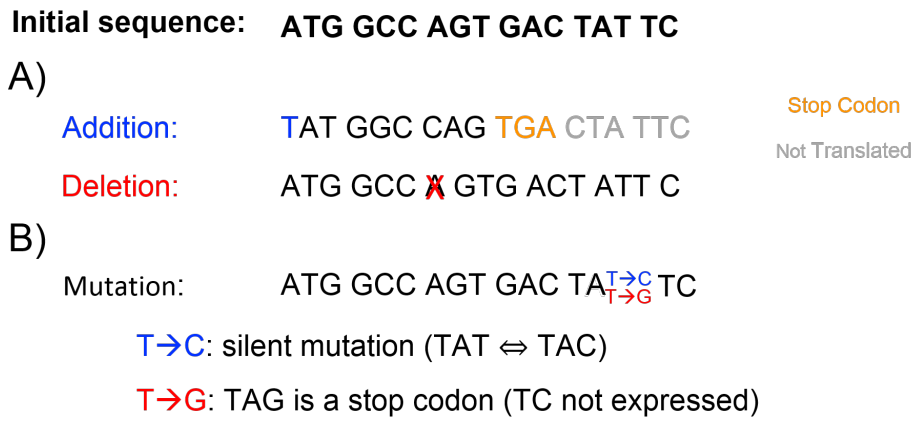

Figure 1.6: Effect of mutations in a protein-coding sequence. A) Additions and deletions. *Adding or deleting a single nucleotide in a protein-coding region changes the rest of the protein sequence since the reading frame is altered.* **B) Substitution mutations.** *A base in the sequence is substituted by another. Such a mutation can have two effects. Either the new codon formed codes for the same amino acid as the original one, or the new codon changes the amino-acid incorporated at this position in the protein. The former is called a silent mutation, whereas the latter is called a point mutation. If the point mutation introduces a stop codon, the protein is truncated and most likely non-functional. A simple amino-acid replacement will modify the protein structure with more or less severe effects depending on the location of the amino-acid and the nature of the substitution.*

Changing a base for another is a mutation that is usually much more manageable and can often be repaired by the cell itself. However, big problems can still arise when the code for an amino-acid is transformed into a stop codon. There are three different triplets that act as a stop codon. Of the 61 other triplets, 20 can potentially become a stop codon with a single point mutation. This shows that the appearance of a stop codon is not a rare mutation. This further emphasizes the difference in robustness between promoters in the genome. Repressors and activators are very sensitive to this type of mutation, with opposite effects on the expression of the target gene. We need to keep these considerations in mind when designing a synthetic genetic control circuit that should remain stable for many generations.

1.2.3 Fluorescent Reporters

In the past two decades, quantitative studies have blossomed in molecular biology. This is due to the technological advancements which, among other things, have allowed us to follow in real time the production of proteins of interest via fluorescent reporters, which otherwise have little impact on the cell.

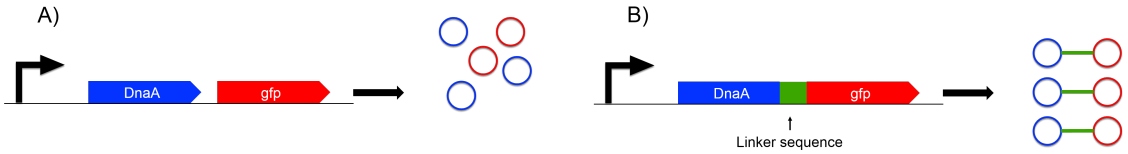

Figure 1.7: Fluorescent reporter systems. A) Transcriptional fusion. *The gene coding for a fluorescent protein is expressed using the signals (promoter, RBS, ...) of the gene of interest. In the example, when dnaA is transcribed, the reporter protein is transcribed on the same mRNA. The rate of production of the green fluorescent protein (gfp) is identical to the rate of production of dnaA. If the half-life of the proteins and the translation efficiencies of the two genes were identical, even the concentration of Gfp and DnaA would be identical.* **B) Translational fusion**. *The fluorescent gene is linked to the gene of interest. A fluorescent reporter is added to the protein of interest via a linker, producing a fusion protein. This ensures an exact 1:1 ratio of DnaA and Gfp, but adds a relatively bulky domain to DnaA, possibly interfering with some of its interactions or functions.*

Fluorescent proteins were a revolution to molecular biology. Fluorescent proteins are naturally found in organisms such as jellyfish, as is the case of the green fluorescent protein, GFP [11]. However, it was not until 1994, when it was shown that they could be used as tags *in vivo*, that fluorescent reporters truly became important tools for biological research [12]. Since then, they have been used in a multitude of ways to visualize proteins in living organisms. These proteins are not part of the genome of *E.coli*, but can be introduced into its genome via genetic modification.

The gene coding for a fluorescent protein can be inserted such that it is transcribed on the same mRNA as the gene of interest (Figure 1.7). In this case, the fluorescent protein will be under the control of the promoter being studied and the rates of production of the fluorescent protein and the protein of interest are proportional to each other. This sort of construction can only give relative concentrations of the expression of a certain gene, since the fluorescent protein is not necessarily translated with the same efficiency as the target gene. This can be dealt with by changing the RBS of the fluorescent gene to match the one of the protein of interest. However, equal protein concentrations would only be assured if the respective degradation rates were also identical. This sort of construction is called a transcriptional fusion.

One can also fuse the fluorescent protein to the protein of interest by cloning it in phase into the open reading frame of the target protein. This is called a translational fusion. Often, a small sequence, called a linker, is inserted between the two proteins in order to minimize steric hindrance. When the protein of interest is expressed, the fluorescent tag will be produced at equal stoichiometry (Figure 1.7). In practice, this type of fusion is challenging and the choice of the location where the gene coding for the fluorescent protein is to be added, as well as the linker

are critical. After being translated, proteins fold into a defined three-dimensional structure. Attaching another module to the protein may hinder the folding of both proteins rending them non-functional. Still, translational fusions allows one to visualize proteins in living cells, but there are some drawbacks. One cannot be sure that the degradation rate of the tagged protein is the same as the one of the non tagged protein and the functionality of the tagged protein can still be affected by the tag.

There are other drawbacks to using fluorescent proteins *in vivo*. There is always some level of phototoxicity during fluorescence measurements,which can stress the cells. Furthermore, the rate of maturation of the fluorescent protein may be too slow for measuring fast kinetics. Maturation rates range from tens of minutes to hours.

Nonetheless the drawbacks are more than manageable and many improved fluorescent proteins have been developed for molecular biology [13]. Fluorescent proteins are particularly useful for the *in vivo* observation of single cells. There are many microscopy setups that allow this. One of the more interesting ones for following many individual cells is the microfluidics device.

Chapter 2

Controlling the replication of DNA

We have focused so far on gene expression, but we still have not addressed one of the main subjects of this thesis: DNA replication. This process is necessary for the cell to divide. We will begin by a brief overview of the process and then analyze how we can take control of it.

2.1 DNA replication

Before cell division, *E. coli* has to make an identical copy of its chromosome in order to produce two genetically identical (minus mutations, see above) cells. There are slight biochemical differences between the mother and daughter cell, but they are identical at the genetic level. DNA replication is therefore crucial for cell division and many studies have shown that this process does not proceed without prior replication of the DNA [14, 15, 16]. Before continuing with the description of the mechanisms of DNA replication, we must clarify that the nucleotides of the circular chromosome of *E. coli*, $4.6 \cdot 10^6$ base pairs, are numbered clockwise, starting from a reference point chosen by convention.

Replication of the chromosome of *E. coli* starts at the origin of replication, *oriC*, a specific sequence located roughly at position 3 925 thousand base pairs (kbp). Replication is initiated by the DnaA protein [17], which binds to thirteen specific binding sites within *oriC*. The assembly of this multiprotein complex on DNA (comprising other proteins in addition to DnaA) facilitates the melting of the double DNA helix near *oriC*, thereby creating a loop formed of two single-stranded DNA [18]. Among the other proteins recruited are two molecules of DNA polymerase, which will proceed to bi-directionally replicate the chromosome.

Just as transcription, DNA replication is a directional process, proceeding in the 5' $\rightarrow$ 3' direction. As in transcription, the incoming nucleoside triphosphate is hydrolyzed, transferring its phosphate group to the 3' hydroxyl of the polymer and releasing a pyrophosphate. The site where a newly created DNA strand separates from the older one is called a replication fork. The replication fork therefore moves on the chromosome as the two DNA polymerases advance from the origin of replication to the terminus, located at 180° with respect to the origin. The description is overly simplified in this report, but is sufficient for the objective of this work.

Even though cell division does not normally occur without prior DNA replication, the opposite is quite possible. In artificially growth-arrested cells, replication continues to occur, resulting in elongated cells that contain multiple copies of the chromosome [19]. In this sense, DNA

replication seems independent of cell division. Fast growing cells contain more than one copy of the chromosome at all times. A new round of DNA replications starts before the previous round of replication has reached the terminus. This rush is necessary since the replication only starts at the origin and needs to proceed until the terminus. Given the speed of DNA polymerase, this process takes about forty minutes. In order for cells to divide every twenty minutes, as is the case for *E. coli* in rich medium, a new round of DNA replication has to be initiated every twenty minutes.

The variable number of chromosomes as a function of the growth rate represents a big challenge for the quantitative analysis of gene expression, and in particular the study of stochasticity *in vivo*. Taking into account the number of chromosomes, the number of gene copies and the variability on this numbers in a population introduces many variables difficult to control. This difficulty was one of the motivations at the beginning of this thesis to construct a strain of *E. coli* where replication is controlled by an external stimulus.

Stopping DNA replication can be done either through genetic modifications of the bacterium or by adding drugs to the growth medium. However, using drugs to suppress DNA synthesis forces the cell into an abnormal state, affecting many physiological parameters. Ciprofloxacin, for example, inhibits DNA gyrase and therefore slows, or completely stops, DNA replication [20]. However, this inhibition leads to the formation of arrested replication forks within the cell. The cell responds to the presence of these structures by producing proteins to repair its DNA via the SOS response [21]. The physiological and metabolic state of the cell is greatly disturbed, since the SOS response activates many functions that are normally unused in the cell [**?**]. Stopping DNA replication this way may compromise the validity of results that aim to study the behavior of the cell in normal, non-stress conditions. The other possibility left is to genetically modify the cell to stop DNA replication in a conditional and reversible manner. Reversibility is needed, since inhibiting DNA replication stops cell division, and therefore long-term viability.

The best candidates for genetic engineering are the proteins that directly participate in DNA replication, most importantly the DNA replication initiation protein DnaA. The DnaA protein is essential for DNA replication. Previous reports show that temperature sensitive mutants of DnaA stop DNA replication in *E. coli* at the non permissive temperature [22, 14]. We also considered other candidates that play a role in DNA replication, such as the proteins DnaB and DnaN. DnaN,he β-clamp subunit of DNA polymerase III, also assumes other roles in the cell, such as participating in DNA repair mechanisms [23]. The absence of DnaB creates arrested replication forks that are lethal to the cell [22]. Arrested replication forks were not reported with DnaA temperature sensitive mutants.

Controlling DNA replication via the DnaA protein allows us to quantify the minimal amount of DnaA proteins needed to initiate DNA replication, a value that has not been unambiguously reported yet.

Depending on how we proceed during the genetic modifications, it is also possible to study the two other genes that are transcribed by the *dnaA* promoter and are part of the *dnaA* operon, *dnaN* and *recF*. If we wish to do all of this, we need to change the location of the DnaA gene and put it under the control of an inducible promoter. The choice and location of such a promoter are important, since the final construction needs to be robust and mimic the expression of DnaA

proteins during exponential growth. During the latter part of this chapter, we will discuss the choice of the promoter and its location.

Before this, let us take a look at the *dnaA* gene and its operon.

2.2 DnaA

2.2.1 The dnaA operon

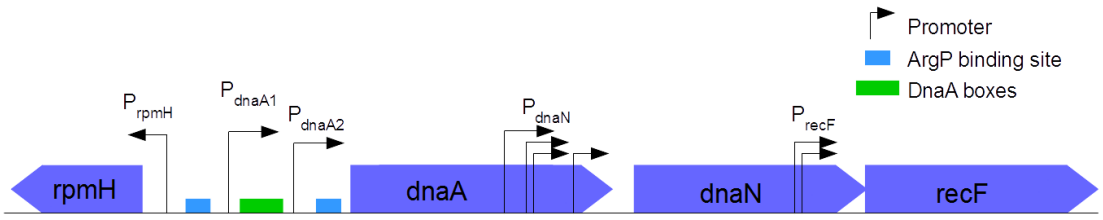

Figure 2.1: The dnaA operon. *dnaA is transcribed from two promoters that are regulated by ArgP and DnaA itself. The operon also includes the* dnaN *and* recF *genes. Both of these genes possess additional promoters within the coding regions of the preceding genes.*

The *dnaA* gene is located at position 3 883 kb in the chromosome of *E. coli*, very close (∼50 kb) to the origin of replication, *oriC*. *dnaA* is transcribed as an operon along with *dnaN* and *recF*. The proteins coded by these two genes also interact with DNA (Figure 2.1).

The two promoters upstream of dnaA, p_{dnaA1} and p_{dnaA2}, transcribe the entire operon [14, 24]. p_{dnaA2}, the strongest of the two promoters, is responsible for 80% of the transcription of dnaA [24]. The promoter activity was measured by Hansen *et al.* [14] to be about 70 Miller units (MU) for different growth rates on different media. The two *dnaA* promoters are regulated by DnaA [25]. In addition to the full length transcript originating at the *dnaA* promoters, *dnaN* and *recF* possess additional promoters.

dnaN possesses four promoters located towards the end of the coding region of *dnaA*. *dnaN* codes for the β subunit of the DNA polymerase III holoenzyme. The β-clamp is responsible for the high processivity of the polymerase [26, 27]. *dnaN* is an essential gene, just as *dnaA,* which means that if it is not properly expressed, cells will not grow. The β-clamp also participates in the SOS response of the cell, mainly during lesions of the DNA provoked by UV radiation [28, 23].

The last gene of the operon, *recF*, is also induced by UV-radiation. Two additional promoters, located in the coding region of *dnaN* transcribe *recF*. The role of RecF is to allow proper arrest and recovery of replication forks damaged by UV radiation [29]. Contrary to *dnaA* and *dnaN*, *recF* is not an essential gene.

An interesting point about *dnaN* and re*cF* is that their individual over-expression has little effect on cell viability, while the over-expression of both genes leads to mortality of the bacteria. Lethality is further enhanced if *dnaA* is over-expressed as well [30]. These observations show the importance of the regulation of the operon and the delicate balance that the promoters of the operon need to maintain. The first gene of the operon, *dnaA*, is itself very delicate. Even small modifications can render the DnaA protein inactive. On the other hand, over-expression of only

dnaA is not lethal, unless it is combined with a deficiency in DNA repair mechanisms [31, 32]. Strong overexpression of DnaA does, however, slow down the growth rate of *E. coli*.

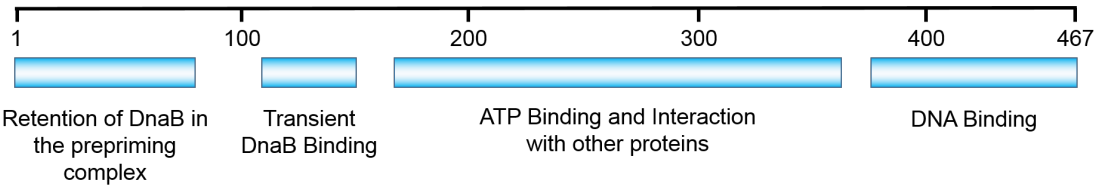

Figure 2.2: Functional domains of DnaA. *The DnaA protein contains four functional domains. Their roles are indicated below the schematic. The numbers on top show the codon numbers of the gene* [15].

The DnaA protein contains four functional domains [15]. The major function of DnaA is to initiate DNA replication. In order to do so, it needs to bind to DNA, recruit other replication proteins and load them onto the origin of replication. Each functional domain shown in Figure 2.2 fulfills one of these functions. The first domain allows recruitment of DnaB, along with other proteins. Without this domain, DnaA could not initiate DNA replication. The second domain allows a transient interaction between DnaA and DnaB. The third domain contains the adenosine triphosphate (ATP) binding site of DnaA. This domain is also involved in the interaction with other proteins, such as RepA. This domain is extremely important, since the properties of DnaA change when bound to ATP or ADP. ATP bound DnaA has a higher affinity for DNA than DnaA without a nucleotide ligand or DnaA bound to ADP. The active, DnaA-ATP form of DnaA, is converted to the "inactive", DnaA-ADP form, by hydrolysis of ATP. The fourth and last domain allows DnaA to recognize specific binding sites on the chromosome, called DnaA boxes. The four *dnaN* promoters are located in the nucleic acid sequence coding for this domain [33, 34].

The role of DnaA in DNA replication

DnaA assumes three different states, depending on the bound nucleotide: the unbound state (no nucleotide bound), the active state in the form of a DnaA-ATP complex, and the inactive state in the form of DnaA-ADP. DnaA can interact with many other proteins and recruit them to the chromosome, which is one of its key roles during DNA replication.

DnaA binds to specific sequences called DnaA boxes. The consensus binding site has the following sequence: $TT^{A}/_{T}TNCACA$ [15, 14]. These DnaA boxes are found all over the chromosome, including the *oriC* region. More than 300 such sites have been identified [35]. The affinity of DnaA for the binding sites depends on the sequence of the DnaA-box; the closer the sequence is to consensus, the stronger the binding. High affinity boxes can bind DnaA in any of its states, however low affinity DnaA boxes are only bound by DnaA-ATP [36]. DnaA-ATP is also the only form of DnaA that can initiate DNA replication.

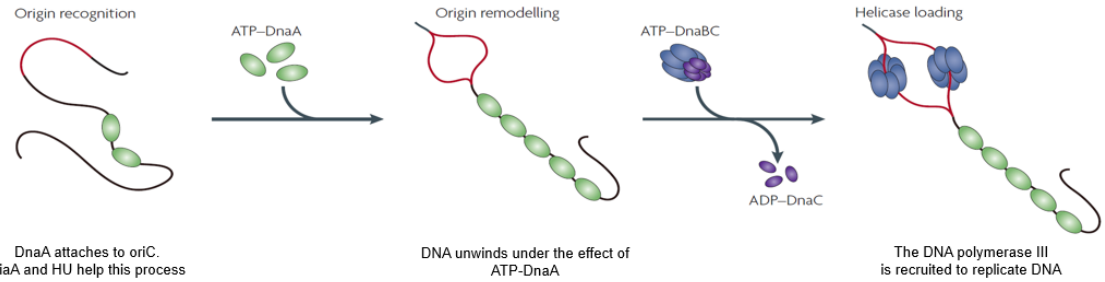

Figure 2.3: Initiation of DNA replication by DnaA. *DnaA binds to DnaA-boxes in the origin of replication. The nucleo-protein complex unwinds the DNA and recruits additional components of the replication machinery, most importantly DNA polymerase III* [18].

At initiation of DNA replication, DnaA-ATP binds to the DnaA boxes in *oriC*. Some models estimate that binding to seven boxes is sufficient to initiate DNA replication [37], but experimental evidence suggests there are at least thirteen possible binding sites for DnaA at *oriC* [38]. The binding of DnaA-ATP is promoted by two proteins called DiaA [39, 40] and HU [41]. DiaA binds several DnaA molecules and stimulates the formation of ATP-DnaA complexes near *oriC*, while HU stabilizes the DnaA oligomers, as well as DnaA bound to *oriC* [42]. DnaA bound to *oriC* recruits DnaB and DnaC [43]. This allows DNA polymerase III to start DNA replication. This process is summarized in Figure 2.3. This mechanism seems to suggest that seven (or may be thirteen) DnaA proteins are sufficient to initiate DNA replication, much fewer than the hundreds of DnaA proteins present in the cell. We will come back to this point below.

Once DNA polymerase has started replicating DNA and the replication fork moves on the chromosome, the original strand of DNA will be methylated, while the newly synthesized strand does not yet carry this modification (the newly synthesized double-stranded DNA is hemi-methylated) [44]. The hemi-methylated sequence near *oriC* has a high affinity for the SeqA protein, which binds to sites overlapping the DnaA boxes and containing the hemi-methylated GATC sequences. DnaA can no longer bind to *oriC*, thus precluding a premature re-initiation of replication. The replication origin is thus sequestrated until the Dam methylase methylates the newly synthesized strand [45]. SeqA dissociates, and the origin is ready for another round of replication.

The mechanism implies that very few DnaA proteins are needed to start DNA replication. Previous publications suggest that a mere 20 copies suffice [46]. However the number of DnaA proteins in *E. coli* is much larger than that. Depending on the measurement method, numbers range from 800 to 2000 in rich medium [47, 46], while in minimal medium, this number was estimated to be as low as 290 molecules per genome per generation [48].

Regulation of and by DnaA

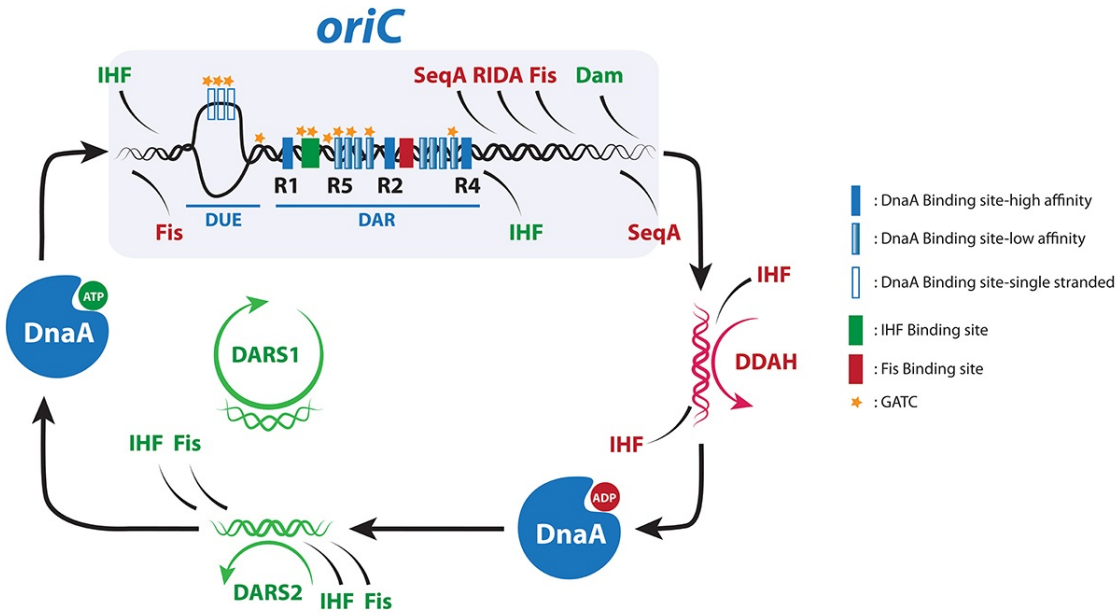

Figure 2.4: The chromosome replication cycle. *The schematic summarizes the major events linked to the replciation of the chromosome of* E. coli. *In green are proteins or mechanisms that promote DNA replication and in red those that inhibit replication. Most of the DnaA boxes are located in a region of oriC called DnaA assembly region (DAR). The four strongest sites, R1, R2, R4, and R5, initiate binding of DnaA to the weaker sites. The assembled replication complex unwinds the the duplex unwinding element (DUE) region, which allows other proteins to be loaded onto the chromosome. DNA polymerase III starts bi-directional replication. Just after initiation of replication, SeqA proteins sequester the origin of replication to prevent re-initiation of the same origin. RIDA and DDAH mechanisms hydrolyze DnaA-ATP into DnaA-ADP, further preventing re-initiation or imposing a delay for new rounds of replication. Dam methylation, IHF and the DARS mechanisms participate in the re-activation of DnaA by promoting the conversion of DnaA to the active DnaA-ATP form. The image was taken from a review by Riber et al.* [38].

While hemi-methylation prevents re-replication, other control mechanisms directly affect DnaA. Hemi-methylation not only modulates the binding of SeqA and DnaA to *oriC*, it may also affect the activity of the *dnaA* promoter, again by SeqA binding to the promoter region when hemi-methylated and therefore terically hindering the access of RNA polymerase [49]. Other forms of regulation involve the many DnaA boxes scattered on the chromosome, the regulatory inactivation of DnaA (RIDA), and more subtle effects on the promoters of DnaA. A diagram summarizing DnaA activation and inactivation mechanisms during a round of DNA replication, reproduced from a recent review [38], is shown in Figure 2.4.

DnaA boxes The numerous DnaA boxes on the chromosome represent a major mechanism for regulating the initiation of DNA replication. To a first approximation, DnaA is produced at a constant rate during the cell cycle. The crucial binding sites at *oriC* will only be occupied once all DnaA boxes on the chromosome are bound by DnaA. Such a mechanism would "count" the number of DnaA molecules in the cell and trigger replication of the chromosome once a critical threshold of number of DnaA molecules per chromosome is passed. The newly synthesized DNA

will then provide additional DnaA-boxes, "sequestering" free DnaA. The sequestration of DnaA is particularly efficient when many consensus sequences are packed tightly next to each other. This is the case at the *datA* locus. This region, at about 450 kb from *oriC,* can sequester up to 300 DnaA proteins [50].

RIDA and DDAH The regulatory inactivation of DnaA, RIDA, consists in hydrolyzing the DnaA-ATP complex to form DnaA-ADP, i.e., converting an active, high-affinity form of DnaA into a low affinity form. While DNA polymerase III is replicating DNA, a protein called Hda attaches itself to its β-clamp subunit (encoded by *dnaN*). When the clamp encounters a DnaA protein, Hda interacts with DnaA-ATP and provokes the hydrolysis of ATP bound to DnaA, forming DnaA-ADP [51, 52]. Since DnaA-ATP is required for unwinding DNA at the origin of replication, the RIDA represses the initiation of DNA replication. Dat-A dependent DnaA-ATP hydrolysis is another mechanisms that converts DnaA-ATP into DnaA-ADP, but is less efficient than RIDA. DnaA-ADP is commonly referred to as inactive DnaA, since in this state, DnaA does not bind to low-affinity specific DnaA boxes on the chromosome [36].

DARS To return to its DnaA-ATP form, DnaA binds to two non-coding chromosomal regions, referred to as DnaA-reactivating-sequence-1 and 2 (DARS1 and DARS2) [53]. These regions bind other regulatory proteins, such as IHF and Fis, to reactivate DnaA into its ATP bound form [54]. There is also evidence showing that DnaA-ADP can be rejuvenated into its ATP bound form when interacting with phospholypids, present in the membrane of the cell. One report showed that the bulk of DnaA in the cell is located near the membrane [46], further supporting the idea of the cell membrane acting as a reservoir for re-activating DnaA-ADP.

DnaA promoters Inactivation and activation of DnaA also regulates the initiation of DNA replication indirectly by changing the activity of the DnaA promoters. Several low affinity and high affinity DnaA boxes between p_{dnaA1} and p_{dnaA2} affect the transcription of *dnaA*. The binding of DnaA-ATP to these sites represses and activates transcription [25, 55, 56]. A recent study has shown that p_{dnaA2} is activated by DnaA-ATP [57]. This publication shows that the activity of p_{dnaA2} decreases when mutating the high affinity DnaA box that is closest to this promoter. This result suggests an activation role for DnaA when binding those DnaA-boxes. Mutating the low affinity boxes or the other high affinity box in this region resulted in an increase of the transcription rate of both promoters, showing the repressor characteristics of DnaA. The model derived from these data proposes an activation of p_{dnaA2} at low DnaA concentrations due to the high affinity DnaA box. When the concentration of DnaA increases, the low affinity sites are bound as well, sterically hindering the access of RNA polymerase to the promoter and therefore repressing transcription. Other regulators affect the transcription of the *dnaA* promoters as well. The net result of these different regulatory mechanisms is a roughly constant transcription rate, and probably intracellular concentration, of DnaA.

As shown in Figure 2.1, there are two binding sites of ArgP (previously known as IciA) close to the *dnaA* promoters. One is upstream of p_{dnaA1}, the second downstream of p_{dnaA2}. Another study showed that, while the latter site did not affect significantly the activity of the promoters, the binding on the first site increases the activity of p_{dnaA1} [58]. This interaction can

counterbalance to some extent the inhibition of p_{dnaA} by DnaA binding to some of the DnaA boxes in the region.

Other roles of DnaA

The capacity of DnaA to affect the transcription of promoters is not limited to p_{dnaA1} and p_{dnaA2}. Since there are many DnaA boxes on the chromosome of *E. coli*, it is quite natural that DnaA not only plays a role as the DNA replication initiator protein, but also as a transcription regulator for many genes of the organism. The following is a non exhaustive list of the genes that are significantly regulated by DnaA. We also comment on potential effects of these genes in conditions when the concentration of DnaA is artificially decreased, a condition that we explore in our experiments.

nrdA and nrdB The *nrdAB* operon codes for ribonucleoside diphosphate reductase, an enzyme that catalyzes the reduction of ribonucleoside diphosphates into deoxyribonucleoside diphosphates. This is one of the first reactions on the pathway to DNA synthesis. The promoter of both of these genes has two DnaA boxes and without them its activity is reduced up to two-fold [4]. However this does not compromise the viability of the cell. A reduced activity of the ribonucleoside diphosphate reductase only slows down the replication forks without any lethal effects.

guaB The *guaB* gene codes for inosine 5'-monophosphate (IMP) dehydrogenase, which catalyzes the conversion of IMP into xanthosine 5'-monophosphate (XMP) during the production of guanine [59]. *guaB* possesses a DnaA box downstream of the start site of transcription. Binding of DnaA leads to a two-fold repression of *guaB* expression [5]. GuaB has been overexpressed in *Bacillus amyloliquefaciensm* without any adverse effect for the cell [60]. Therefore, any increase in expression due to the absence of DnaA should not have an impact on our experiments (see below).

rpoH *rpoH* codes for the sigma factor σ^{32}, the expression of which is induced by heat shock. The RNA polymerase holoenzyme containg σ^{32} directs the transcription of "heat shock" proteins that *E. coli* produces to cope with a temperture shift, for example from 30° to 42° [61]. The expression of σ^{32} has a complex regulation which includes CytR [62], Crp (cAMP receptor protein), as well as DnaA and other global transcription factors. DnaA, in particular, is the only repressor of the *rpoH* gene and affects two of its five promoters [63]. However, given the complex regulation of this gene, as well as the lack of harmful effects when the gene is expressed from multicopy plasmids [64], we can assume that an increased expression of *rpoH* due to the loss of DnaA repression will not impact our results.

mioC *mioC* is a gene located next to *oriC* and is essential for biotin synthesis [65]. Because of its location, there is a possibility that this protein plays a role in DNA replication. However, so far results are inconclusive [66, 67]. *mioC* is repressed by DnaA, but even if this regulation is lost, no drastic effect for the cell were reported [68].

uvrB UvrB, along with UvrA and UvrC form the nucleotide excision repair (NER) system, one of the mechanisms for repairing damaged DNA. UvrB forms a complex with UvrA, binds DNA, and facilitates the action of UvrC [69]. The *uvrB* gene is regulated by LexA and therefore activated during the SOS response. There is also evidence that DnaA represses this gene [70]. Since there is a LexA regulation, DnaA repression most likely serves to fine-tune the expression or *uvrB*. We can assume that even without DnaA repression, the increased quantity of UvrB will not adversely affect the cell.

proS *proS,* also known as *drpA*, codes for a protein that is part of proline-tRNA synthetase (ProRS), which, among other proteins, allows the faithful translation of mRNAs into proteins [71]. It was reported that over-expression of *dnaA* represses the expression of *proS* by eight-fold [72]. No other publications mention a regulation of *proS* by DnaA, leading us to believe that the regulatory effect of DnaA is probably negligible compared to other interactions, such as with amino acids [73].

Beside these genes, several other proteins are affected by mutations in DnaA [74]. However, although DnaA regulates many genes, its absence does not compromise the viability of cells so long as DNA can be replicated via a different pathway [75]. After this review of available literature data, it is clear that DnaA is a perfect protein to control DNA replication externally, since its absence should not be detrimental to most other processes of the cell.

As described in detail below, we will control the expression of *dnaA* using an inducible promoter. However, the type of promoter needs to be defined. Stopping the expression of DnaA stops cell division. Growth arrest is a very severe interference with the cell and any mutation that circumvents such a control is very strongly selected. Any such mutation would overtake a population very rapidly because the mutated cells would grow exponentially, while the rest of the population does not grow at all. On the other hand, we have to choose the inducible promoter used for the controlled expression of DnaA such that maximal induction does not compromise cell viability. Remember that a large overexpression of DnaA is lethal [32]. Given these design criteria, we need an activated promoter (which, compared to a repressed promoter, greatly reduces the probability of mutations that lead to constitutive expression) that has a promoter strength exceeding the one of the natural *dnaA* promoter.

2.3 The uhpT promoter

Being an activator and having the correct promoter activity are not the only characteristics needed for our inducible promoter. The promoter should also have a low basal activity to avoid DnaA accumulation when it is not induced and it should be induced by a single, specific molecule, that is not metabolized by our cells. Ideally, in order to avoid any interference with transport proteins, the inducer molecule should remain in the medium, i.e., be recognized by a membrane-bound receptor.

2.3.1 The uhp pathway

E. coli cells possess a response pathway sensitive to glucose-6-phosphate (g6p) that suits all our requirements: the *uhp* system. The genes coded by the *uhp* operon allow the intake of sugar phosphates [1], which are not present in the media we use. The operon is located close to *oriC* and the *dnaA* operon, at 3,85 Mb, i.e., at a position that allows strong expression at high growth rates.

Four genes make up the *uhp* pathway. *uhpB* and *uhpC* code for transmembrane proteins that sense g6p in the medium. The binding of g6p to the receptor is transmitted to the interior of the cell via conformational changes that lead to the phosphorylation of the third protein of the regulon, UhpA. In the phosphorylated form, UhpA∼P, binds to the promoter of the last gene of the system, *uhpT*. This gene codes for a transmembrane protein that transports g6p inside the cell.

To summarize the *uhp* pathway: first, the UhpB-UhpC couple recognize the presence of the inducer molecule g6p. This leads the phosphorylation of UhpA, converting it into UhpA∼P, which in turn activates the expression of UhpT, which will allow g6p to enter the cell (Figure 2.5).

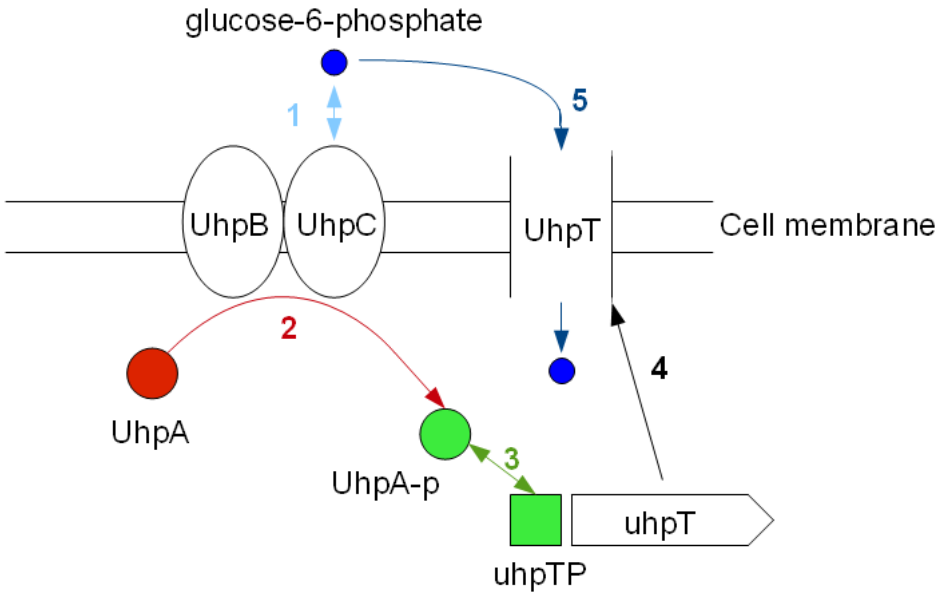

Figure 2.5: The *uhp* pathway:(1) *UhpC recognizes glucose-6-phosphate.* (2) *After recognition, UhpC and UhpB phosphorylate UhpA.* (3) *The phosphorylated UhpA∼P binds to the promoter of uhpT.* (4) *UhpT is produced and inserts into the cell membrane.* (5) *UhpT allows glucose-6-phosphate to enter the cell.*

UhpT is an antiporter that actively transports g6p from the outside of the cell to the inside and obligatorily transfers an inorganic phosphate to the outside of the cell in this reaction. Without the expression of *uhpT*, g6p can not enter the cell. The *uhpT* gene is seemingly isolated from the other genes of the regulon. While the promoter of *uhpT* is activated by UhpA∼P, the UhpT protein does not interact with any of the others. The overall characteristics of the *uhp* system can therefore be described as an inducible promoter, p_{uhpT}, that responds to the concentration of g6p in the medium.

We will now describe in more detail, gene by gene, the characteristics of the *uhp* system and

we will argue that the system is suitable for our study of *dnaA*.

uhpT

uhpT codes for a transmembrane protein transports hexose phosphates (sugar phosphates with six carbons) across the plasma membrane via phosphate exchange, i.e., for each hexose-phosphate entering the cell, an inorganic phosphate molecule leaves the cell [76]. The gene is not essential [77] and has no function in cells grown in a medium containing a non-phosphorylated carbon source such as glucose.

The activity of the *uhpT* promoter (p_{uhpT}) was studied by Verhamme *et al.* in different media [1]. In MOPS minimal medium, the activity of the promoter ranges from 0 to ∼10 $MU \cdot min^{-1}$ and attains twice that value in rich, LB-medium. The transcription of *uhpT* is very weak in the absence of inducer; i.e. the baseline activity of the promoter is negligible [78, 79]. The promoter of *uhpT* is activated by the phosphorylated form of UhpA, which binds at two sites in the promoter region, respectively at position –80/-50 and -50/-32 with respect to the transcription start site [79]. The cAMP receptor protein, Crp, binds at position -103.5, located just downstream of *uhpC*, and participates in the activation of transcription [80].

In summary: the regulation of *uhpT* is entirely dependent on the presence of phosphorylated UhpA (UhpA∼P), its promoter is tightly regulated, its basal activity is negligible, it has previously been characterized, and the mode of regulation is activation.

uhpA

UhpA is the regulator of *uhpT*. Although the protein binds weakly to its binding sites upstream of the promoter of *uhpT* [81], it has a much higher affinity when phosphorylated (UhpA∼P). In the canonical signal transduction pathway of this two-component system, UhpA is phosphorylated by UhpBC [82]. As many response regulators, UhpA can also be phosphorylated by the generic phosphate donor acetyl-phosphate [1]. Although the three *uhp* genes, *uhpABC*, are co-transcribed, UhpA has a higher translation rate than the other two genes [48]. TheUhpA∼P complexes can be dephosphorylated either naturally with a half life of ∼60 minutes [83] or by the action of UhpB, which can act as a phosphatase [84].

uhpB-C

The role of UhpB is to phosphorylate and dephosphorylate UhpA; in other words, it acts both as a kinase and phosphatase of UhpA [84]. Other reports suggest that UhpB may also sequester UhpA [82]. This effect is probably negligible due to the difference in abundance between UphA and UhpB [48]. UhpB works in tandem with another transmembrane protein, UhpC. This component of the signal transduction system detects the presence of g6p in the medium [1]. The conformational change upon binding of g6p is transmitted to UhpB, stimulating its kinase activity and thereby increasing the rate of phosphorylation of UhpA. UhpC does not directly interact with UhpA.

The phosphatase activity of UhpBC is stronger than its kinase activity [84]. As a result, *in vivo,* the ratio of UhpA∼P to UhpA is always in favor of the unphosphorylated form, even

in the presence of g6p in the growth medium. This could explain why the activity of the *uhpT* promoter on multicopy plasmids s not significantly increased when compared to the single, chromosomal copy of the promoter [85]. Nonetheless, the few molecules of UhpA∼P are sufficient to activate the *uhpT* promoter at relatively low concentrations of g6p, provided the phosphate concentration in the medium is low [1]. Phosphate inhibits the allosteric activation of UhpB by UhpC. In a medium containing high concentrations of phosphate, the uhp system is only activated at correspondingly higher concentrations of g6p.

2.3.2 Converting the *uhp* signal transduction system into and inducible expression system to control the expression of DnaA

If we want to convert the uhp system into an inducible expression system we have to first prevent the utilization of the inducer, g6p. Deletion of *uhpT* accomplishes this goal because g6p can no longer enter the cell. The activity of the *uhpT* promoter is now monotonically controlled by the concentration of g6p in the medium. The simple deletion of *uhpT* has converted the promoter of this gene into an expression system inducible by g6p.

To control the expression of DnaA using this system, we have two possible solutions: (i) after deleting *uhpT,* we replace the promoters of *dnaA* with the *uhpT* promoter, or (ii) we move the *dnaA* gene from its original location to the position of the deleted *uhpT* gene, effectively replacing *uhpT* by *dnaA*. Both solutions have their merits. The first construction will put the entire *dnaA* operon under external control, i.e., the two other genes of the operon, *dnaN* and *recF*, will be controlled in the same way as *dnaA*. The second option runs the risk of subjecting the cells to lethal effects during dnaA arrest due to the de-repression of the *dnaA* promoters, which will lead to the accumulation of DnaN and RecF. As mentioned above, over-expression of these two genes is lethal to the cell. The lethality will not be a problem for short experiments.

Since both solutions allow us to observe unique features, we decided to construct both. This strategy allows us to study and quantify not only the dependence of DNA replication on the concentration of DnaA, but we can also assess the effects of varying concentrations of DnaN and RecF, which was previously only measured through over-expression of plasmid-borne copies of these genes [30].

An important advantage of using this expression system for controlling the expression of DnaA is the close proximity (distance of 48 kb) of *uhpT* to oriC. Thus, the gene copy number as a function of growth rate of *dnaA* cloned at the *uhpT* locus will be identical to the one at the original locus.

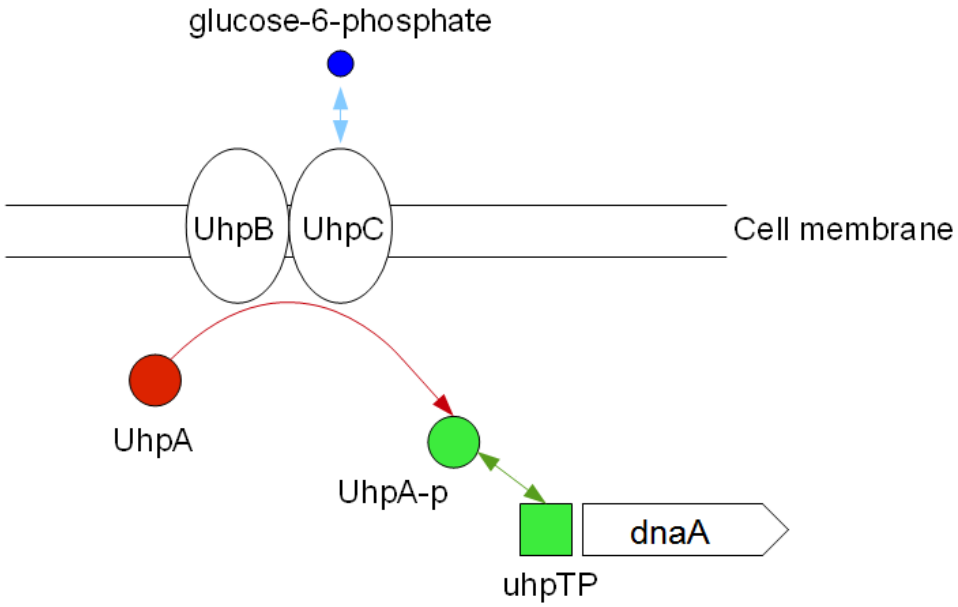

Figure 2.6: The *uhp* pathway:(1) *UhpC recognizes glucose-6-phosphate.* (2) *After recognition, the UhpBC complex increases the rate of UhpA phosphorylation.* (3) *Phosphorylated UhpA binds to the promoter of uhpT.* (4) *dnaA is produced and external glucose-6-phosphate remains constant.*

Figure 2.6 summarizes the pathway used to control the *uhpT* promoter, p_{uhpT}. UhpC recognizes the presence of the inducer molecule g6p in the growth medium. This triggers the kinase activity of UhpB, which increases the amount of UhpA∼P. The concentration of UhpA∼P and UhpA rapidly reaches a new equilibrium value due to the dual effect of UhpB as a phosphatase and a kinase. This increase is sufficient for UhpA∼P to bind with high affinity to the *uhpT* promoter, activating the expression of *dnaA*.

The *uhpT* promoter has been characterized *in vitro* [1]. The promoter activity is described by a Hill function with a Hill coefficient of two. Thus, the promoter activity obeys the following functional relationship:

$$p_{uhpT} = P_{max} \cdot \frac{[g6p]^2}{K_d^2 + [g6p]^2}$$

where [g6p] is the extracellular concentration of glucose-6-phosphate, P_{max} is the maximum promoter activity and K_d is the dissociation constant of UhpA∼P for its binding site at p_{uhpT}. This equation will be used to predict the activity of the promoter as a function of the concentration of g6p. The promoter activity, in turn, will yield information about the amount of DnaA proteins in the cell. Mathematical equations are also used to describe cell growth and the internal concentration of proteins, such as DnaA.

2.4 Writing equations

The cell is a complex system comprising many component that interact dynamically. Protein expression depends on the number of mRNAs and RNA polymerases. Their number, in turn, depends on the number of RNA polymerases that interact with DNA to initiate transcription. This in turn depends on the number of chromosomes in the cell. To make matters even more

complicated, RNA polymerases and ribosomes being proteins, they affect their own expression.

In summary, protein production is conditioned by a sequence of events ruled by probabilities making the cell a stochastic system. This makes it very difficult to know the exact number of proteins in a single cell, even when the average number of a population of cells is known. At the population level, we can write an equation describing the average behavior of the population by balancing production and degradation terms.

In the case of DnaA we can write the following equation:

$$\frac{d\,[DnaA]}{dt} = p_{dnaA}(t) - [DnaA]\,(\gamma_d(t) + \gamma_i) \tag{2.1}$$

DnaA is expressed by the cell at a rate equal to $P_{DnaA}(t)$. The production rate can be measured by combining promoter activity and translation efficiency, and can be expressed in Miller Units, as was done with p_{uhpT} [1].

The degradation term contains two components: the half-life of the protein, described by the degradation constant γ_i, and the dilution of the intracellular protein concentration due to growth dilution, described by the growth rate, γ_d. The growth rate γ_d can be measured experimentally and is constant during exponential growth. γ_i takes into account the spontaneous degradation of DnaA molecules, as well as active degradation by proteases. In the case of *E. coli*, γ_i is much smaller than γ_d during exponential growth since DnaA is a stable protein and there have been no reports on proteases specifically targeting this protein in *E. coli*.

This equation can thus be used to estimate of the intracellular concentration of DnaA. In order to exploit the model and make predictions about the system, we need to determine the parameters involved. A considerable part of the work of this thesis consists in estimating these parameters.

2.5 Motivation and Goal

Motivation

There have been many studies of DNA replication or the DnaA protein of *E. coli*. However, most of them deal with DNA replication as a global phenomenon, or the global role of DnaA in this process [37]. While the effects of over-expressing DnaA have been studied by several groups [32, 31], the effects of low concentrations of DnaA are much less understood. The precise number of DnaA molecules needed for DNA replication has not yet been determined experimentally and the possibility of restarting DnaA production after a period of arrest have never been explored.

Arresting cell division by decreasing the production of DnaA is interesting for fundamental research as well as industrial applications. For fundamental research, quantifying the number of DnaA molecules needed to initiate DNA replication and cell division will add valuable information to existing models of these processes, while industrial applications could benefit from a novel way to control resource allocation, directing resources away from biomass production (i.e., cell division) and toward the formation of a biotechnological product.

E. coli is considered in the industry not so much as a source of biomass, but rather as a microscopic factory that can produce molecules of interest [86, 87, 88, 89]. A drawback of using cells in such a way is that most of the resources given to the culture will be used for biomass

production, cells will grow and divide. However if cell division is impaired, while the metabolic activity of the cell is maintained or better yet, optimized for protein or metabolite production, the efficiency and usefulness of this approach is greatly increased. Proofs of concept of this approach already exists. For example, stopping the synthesis of RNA polymerase in *E. coli* maximizes the yield of an artificial pathway producing glycerol [19]. In this study, cell division was also arrested by limiting the production of RNA polymerase. However, DNA replication remained functional. This leads us to believe that by stopping DNA replication it should be possible to further increase metabolic production. Going even further with this reasoning, if cell division could be switched on and off at will, this would also reduce the frequency at which an industrial bioreactors would have to be emptied and restarted, further increasing the performance of the process.

Creating a strain, the growth of which can be stopped for a certain amount of time, is also of great interest to fundamental research. One of the reasons that make cell biology such a complex field is that many reactions and metabolic pathways constantly change and interact with each other. Thus, trying to identify the main factors that drive a cellular activity, or even just trying to identify the proteins involved in a given pathway, becomes a challenging task. Removing one important process, cell division and DNA replication, might facilitate studies by eliminating one global confounding process.

Goal

The goal of this thesis is to quantify the influence of varying concentrations of DnaA on DNA replication, cell division, and viability. We focus on characterizing the effects of *limiting* concentrations of DnaA. We construct a new inducible promoter, based on the *uhp* system, and use it to externally control the production rate of *dnaA*. In other words, we aim to estimate the number of DnaA proteins required to initiate DNA replication and cell division. We will also evaluate, in a quantitative manner, the effects brought about when cells stop dividing due to DnaA depletion. This requires a thorough characterization of the *uhp* induction system we use. We need to understand the number and nature of the interactions between the proteins that make up this expression system.

In order to attain our goal, and due to the relatively small number of molecules involved in the induction system and the system controlling DNA replication, we need to combine population analysis, single cell observation and mathematical modeling. In more detail, our objectives are:

- Modify *E. coli* in order to use p_{uhpT} as an inducible promoter of the *dnaA* gene.
- Quantitatively characterize the average behavior of both the *uhp* expression system (promoter strength as a function of inducer concentration, ...) and the physiology of our modified strain (viability, growth rate, ...).
- Construct a quantitative model of signal transduction by the *uhp* expression system, in order to obtain a predictive tool.
- Use single cell observations, utilizing time lapse microscopy and microfluidics among other techniques, to study the behavior of individual cells when their DnaA concentration is low.

- Explore the possibility of freezing cell division without blocking the rest of the metabolism.
- Extract numerical values for the concentration of DnaA in different growth regimes and determine the threshold value of the concentration of DnaA below which replication ceases.

Due to the prescribed duration of a PhD thesis, we limited our objectives to the list above. However, extensions of the project include the addition of a model of DNA replication, the construction of additional control modules, the exploration of the connection between the arrest of DNA replication and metabolism, and many others. These perspectives will be discussed in the last chapter of this thesis.

Part II

Experimental techniques and methods

Chapter 3

Experimental techniques and setups

In this section, we will describe the different methods and protocols that we will use for data acquisition, data analysis and molecular biology. In addition to the simple description of the techniques, we will point out their advantages and limitations. Microfluidics and time lapse microscopy experiments allow us to study the details of our cells, while micro-plate experiments give us a broad view of how the population behaves.

.

3.1 Experimental setups and cell growth

Microfluidics setups are good tools for studying single cells, but they are not the only ones. There are other systems that also allow the study of a large number of single cells by time lapse microscopy, such as agar pads. We will use both systems in this work.

3.1.1 Microfluidics systems

In order to monitor bacterial growth and physiology, the microfluidics system needs to trap cells so they can easily be observed individually without hampering their growth. This type of device is called a "mother machine". A schematic of the device is shown in Figure 3.2. The device is fabricated by polymerizing a mixture of PolyDiMethylSiloxane (PDMS) and a crosslinker in a ratio of 10:1 as specified in the mother machine handbook [90]. The mixture is cast into a mold that has the complementary shape of the device. The device is baked at $65°C$ for 1 hour in order to solidify and is then cut out in the form of a small $2\,cm^2$ rectangle as shown in Figure 3.1. Next, the device is punctured; two holes are made through the central channel to allow connection to the bottle containing the growth medium and the waste by way of small tubes.

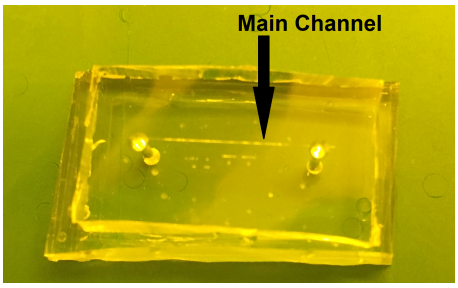

Figure 3.1: Photo of our microfluidics PDMS device before being bonded to a microscopy slide. *The central channel is visible as well as the two holes used to connect to the medium intake and waste.*

The devices need to be washed with pentane once and acetone at least twice for two hours to clear any monomers left, as they can be toxic. The washing protocol, as well as the exact method of fabricating the device, is detailed in the mother machine handbook [90].

Cleaned devices are then bonded onto clean microscopy slides. Cleaning and the preparation for bonding the two parts is done by a plasma cleaner (Harrick Plasma plasma cleaner and plasmaflo PDC-FMG-2), which does an oxygen plasma treatment on both surfaces allowing them to be bonded together. In order to form the plasma, a weak oxygen flow traverses the chamber containing the device and microscopy slide. When the chamber is filled with oxygen at a pressure of 0.7 mbar, the plasma treatment is run for 40 seconds (medium plasma settings), allowing both surfaces in contact with the plasma to be active in order to make strong bonds with each other. The device bound to the slide is then put into an oven at $65°C$ for 20 minutes to finish the process.

The device turns hydrophobic some hours after the oxygen plasma treatment. Therefore, once it is out of the oven, a solution of bovine serum albumin (BSA, 50 $mg \cdot ml^{-1}$) is injected into the device. This fills the channels with liquid that will be easier to replace later with cells. At the same time, BSA will coat the internal walls of the device to avoid bacterial adhesion to the walls. The solution stays in the device for one hour, after which a concentrated solution of bacteria is injected. The cells will naturally enter the channels, as long as the concentration of cells is sufficiently high (40 ml of stationary phase cells concentrated to a final volume of 0.2 ml) and the two entries to the central channel are blocked (with adhesive tape, for example).

The final device possesses 4 000 channels with a spacing of 3 μm. Each secondary channel has a length of 25 μm and a width of 1 μm. The central channel is 2 cm long and 100 μm wide. These specifications are summarized in Figure 3.2.

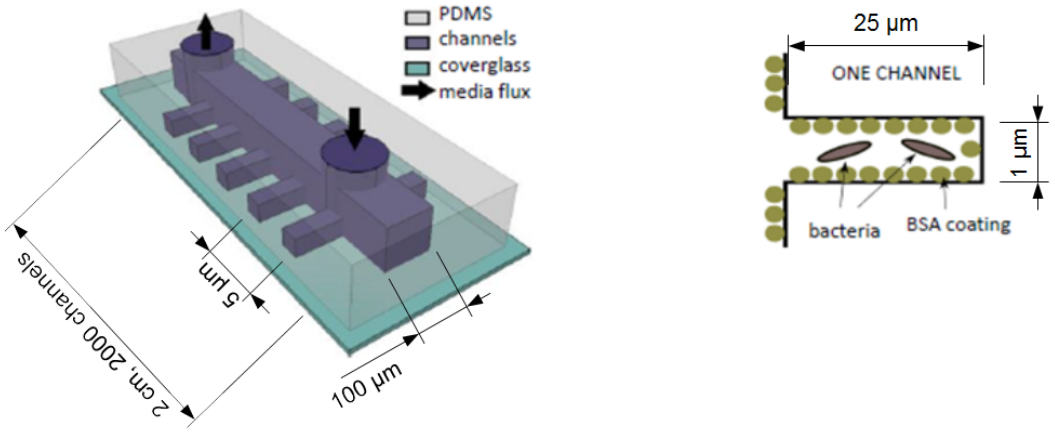

Figure 3.2: Custom mother machine. *The PDMS device is glued to a coverslip using an oxygen plasma. Our central channel has a length of 2 cm and a width of 100 µm. There are 2 000 channels on each side of the central channel, each 25 µ long and 1 µm* wide.

The device with the cells is then placed in an inverted microscope at a temperature of $37°C$, as shown in Figure 3.3 B.

3.1.2 Microscopy setup

Our microscope is a motorized, inverted microscope (Zeiss Axiovert 200M) with a phase contrast objective lens (Zeiss PlanNeofluar, Ph3 100x/1.3). The entire microscope is placed inside a thermostated box, keeping all components at a constant temperature of 37°C. For fluorescence imaging, we use a mercury lamp (Osram, 1xHBO 103X/2) as an excitation light source. Using narrow bandpass excitation and emission filters, we choose the appropriate wavelengths. For the current work, we only used the two cubes corresponding to red fluorescence and green fluorescence (Chroma, #49002 ET-GFP and Chroma, #49005 TR/DsRED ET). To avoid illuminating our cells for too long and inducing an unwanted stress, we used mechanical shutters (Uniblitz-VS35) to keep the light exposure time to a minimum. A photo of our microscope is shown in Figure 3.3 A.

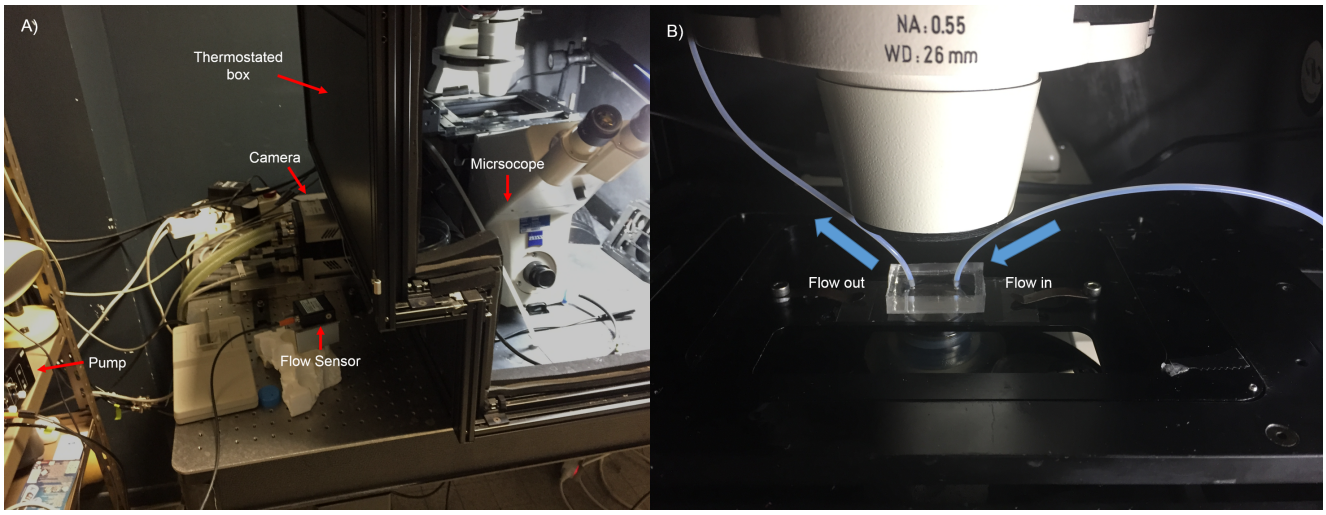

Figure 3.3: A) Photo of our microscopy setup. *The microscope is placed inside a thermostated, black box. A halogen and mercury lamp provide light sources (both located behind the box). The miscoscope is also connected to controllers of the mechanical shutters (not shown), themselves controlled by a computer (not shown). The camera islinked to the microscope, the computer, and a cooling system (not shown). The photo also shows our flow sensor connected to a pump and controlled by the computer. All these elements are used for a* microfluidics *experiments.* **B) Photo of the device placed under the objective.** *The microfluidics device is connected by tubing to the in- and out-flow reservoirs.*

The principle of time lapse experiments is to let the sample grow under the microscope in a suitable environmental conditions. In our case, this involves as little light exposure as possible, a stable temperature of $37°C$, and little to no external perturbations. For this reason, our microscope is not only placed in a thermostated box, but it is also connected to a computer and controlled externally by the Winview software (Princeton Instruments). Our program allows us to move the sample inside the microscope as well as take images without having to open the thermostated box. In our experiments, we took 20 to 40 frames in both phase contrast and fluorescence mode every 10 minutes for extended periods of time, from several hours up to three days. The images were acquired with a 16-bit grays level CCD camera cooled to -80 $°C$ (Roper Scientific, Princeton Instruments PHOTOMAX 512), also controlled by the Winview software.

The big advantage of using a mother machine device is that the cells are organized in the channels, which allows us to monitor many cells in each frame without any interference from overlapping cells. We are able to measure up to 15 channels per frame with $\sim$10 cells per channel during normal exponential growth. Even if we analyzed only one cell per channel, we could trace several hundreds of cells over many generations, giving us enough data to generate trustworthy statistics in a dynamical environment.

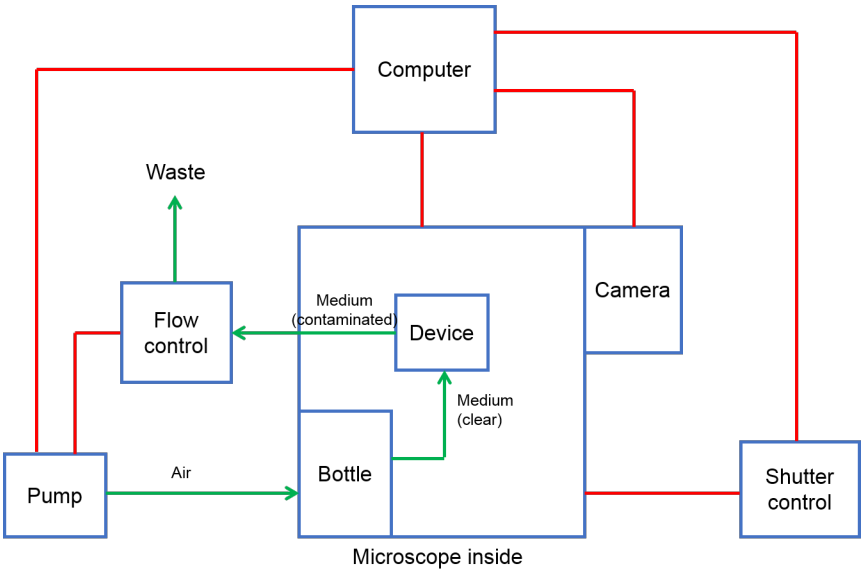

Figure 3.4: Full diagram of the microfluidics setup. *The computer controls the camera, pump and microscope, and the shutters. The pump collects a feedback from the flow monitor to ensure a constant flow of medium. Both the medium bottle and the device are inside a thermostated box in order to keep cells growing at an optimal temperature. The flow of air goes into the medium bottle from the pump, pushing the growth medium into the device. After traversing the device, the contaminated medium, containing cells that are expelled from the channels, exits the microscope, passes through the flow monitor to maintain a constant flow of* $20\mu L \cdot min^{-1}$*, and is collected in the waste* bottle.

The microfluidics device is connected to a bottle that supplies the sterile growth medium. By changing the bottle, we are able to change the conditions of the cell environment without stopping the time lapse measurements. The bottle itself is also inside the thermostated chamber so that the medium remains at a constant temperature. The medium in the bottle is prone to contamination, so all connections that go into the bottle have a 0.2 μm filter at their extremities to ensure sterility. As for the microfluidics device, its inlet is connected to the bottle and the outlet goes through another tube, past a flow controller and into a waste bottle.

To force the medium though the tubes, we use a pump (ELVEFLOW OB1 pressure controller), located next to the thermostated box, connecting to the bottle of growth medium via tubing and a filter. The bottle containing the medium is tightly sealed, with an inlet connected to our pump and the outlet connected to the microfluidics device. Thus, by pumping filtered air into the bottle, the increase in pressure forces the medium to go though the outlet into the device. The flow is directly proportional to the pressure exerted by the pump.

In order to have a constant flow all over the experiment we use a flow regulator (ELVEFLOW FLOW-03-3S), which is connected to the pump. Before the medium goes into the collection tube after having left the device, it goes through our flow regulator, which measures the flow speed and gives a feedback to the pump in order to increase/decrease the pressure as needed. The full setup is summarized in the diagram of Figure 3.4.

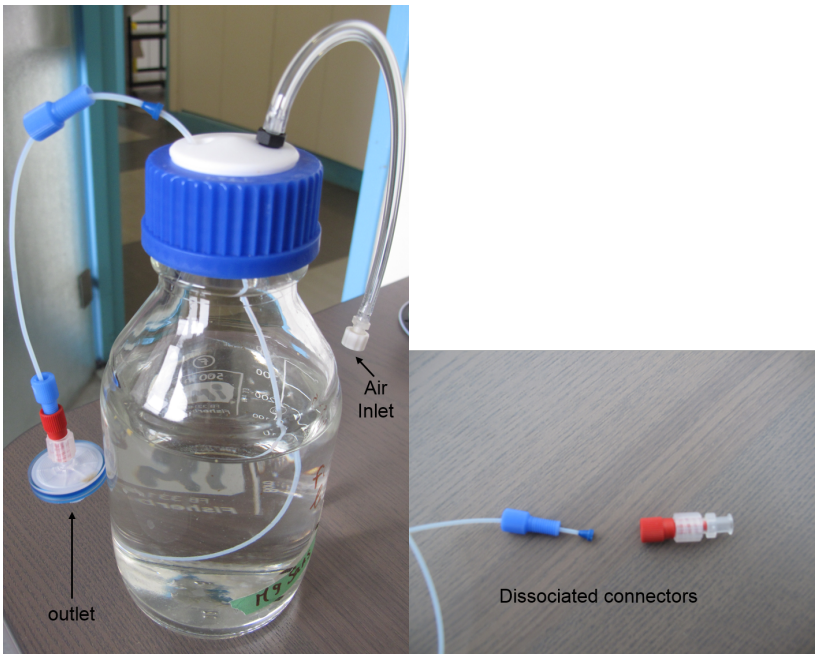

Figure 3.5: Bottle for microfluidics experiment and connectors. *The air going into the inlet passes through a 0.2 µm filter, equivalent to the one located in the outlet tubing. During experiments, both connectors are tightened on the bottle cap to ensure sterility.*

It is very important that the bottle and its contents are kept sterile during the entire experiment. To ensure the sterility of the bottle, it is autoclaved along with the lid and the outlet tubing (Figure 3.5). The assembly is wrapped in aluminum foil to ensure sterility after going through the autoclave and then put inside a sterile, laminar flow cabinet, where the connectors that seal the bottle, the inlet tube, the filters and the growth medium are added or connected. At this point, the bottle is ready to be connected to the pump and the device (Figure 3.5). This is a key step, since contamination of the medium can compromise the entire experiment. We have to be particularly careful because we add g6p to our medium in order to enable DNA replication. This phosphorylated sugar is metabolized by many other organisms.

Coupling the mother machine with the microscope provides us with a stable system, a large amount of single cells, and the possibility to change their environment without stopping the observation. However, the analysis of the data is often long and difficult since we do not have a dedicated program to analyze cells with abnormal growth behavior. Therefore, other types of experimental setups are also useful, even if the amount of data recorded is smaller. Experiments that did not require the observation of transitions between states were carried out using agar pads, which still allows observation at the single-cell level.

3.1.3 Agar pads

Agar pads consist of a small, rectangular gels made of agar as the gelling agent, supplemented with growth medium, including inducers or other components [91]. Such agar pads can sustain bacterial growth in a similar way as a a Petri dish. The agar pad is placed on a microscope slide, placed under the microscope. Time-lapse experiments consist in observing the agar pad in the same way as the microfluidics device.

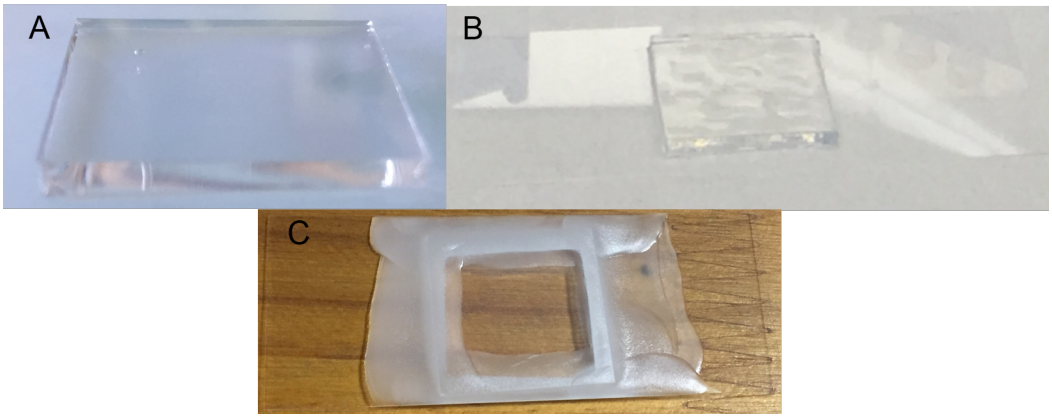

Figure 3.6: Agar pads. A) *1 ml of MOPS medium + 1.5% agar solidifying between two 22 mm cover slips.* B) *Solidified pad on microscope slide with cell droplets drying on its surface.* C) P*ad loaded with cells, sealed with wax and a 20 mm cover slip on a microscope slide, ready to be observed under the microscope.*

Agar pads are made by mixing low melting agar with growth medium and heating the mixture in a microwave until the agar completely dissolves. Afterwards, additional moleculesm such as inducers can be added and the mix is poured between two microscope cover slips in a sterile atmosphere. After 45 minutes of solidification, small droplets of a cell culture are added onto the surface of the pad. The optical density (OD) of the culture must be between 0.08 and 0.2, otherwise there will be too many cells that will very quickly overlap. The agar pad is left to dry for 15 minutes, covered with a cover slip, and sealed entirely with wax to stop it from drying too quickly. The key steps are show in Figure 3.6.

Observing cells with a pad requires a simpler setup than the one necessary for a microfluidics experiment. Furthermore, the analysis of the data is easier and quicker. In this setup, no cells are lost, which is an advantage when counting the total number of divisions and using cell analysis programs. However, this characteristic is also a limitation, since in the long run, cells will overgrow the field of view of the microscope, making it impossible to distinguish individual cells. Also, it should be noted that an agar pad is a closed system. The growth conditions can not be changed during the experiment. This restriction make agar pads less flexible than a microfluidics device. For the same reason, agar pads are much less prone to contamination of the growth medium during the experiment.

Agar pads can be prepared with all kinds of culture media. We used MOPS minimal medium with 0.2% glucose in our experiments.

3.1.4 Growth Media and incubators

Choosing the growth medium, such as the above mentioned MOPS, is one of the first steps for any microbiology related project. The growth medium can be liquid or incorporated into a gel, such as an agarose gel. Growth media can be rich in nutrients or only provide the minimum necessary for cells to grow. In our experiments, we mostly use minimal media because of their well defined composition. Growth medium are either filtered after being prepared or autoclaved in order to ensure sterility.

In this thesis, we used three different types of media:

- Lysogeny broth (LB), which is a rich medium with no added glucose. This medium allows *E. coli* to grow very rapidly with a doubling time of ∼20 minutes. The recipe is very simple (Table 3.1). However, since the yeast extract is different from one batch to another, two different preparations of LB are never truly the same. This makes it inappropriate for quantitative observations. Nonetheless, the medium is useful for growing cells quickly, which is handy for experiments of genetic engineering.

- M9 minimal medium only contains the mineral salts necessary for growth. As the only carbon source, we add 0.2% glucose. This medium requires more ingredients than LB and *E. coli* grows more slowly, with a doubling time of ∼60 minutes. M9 is a defined medium and the concentrations of all components are known. This reproducibility is important for quantitative experiments. The M9 minimal medium uses high concentrations of phosphate as the buffering agent for keeping the pH at ∼7.5, ideal for our cells. The recipe is shown in Table 3.2.

- MOPS minimal medium only contains mineral salts, similar to M9. As the sole carbon source, we add 0.2% glucose. This medium is a defined medium, similar to M9 and bacteria grow with a doubling time of ∼60 minutes. The main difference between MOPS minimal medium and M9 minimal medium is the buffering agent. MOPS (3-(N-morpholino)propanesulfonic acid) is the buffer of the eponymous medium. The phosphate concentration is thus much lower in MOPS medium than in M9 medium. This is important, since the activation of the *uhpT* promoter (p_{uhpT}) is inhibited by phosphate. The components of this medium are listed in Table 3.3.

In addition to the liquid media, we also used gel versions of these media. Petri dishes of LB and MOPS were prepared by adding between 1 and 1.5% of agar to the liquid medium. The microscopy pads were prepared by adding 1.5% of low melting agar to our MOPS medium.

Petri dishes and culture tubes are placed inside incubators at a suitable temperature to allow cells to grow. In most cases, this temperature is 37°C. Culture tubes also require to be shaken at 200 rpm in order for the cells to be properly oxygenated. A container with some water is also present inside the incubators to reduce drying of the solid agar media.

Product	**Quantity**
NaCl	10 g
Tryptone	10 g
Yeast extract	5 g
Distilled water	1 L

Table 3.1: Recipe for 1 L Lysogeny Broth.

M9 + 0.2% glucose	
$CaCl_2$, 1 M	5 μL
$MgSO_4$, 1 M	100 μL
20% glucose	750 μL
1000x trace elements	45 μL
$FeSO_4$ solution	5 μL
1% Thiamine	25 μL
5x Salts	10 ml
H_2O	to 50 ml

5x Salts to autoclave	
H_2O	10 ml
$Na_2HPO_4 \cdot 2H_2O$	425 mg
KH_2HPO_4	150 mg
NaCl	25 mg
NH_4Cl	50 mg

1000x trace elements	
H_2O	200 μL
$Na_2EDTA \cdot 2\,H_2O$	100 μL
$ZnSO_4 \cdot 7\,H_2O$	100 μL
$CaCl_2 \cdot 6\,H_2O$	100 μL
$MnCl_2 \cdot 4\,H_2O$	100 μL
H_3BO_3	100 μL
$Na_2MoO_4 \cdot 2\,H_2O$	100 μL
$CuSO_4 \cdot 5\,H_2O$	100 μL

$FeSO_4$ solution	
H_2O	1 ml
$FeSO_4$	30 mg

20% glucose	
H_2O	10 ml
D-glucose	2 g

Table 3.2: Recipe for 50 ml M9 minimal medium.

MOPS + 0.2% glucose	
10x MOPS mixture	100 ml
K_2HPO_4, 132 mM	100 μL
20% glucose	10 ml
H_2O	880 ml
1% Thiamine	100 μL

Micronutrient stock	
$(NH_4)_6\,Mo_7O_{24} \bullet 4H_2O$	9 mg
H_3BO_3	62 mg
$CoCl_2$	18 mg
$CuSO_4$	6 mg
$MnCl_2$	40mg
$ZnSO_4$	7 mg
H_2O	to 50 ml

10x MOPS mixture	
$FeSO_4 \bullet 7H_2O$, 0.01 M	10 ml
NH_4Cl, 1.9 M	50 ml
K_2SO_4, 276 mM	10 ml
$CaCl_2 \bullet 2H_2O$, 20mM	0.25 ml
$MgCl_2$, 2.5 M	2.1 ml
NaCl, 5 M	100 ml
Micronutrient stock	200 ml
Autoclaved miliQ H_2O	387 ml

20% glucose	
H_2O	10 ml
D-glucose	2 g

Table 3.3: Recipe for 1L MOPS minimal medium.

3.1.5 Growth and washing protocols

For all our experiments, we grow a dense culture of cells overnight in the incubators. This dense culture is referred to as "pre-culture" and its only goal is to provide a high initial cell concentration before quantitative experiments. Cells in pre-cultures are in stationary phase, having consumed all the available carbon sources. In this state, the bacteria need some time to adjust to another

medium and start growing again. This is the case even if the new medium is the same as the one used for the pre-culture. The drawback of using pre-cultures is that the metabolic state of the cells and the protein concentration in their cytoplasm can vary from one pre-culture to another. Nonetheless, after 3-4 divisions, this variability becomes negligible.

We either used the preculture directly or, for more precise quantitative experiments, we highly diluted the preculture to an OD of $1 \cdot 10^{-5}$ to $1 \cdot 10^{-6}$ and let it grow in our incubators overnight. This way we had cells growing exponentially the next morning at an OD of $\sim 5 \cdot 10^{-2}$.

We sometimes needed to transfer cells from one growth medium to another. In such cases, we first washed our cells by centrifuging 1 ml at 11 000 r.c.f. for one minute. The cells were then re-suspended in 1 ml of the new medium. We repeated this process a second time before measuring the OD of the cells and diluting them to the desired concentration with the new medium.

When performing population experiments, we first diluted the cells into fresh medium, usually to an OD of $5 \cdot 10^{-4}$ to $5 \cdot 10^{-3}$, depending on the experiment. The first generations are below the detection limit of the absorbance measurements, but this pre-growth ensures that the bacteria have reached a steady state during the subsequent measurements in case washing them or changing their growth medium affected their initial growth.

3.1.6 Microplate reader

We use microplate readers in order to measure the growth and gene expression of several independent cultures of cells simultaneously. The plates we use have 96 identical wells arranged in 8 rows and 12 columns (Figure 3.7). Each well can hold up to 200 μL of cell culture, leaving about half of the total volume of a well unused in order to ensure sufficient aeration of the culture. The plates are purchased from Thermo Fisher Scientific (NUNC 165305) and have a transparent lid.

The plate reader (Tecan Infinity Pro M200) keeps the plates at a constant temperature and shakes the microplates either with an orbital or linear motion. We grow *E. coli* in the same conditions as in the incubator. The plate reader periodically measures the absorbance or/and fluorescence of each well and records the values.

Our shaking and measurement settings allows for optimal cell growth at 37°. We shake the plates for 120 seconds with an amplitude of 3 mm and alternate between linear and orbital modes. For an optical density (OD) below $0.3\,cm^{-1}$ we observe growth rates equivalent to the ones in the incubators. Beyond this concentration, the cells begin to aggregate at the bottom of the well, introducing a bias into the absorbance measurements.

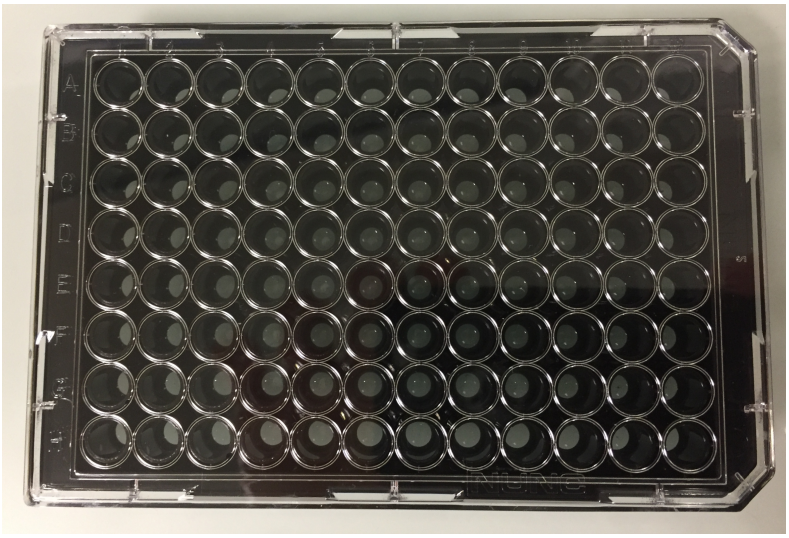

Figure 3.7: 96 well micro-plate. *Thermo Fisher Scientific (NUNC 165305) with transparent lid, used for our measurements in the Tecan plate reader.*

The plate reader is one of the most widely used tool for carrying out experiments at the population level. Using the same conditions on several wells is equivalent to repeating an experiment. This allows us to save time and acquire data more rapidly. Coupling population data with time lapse microscopy measurements, using both our mother machine and agar pads, allows us to observe single cells and link these observations to population measurements. These complementary techniques produce a rather comprehensive picture of the kinetics of cellular processes at the population level and in single cells.

3.1.7 Data processing

The data from the plate reader are raw measurements of absorbance and/or fluorescence, obtained at regular intervals for each of the 96 wells. In order to remove outliers or experimental errors, we systematically used between six and thirteen independent wells to study a single set of conditions. An example of a set of wells representing the same condition is shown in Figure 3.8. We therefore always had at least six data sets for calculating the average value and its standard error. The large number of replicas also allowed us to spot abnormal data points. On each plate, we included between four and eight wells containing only the growth medium. These control wells serve both as a control for contamination and as a background signal, which was subtracted from the final data sets.

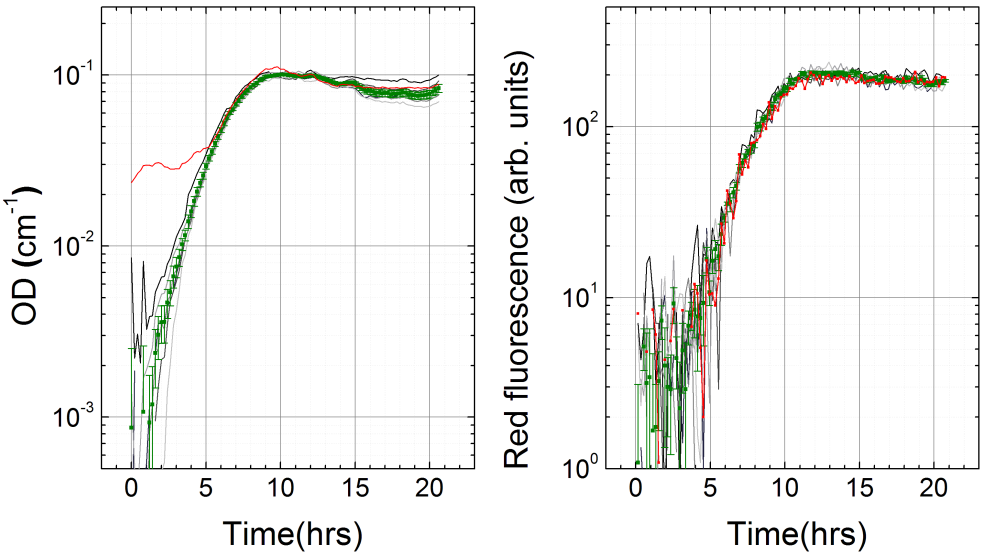

Figure 3.8: Example of a set of seven individual wells from the microplate reader. *Left: the OD measurements and Right : the corresponding red fluorescence. The grey and red lines are the individual well measurement. Visual inspection identifies one well (the red line) as an outlier in the OD measurements, but not in the red fluorescence. This abnormal measure has been excluded from the estimation of the average and standard error of the mean (dots and error bars in green).*

Once the average and standard error of the mean (represented by error bars) of one condition are determined, we can process the data in order to compare the results of different experiments. The initial cell concentration in different experimental conditions are usually slightly different, even when we intended otherwise. These discrepancies were more pronounced after the cells were washed and re-diluted. In order to be able to compare two curves where g6p was removed at time zero, we compensated for this variation in the initial number of cells by realigning the curves using a multiplicative correction factor. On the same plate, we measured replicas of an initial culture (the same tube, therefore the identical concentration); one part of the culture containing g6p (which leads to normal, exponential growth), the other part lacking g6p (leading to growth arrest after run-out of DnaA). The correction factor was extracted (see Figure 3.9) by comparing the the measurements derived from the same initial sample. The exponentially growing cells were aligned by the multiplicative factor (a shift in the y-axis of a logarithmic plot. Figure 3.9 shows an example of how we re-aligned two curves using an OD signal. This same correction factor was then applied to the data where growth is limited by the removal of g6p at time zero. In principle, the growth-limited data could have been mutually aligned on their exponential part. However, in practice, there are too few data points for a reliable alignment. Once determined, the multiplicative correction factor for aligning the data from two experiments can be applied to all measurement types. For example, determining the multiplicative factor using OD measurements allowed us to also align fluorescent plots of the same wells.

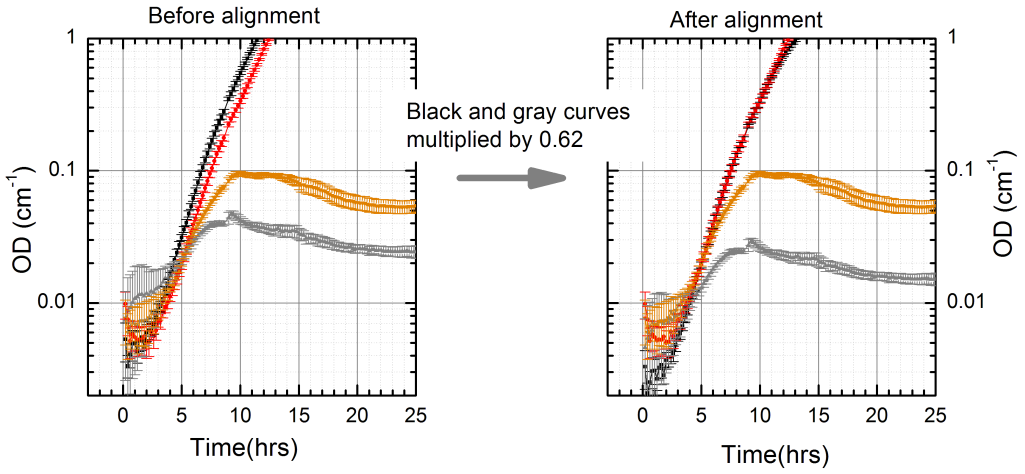

Figure 3.9: Alignment of measurements from the microplate reader. *Left: the four measurements (black, gray, red and orange) should contain the same initial cell concentration. However, experimental errors lead to somewhat different initial conditions. The black and gray curves originate from the same preculture, but one contains g6p, leading to normal, exponential growth (in black), and the other one does not (in gray). The same happens for the two other conditions, respectively red and orange. In order to compare the significant parts of the experimental curves with limiting g6p (their maximum values for example), we need to multiply the black and gray curves by 0.62 in order to shift them on the y-axis (logarithmic scale). The correction factor is determined from the alignment of the exponentially growing cells (black and red curves).*

Chapter 4

Molecular biology methods

The molecular biology techniques used in this study aim at modifying the genome of *E. coli* to either control the expression of already existing proteins, such as DnaA, or to introduce new genes, such as those coding for fluorescent proteins. We used PCR (Polymerase Chain Reaction) to synthesize the DNA fragments and verified them using gel electrophoresis and sequencing. The latter was done by the Eurogentec corporation. When we were satisfied with our fragments, we inserted them into the cell by electroporation or TSS transformation [92].

There are two ways to make cells express genes they do not normally have, i.e., trans-genes.

The first possibility is to clone the gene on an autonomously replicating, extra-chromosomal, circular, double strand DNA, called a plasmid. A plasmid, once transformed into the cell is maintained and replicated independently of the chromosome. The genes on the plasmid are expressed as any other gene on the chromosome. However, contrary to the chromosome, most plasmids are present in the cell in multiple copies.

The second possibility consists in directly modifying the chromosome of *E. coli* and to remove, substitute or add new DNA sequences. This also requires DNA stands to enter the cell, and to integrate into the chromosome. Integration into the chromosome happens via homologous recombination, i.e., the entering fragments need to share part of their sequence with identical sequences present on the chromosome. Natural recombination in *E. coli* is relatively inefficient. In order to improve the success rate of this reaction, we express additional proteins that catalyze efficient homologous recombination.

4.1 Plasmids

Plasmids are circular DNA sequences that possess an origin of replication. They exist as independent units inside the cell and replicate along with the chromosome of *E. coli*. Plasmids contain three essential elements (Figure 4.1):

1. **An origin of replication.** This part of the plasmid contains all functions necessary for its replication, independently of the replication of the chromosome. The origin of replication determines the average number of plasmids per bacterium. Some plasmids have low copy numbers, ranging form 1-10 copies per cell, others, such as plasmids carrying the *colE1* origin of replication can have ∼50 copies per cell. High copy number plasmids can be

present in the several hundreds of copies per cell. For some of the experiments we use a temperature sensitive version of the origin of replication pSC101. A replication protein, RepA, is part of this origin of replication. The temperature sensitive version of the plasmid contains a temperature sensitive RepA. The protein is unstable at high temperatures and degraded by the cell. This allows eliminating the plasmid by growing the bacterial culture at elevated temperatures. For example, the pSIM5 plasmid, which we will use for providing the lambda Red recombination functions, only replicates at temperatures below 37°C.

2. **A selection cassette.** This part of the plasmid codes for a constitutively expressed gene conferring the resistance to an antibiotic. By selecting for the resistance to the antibiotic carried by the plasmid, we can easily verify that the bacteria contain the plasmid.

3. **A gene of interest**, along with a suitable promoter. In the case of an expression plasmid, we want to clone the gene to be expressed behind an inducible promoter. In the case of a reporter plasmid, we want to assess the activity of a promoter by monitoring an easily measurable signal. In this case, the promoter under investigation generally directs the transcription of a fluorescent protein.

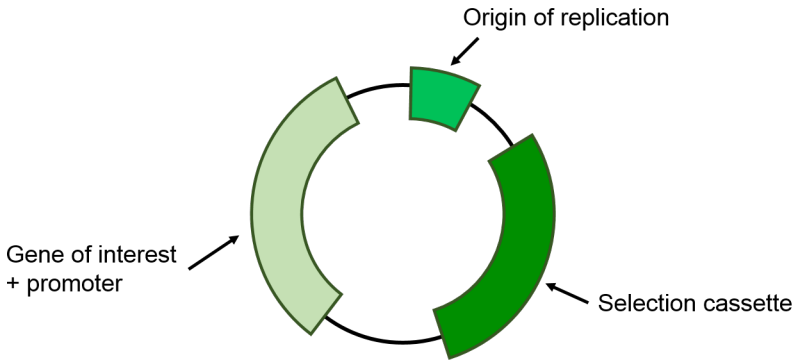

Figure 4.1: Diagram of a Plasmid. *Plasmids posses three main components. An origin of replication, which ensures the replication of the plasmid independently of the replication of the chromosome. An antibiotic resistance gene, which is used for selection. A sequence of interest, which usually comprises one or more genes and promoters that we want to study.*

Most of the plasmids we use confer resistance to the antibiotics ampicillin or chloramphenicol. The list of plasmids used in this study is shown in Table 4.1.

Plasmid	Origin	Resistance	Relevant genotype
pSIM5	*pSC101 ts*	Cm^R	*gam exo bet*
pGLYC	*colE1*	Amp^R	*gpd1 gpd2*
pGFPR01	colE1	Amp^R	P_{BAD}- pH sensitive gfp (GenBank AF058694.2) [93]

Table 4.1: Plasmids used in this study

Genetic engineering using plasmids is rapid, because the DNA can be assembled *in vitro* using molecular biology techniques. However plasmids have their drawbacks when used in quantitative analysis. The most evident problem of plasmids is that they only introduce new gene and

promoter combinations into the cell. Often, it is interesting to remove genes from the cell. Another drawback of plasmids is that they introduce additional variability if their copy number is low. Even though during cell division, the plasmids will be distributed more or less evenly between the two daughter cells, the partition is random and will therefore follow a binomial distribution. Plasmid replication also involves some stochasticity, because the origin of replication only determines the *average* copy number of the plasmid in the cell. The deviations from average cause further differences between cells. In the end, the number of plasmids, and therefore the copy number of genes carried by the plasmid, will be different between cells, thus introducing significant variability due to plasmid "noise". On the other hand, if the number of plasmids in the cell is too high, the behavior of the genetic system under investigation may change due to the high number of extra binding sites for transcription factors. For quantitative analyses, it is therefore preferable to modify the chromosome.

4.2 Homologous recombination

To modify the chromosome, we use homologous recombination. This process, present in almost all organisms, is based on base pair recognition of complementary DNA strands. The recombination machinery "exchanges" homologous DNA strands. The foreign DNA to be inserted into *the E. coli* chromosome must bear sequences that are identical to sequences on the chromosome. The recombination machinery catalyzes the strand exchange with a certain probability.

The natural recombination frequency of *E. coli* being relatively low. We therefore express recombination enzymes derived from the bacteriophage λ to make the process more efficient. These enzymes catalyze the recombination of a linear fragment of DNA with the chromosome, provided that at least 50 bp at each end of the fragment are homologous to the target sequence on the chromosome. The reaction is summarized in Figure 4.2.

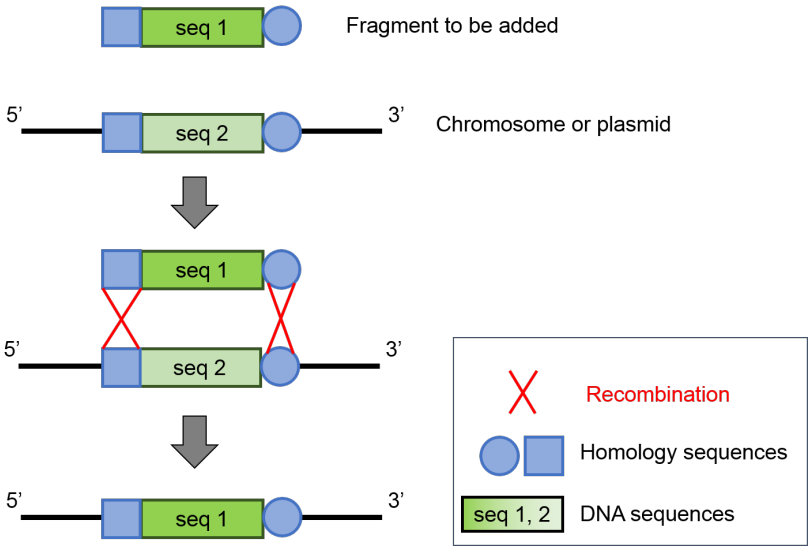

Figure 4.2: Genetic engineering of the chromosome by homologous recombination. *Two homologous regions can recombine, thereby swapping the intervening regions. The homologous sequences are aligned by the recombination machinery and cellular enzymes replicate the chromosome by switching from the chromosome template to the incoming fragment and back. The result is the exchange of seq2 on the chromosome by seq1, initially on the linear fragment.*

4.2.1 The lambda red system

The natural frequency of recombination being low in *E. coli*, we express dedicated recombination enzymes to make the process more efficient. The λ_{red} recombination system encodes three bacteriophage genes that were cloned onto the plasmid *pSIM5*. These genes, *gam*, *exo*, and *bet*, code for proteins that strongly enhance the recombination reaction. The function of these genes are:

- The Gam protein inhibits the *E. coli* nucelase RecBCD, which would otherwise rapidly degrade the incoming, linear DNA fragment.
- The Exo protein possesses a 5'→3' exonuclease activity. When acting on a double-stranded DNA fragment, this activity leaves single-stranded 3'-regions at both extremities of the fragment. The single-stranded regions are the recombinogenic substrates that can pair with transiently single-stranded regions of the chromosome. During DNA replication, the lagging strand remains transiently single-stranded and highly recombinogenic.
- The Beta protein promotes the annealing of two complementary single stranded DNA sequences.

We only want these recombination functions to be active during the genomic engineering reaction. Having permanently an increased frequency of recombination would jeopardize the integrity of the chromosome, leading to unwanted chromosome re-arrangements. For this reason, the λ_{red} genes on *pSIM5* are induced at 42°C just before electroporating the target fragment into the cell. The plasmid itself is temperature sensitive and will only replicate at temperatures below 37°C. After the recombination reaction, we grow the recombinant strains at 37°C to eliminate the plasmid.

The protocol is described in detail by Sharan *et al.* [94]. Transformation of the cells with the target fragment is done by electroporation [94]. Briefly: the cells are washed several times with distilled water and concentrated. The purified DNA fragment is added to these electro-competent cells and a capacitance charged to about 2 500 Volts is discharged across the sample. The electrical field and current destabilize the membrane and allow DNA to enter the cell. Our team had prior experience with this technique and all the necessary equipment was ready and available. Our electroporator is a micro pulser from BioRad, the centrifuge is a Bioblock scientific 1-15K from Sigma.

4.2.2 Selection cassette

Even though λ_{red} increases the frequency of recombination, we still need to select the clones that have incorporated the foreign DNA. The frequency of positive clones is on the order of one in a million. Since we do not want to leave an antibiotic marker on the chromosome, all genomic engineering will proceed through a two-step process (Figure 4.3).

In the first step, we introduce a "selection cassette", from now on referred to as "cassette", at the target locus. Successful recombination is selected through and antibiotic marker carried on the cassette. The second recombination will replace the cassette with the DNA fragment that contains either the new gene, a mutated version of an already existing gene, or simply the

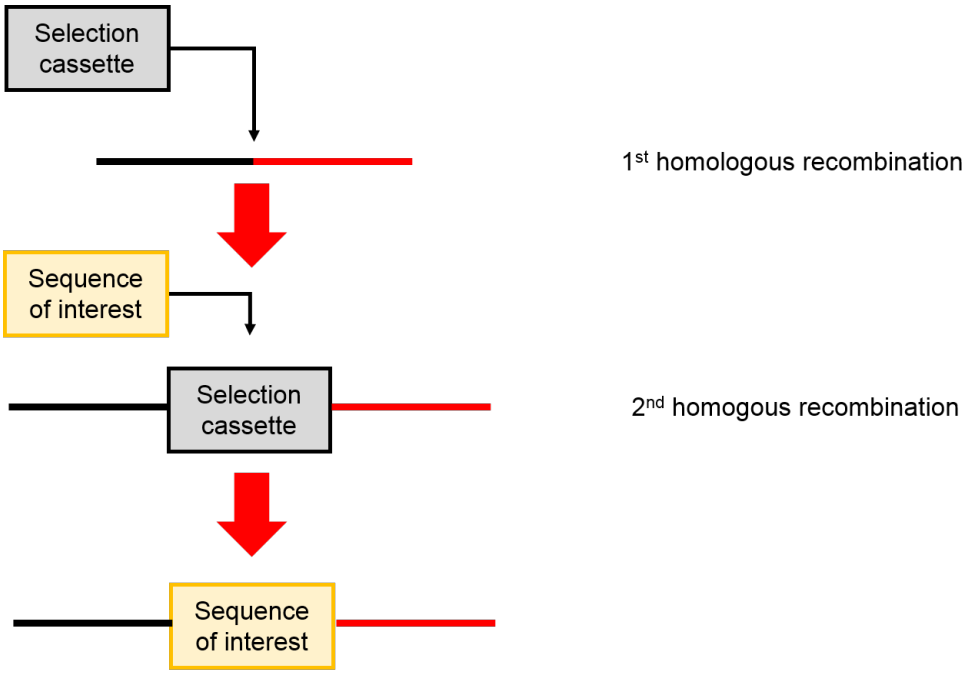

Figure 4.3: Two-step genomic engineering. *In order to change the sequence of the E. coli chromosome, we first recombine a "cassette" into the target locus, selecting an antibiotic resistance coded by the cassette. In a second recombination step, we replace the cassette with the sequence of interest. We can select this event because the cassette also contains a conditional toxin. In fact, in the second recombination we select the absence of the toxin, which indirectly proves the presence of the target sequence.*

homologous regions flanking the selection cassette, which will led to a deletion. We will also rely on the selection cassette for isolating the successful clones. The trick is that the cassette not only contain a resistance gene, but also a toxin controlled by an inducible promoter (Figure 4.4). By plating the cells on a Petri dish containing the inducer, only cells that have successfully removed the cassette in the second recombination step will survive and form colonies. In practice, some colonies that still have the cassette will also grow. The reason for these false positives is generally a mutation in the toxin. Assessment of the antibiotic resistance and sequencing of several positive clones is necessary to distinguish between false and true positives.

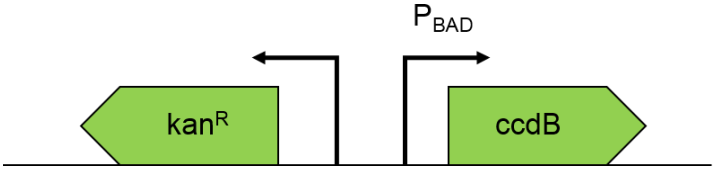

Figure 4.4: Selection cassette. *Our cassette carries two genes with their respective promoters. The kan^R codes for the resistance to kanamycin and is constitutively transcribed. The second gene is ccdB, a toxin that inhibits DNA gyrase. This gene is transcribed from the p_{BAD} promoter, which is only activated in the presence of arabinose in the growth medium.*

The selection cassette is shown in Figure 4.4. The positive selection uses a kanamycin resistance gene, kan^R, controlled by a constitutive promoter, while the negative selection uses the *ccdB* gene [95], controlled by an arabinose-inducible promoter, p_{BAD} [96]. When cells are grown

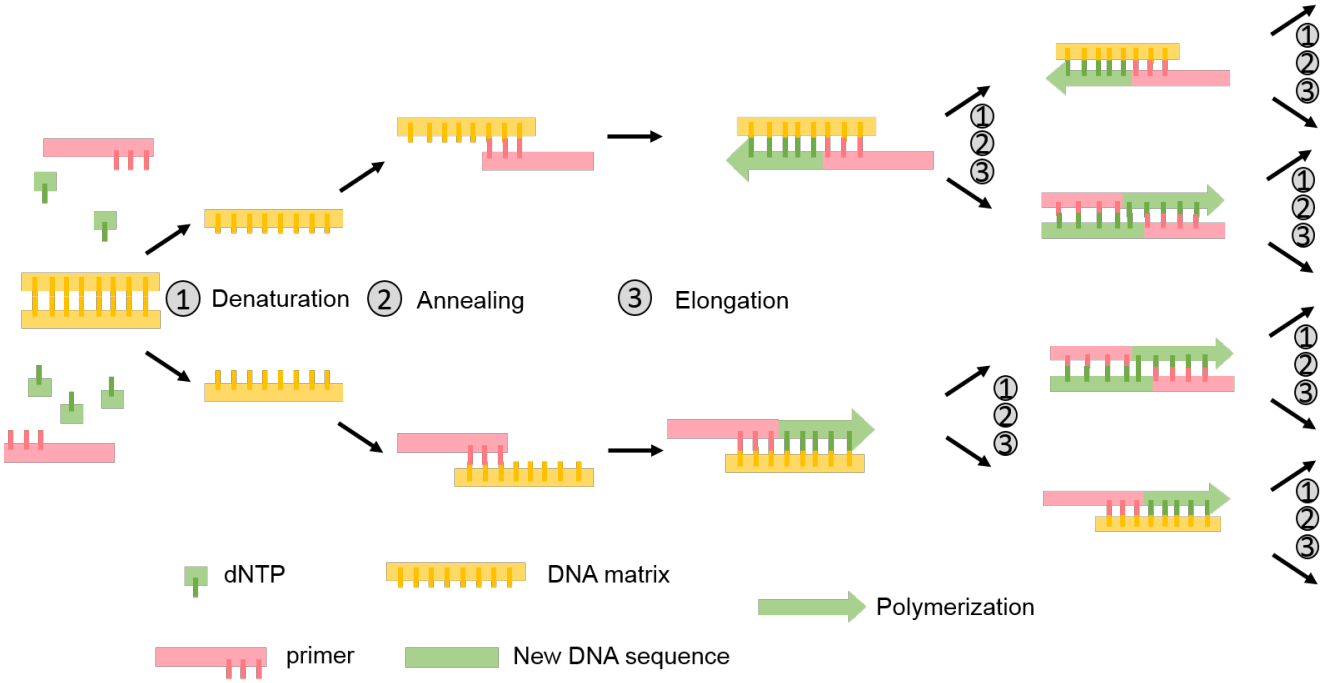

Figure 4.5: Polymerase Chain Reaction. *The reaction proceeds through thermal cycles leading to the denaturation of the matrix, annealing of the primers, and elongation of the primers by the thermostable DNA polymerase. In the example shown, the primers have an overhang: the 5'-part of their sequence is not complementary to the matrix. After amplification, the sequence of the 5'-overhang is incorporated into the DNA fragment.*

in the presence of 1% arabinose, the CcdB toxin is produced, leading to cell death.

4.3 PCR amplification

The selection cassette, as well as plasmids or other linear DNA fragments, need to be synthesized according to our needs. In order to do so, we used the polymerase chain reaction (PCR).

4.3.1 Principle

The PCR reaction replicates a matrix starting from primers that anneal to the end of the fragment. Since at each step, the number of double-stranded fragments doubles, we obtain an exponential amplification of the initial sequence. We typically perform thirty PCR cycles, leading to an amplification of up to $2^{30} \simeq 10^9$-fold. This process is summarized in Figure 4.5.

To add sequences to the extremities of the fragment, we can use primers with a 5' overhang of some bases. This means that the 5'-sequence of the primer does not correspond to the template sequence. However, after the first round of amplification, the extra bases have been incorporated into the fragment, which will serve as a template for subsequent amplification cycles (see Figure 4.5). All of the primers used in this study, as well as the fragments created and their function are presented in Table 5.2.

We used a Biometra T3000 Thermocycler for the PCR reactions. Reagents were purchased from Millipore when using the KOD xtreme hot start Taq polymerase and from Biogen when

using the Phusion Taq polymerase. We followed the protocol provided with each kit and used the program listed in Table 4.2 to perform the amplification.

step	Temperature $°C$	Time sec	Description
1	94 (98)	120	Initial denaturation of double strand DNA
2	98	10	Denaturation of double strand DNA
3	50-70	60 (30)	Annealing
4	68 (72)	30-180	Elongation
Repeat steps 2 - 4 30 times			
5	68 (72)	300	Final elongation

Table 4.2: PCR program . *The parameters used for Kod Xtreme hot start Taq polymerase values are presented in the table. The corresponding values for the Phusion Taq polymerase are in parentheses. Template DNA is amplified selectively using a primer pair. In the first step, the template is denatured. The following three steps - denaturation, primer annealing, and elongation - are repeated 30 times. A final elongation step ensures complete polymerization of the fragment. The annealing temperatures are chosen in accordance with the primer characteristics and the elongation time depends on the length of the target fragment.*

4.3.2 Methods

Overlapping PCR

PCR can also be used to assemble different DNA fragments. Combining different DNA fragments in a PCR reaction will generate the concatenation of the fragments, provided that successive fragments are homologous for about 20 bp at their ends. One fragment can thus serve as a primer for the next fragment. In order to amplify the joint fragment, the primer pair has to correspond to the extremities of the assembled fragment. This is called Overlap Extension - PCR.

Gibson assembly

An alternative to Overlap-Extension-PCR is the so called Gibson assembly. This method is also based on the partial overlap of the ends of the fragments that are to be assembled. Annealing of these ends is favored by including a 5'-3' exonuclease in the reaction. The free 3'-ends serve as primers for a Taq polymerase which will fill the gap created by the exonuclease. A DNA-ligase connects the fragments covalently. All three enzymes are thermostable. The reaction does not involve any temperature cycling, but needs to be carried out at a temperature corresponding to the melting temperature of the overhangs, i.e., about 50°C. The principle is illustrated in Figure 4.6.

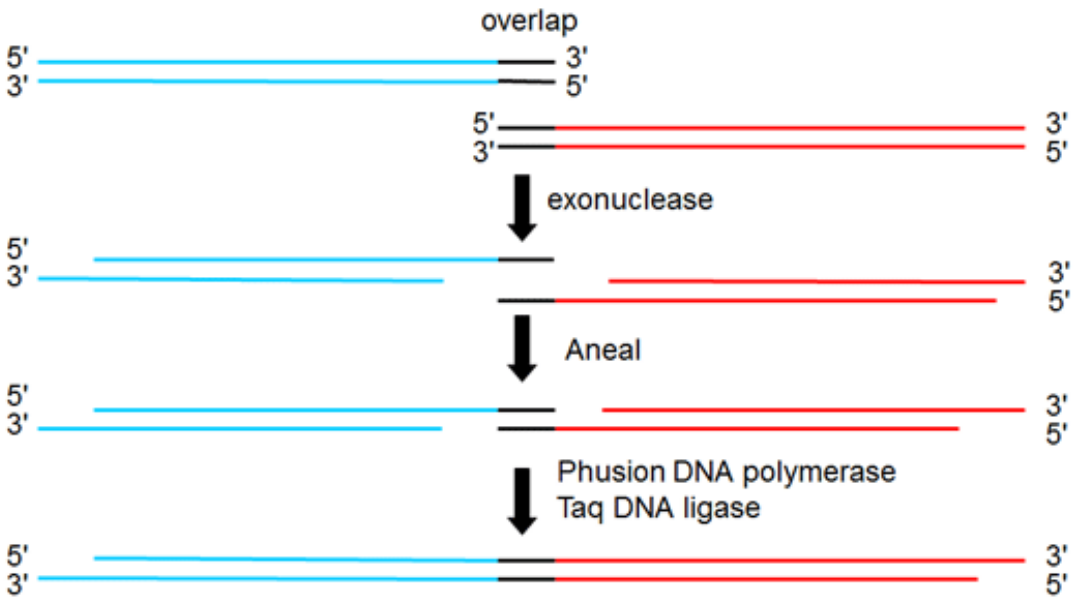

Figure 4.6: Gibson assembly. *Two double strand DNA fragments are joined via base pairing. An exonuclease rends the edges of the fragments into single strand DNA which can be annealed. Then a DNA polymerase synthesizes the missing gaps on the newly formed double strand which are connected covalently via a ligase. If the joined DNA fragment is circular then the exonuclease can no longer attach to it.*

Gibson assembly is generally used to assemble plasmids. However the protocol can be adapted for the construction of linear fragments. In order to compensate for the degradation of the ends of the assembled fragment by the exonuclease, we have to include a primer pair corresponding to the extremities of the assembled fragment. These primers will be extended by the Taq polymerase and thus "repair" the ends. We used the Gibson assembly kit from New England Biolabs as well as their protocol with the addition of the primer pair at a concentration of 10 μM each.

4.4 Verification of successful genetic engineering

Gel Electrophoresis

In order to analyze the PCR reactions, we use gel electrophoresis. We use 1% agarose gels with Tris-Acetate-EDTA (TAE) solution, purchased from Biosolve, as the running buffer. Agarose was purchased from Euromedex.

The sample is loaded into the wells after mixing at a 5:2 ratio with gel loading buffer: 30% glycerol (from Sigma), 0.25% of xylene cyanol FF (Sigma) and 0.25% of bromophenol blue (Sigma) in distilled water. We also added 0.001% of GRGgreen dye to the gel loading buffer. The fluorescent GRGreen dye interacts with DNA, rendering it visible when illuminated with blue light from a DarkReader Blue light Transilluminator.

All gels contain at least one lane of a molecular weight marker: generally the 1 kb Smart ladder (CSL-MDNA-1kb). This ladder contains DNA fragments of known sizes and concentrations. This allows us to estimate the size and concentration of the DNA fragment in the samples. A photo of one of our gels is presented in Figure 4.7.

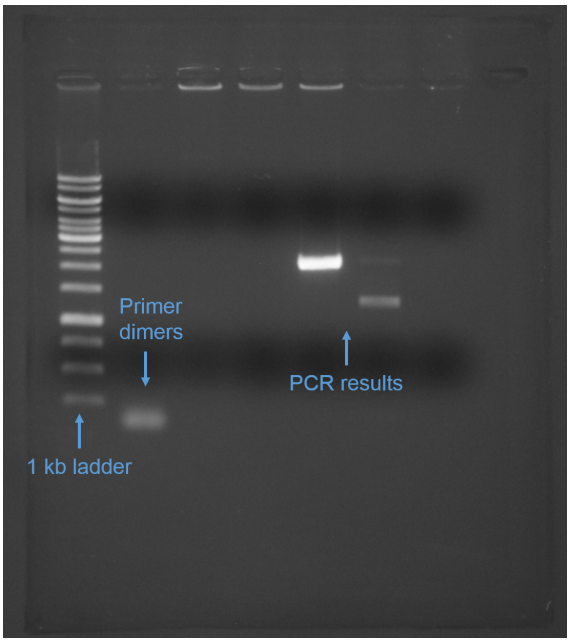

Figure 4.7: Example of PCR analyzed by gel electrophoresis. *The gel contains TAE + 1% agarose and the image was taken using a Blue light Transilluminator. The first well contains a 1 kb ladder. The second well is used as a control and contains the PCR mix with both primers but no matrix DNA, which often leads to primer dimerization. The fifth well shows the results of the PCR with an appropriate matrix (one bright band), while the sixth well shows the results with an inappropriate matrix (two bands).*

In case we obtain more than one DNA fragment in our PCR reaction, we excise the appropriate band from the gel. All PCR reactions, extracted from the gel or not, are purified using the Nucleospin gel and PCR Clean-up from Macherey-Nagel. The purification removes all the enzymes, the buffer component, and dNTPs from the solution.

Sequencing

After cloning of a PCR product into the chromosome, we verify proper integration by amplifying the region using PCR. This analysis provides information about the correct size of the insert. However, all chromosomal (and plasmid) constructions were also verified by sequencing the PCR product of the modified chromosomal region. Sequencing was performed by Eurofins. One read generally provides 900 bases of exploitable sequence information. We took the mapped sequences and aligned them with the theoretical sequence we expected. Matching sequences guaranteed that our modifications had been successful and we could carry on with the experiments using our successfully constructed strains.

Part III

Results

Chapter 5

Strain construction

The first step of this project consists in constructing a strain of *E. coli* in which the expression of *dnaA* is under external control. Even though we dispose of all the molecular genetics techniques necessary for this construction, this part of the project was certainly the most time-consuming, lasting over a year. In addition to the intrinsic difficulty of constructing a strain that leaves the only copy of an essential gene under external control, several pitfalls and complications (such as viral infections, etc.) have slowed the construction. I will not describe the difficulties, but focus on the strains we managed to construct.

As explained in the Introduction (section 2.3), we will use p_{uhpT} as our inducible, activatable promoter for expressing *dnaA*. The *uhpT* gene is removed; the inducer g6p will therefore remain in the growth medium and will not be consumed. Apart from these basic design decisions, an optimal choice of the starting strain is also very important. As mentioned above, the constructions are time consuming, making the design phase just as important as the actual construction.

Name	Relevant genotype
BW25113[97]	K-12, $rrnB_{T14}$, $\Delta lacZ_{WJ16}$, hsdR514, $\Delta araBAD_{AH33}$, $\Delta rhaBAD_{LD78}$
IJ39	BW25113, lacI::intS, lacI::galK
IJ40	IJ39, $spec^R$and two lac operators upstream of rpoBC*[19]
PP1	IJ39, rpoC-mcherry***
PP2	IJ40, rpoC-mcherry***
AC7	PP1, $\Delta dnaA_{4-909}$, uhpT::dnaA
AC8	PP2, $\Delta dnaA_{4-909}$, uhpT::dnaA
CL1	PP1, kan^R, P_{dnaA}::P_{uhpT}, $\Delta uhpT$****
CL2	PP2, kan^R, P_{dnaA}::P_{uhpT}, $\Delta uhpT$****

Table 5.1: *E. coli* strains used in this study. All strains are derived from *E. coli* K-12. * This construction is as shown in Figure 5.1. ** Bases 4 through 909 have been deleted as shown in Figure 5.2. *** Fusion protein between rpoC and mcherry. **** P_{uhpT} has been deleted from its original location, see Figure 5.3.

5.1 Starting strain, control of the concentration of RNA polymerase

We use derivatives of *E. coli* because of the ease of genetic manipulation and the ample previous experience of the laboratory with this organism. In particular, all of our strains are derived from BW25113, partly because Baba *et al.* [77] have constructed a library of all single-deletion of non-essential genes. Even though we do not explicitly use this strain collection in this work, the availability of the deletion strains will facilitate future, functional analyses. The characteristics of the strains we constructed are listed in Table 5.1. BW25113 strain does not metabolize arabinose, which allows us to use the selection cassette developed in the laboratory for scarless gene replacements using homologous recombination. The ideal growth conditions of *E. coli* are at 37°, but the bacterium can also grow at lower and higher temperatures, such as 30° or 42°. All these characteristics make it a good candidate for a biophysical study.

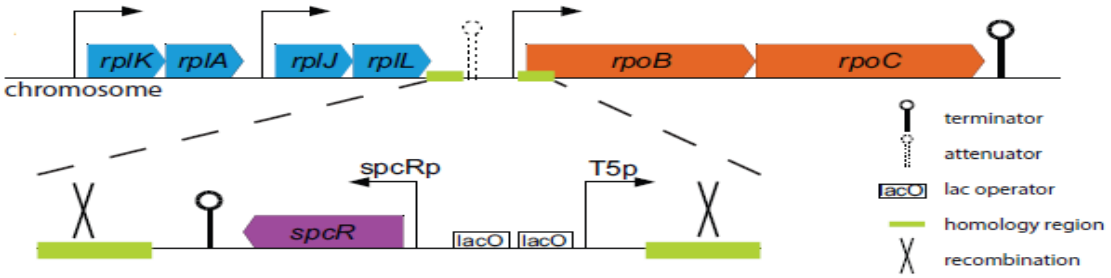

Figure 5.1: Controlling the expression of RNA polymerase. *The starting strain for our constructions contains a system that allows the conditional expression of the two large subunits of RNA polymerase [19]. The strains IJ39 and IJ40 also contain two additional copies of the lac repressor, LacI, on the chromosome (replacing the* galK *and* intS *genes). The transcription of the* rpoBC *operon is under the control of a strong promoter, repressed by LacI. The rate of production of RNA polymerase is therefore determined by the concentration of the inducer, IPTG, in the growth medium.*

We decided to base our experiments on strains constructed by Jérôme Izard [19].The IJ40 strain puts the expression of the two large subunits of RNA polymerase under the control of the inducer IPTG (Figure 5.1). Upon removal of IPTG, RNA polymerase is no longer produced and growth stops. This phenotype is very close to the phenotype we want to produce, but for very different mechanistic reasons: stopping transcription versus stopping DNA replication. In order to avoid "escapers", essentially mutations in lac repressor that circumvent the imposed control, two additional copies of the lac repressor gene were put onto the chromosome. The strain, otherwise wild type, containing these two copies of lac repressor is called IJ39. The strain containing the additional control of *rpoBC* expression is called IJ40. These strains, were further modified in order to quantify the number of RNA polymerases in the cell. A fluorescent protein, mCherry, [13], was fused to *rpoC*. These translational fusions resulted in the strains PP1 and PP2 [19]. At full induction, these strains produce more RNA polymerase than the WT stain. Apart from this difference, the physiology of PP1 is indistinguishable from the WT. PP2 also possesses WT characteristics when IPTG is present in the growth medium, but completely stops production of RNA polymerase, and therefore growth, in the absence of IPTG.

By controlling the concentration of RNA polymerase we modify gene expression. By controlling dnaA, we modify the replication of the chromosome. Combining these two controls will give us two independent ways of modulating two of the most fundamental functions of the cell: transcription and replication.

5.2 Design of stains controlling the expression of *dnaA*

We can control the transcription of *dnaA* in two different ways. Either we move the *dnaA* gene to the *uhpT* locus, replacing *uhpT* by *dnaA* (these constructions will be called "AC"), or we replace the promoter of the *dnaA* operon by p_{uhpT} (these constructions will be called "CL"). Both strategies require the deletion of the native *uhpT* gene to avoid the uptake of g6p. Since we replace the entire promoter region, including the start codon of *dnaA,* by the corresponding region of *uhpT*, we also change the RBS (ribosome binding site) of *dnaA*. The rate of protein expression is affected, among other factors, by the sequence of the RBS and of the start codon of the gene. As a result, the translation efficiency of DnaA in our construction will not be identical to the ones of wild type *dnaA* or *uhpT*. The natural start codon of *uhpT* is AUG, while the start codon of *dnaA* is GTG. In all our construction, we retained the start codon of *dnaA*.

Both strain designs have advantages and disadvantages and we decided to construct them all. Due to time constraints, most of the work was done using the construction that movs *dnaA* to the *uhpT* locus, i.e., the AC-strains.

5.2.1 Design of the AC-strains

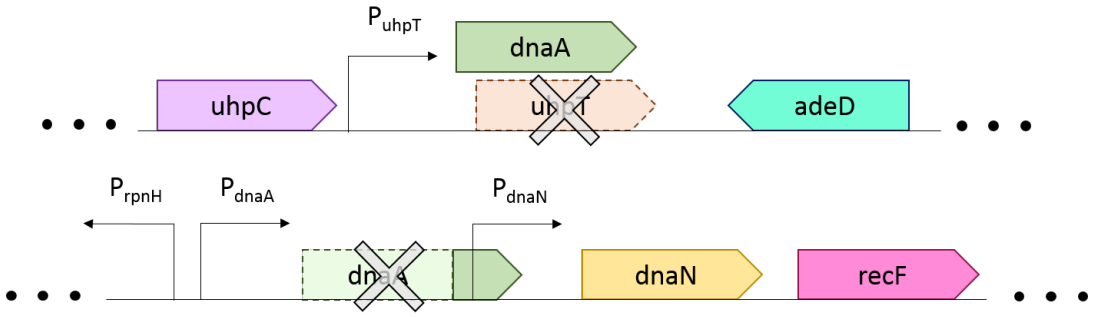

Figure 5.2: Design of the AC-strains. *In a first step, the uhpT gene was removed from the chromosome and replaced by a copy of dnaA. In a second step, the original dnaA gene was partially deleted. We kept the C-terminal region of DnaA in order to retain the four dnaN promoters (symbolized here by a single broken arrow). The partial sequence of dnaA does not code for a functional* protein..

The AC-strains have the advantage that the genome of the cell is minimally perturbed. The remaining genes of the *dnaA* operon, *dnaN* and *recF*, keep all of their native regulatory signals. The region upstream of *uhpT* remains equally unchanged, including the RBS. The *adeD* gene is located downstream of the new location of *dnaA*, but transcribed in the opposite direction. No perturbation of the expression of *adeD* is to be expected.

Additional considerations need to be taken into account for the way to remove the native *dnaA* gene. *dnaA* is the first gene of the operon. In order not to perturb the expression of

the other genes, we need to keep all of the promoters of the operon. Unfortunately, four of the promoters transcribing *dnaN* are located within the coding region of *dnaA*. These promoters are located in the fourth region of DnaA, responsible for binding to DNA (Figure 2.2). We therefore decided to leave this region intact by only removing the N-terminal parts of DnaA in such a way as to maintain the reading frame. Not expressing the protein fragment would probably lead to premature termination (by Rho) of transcription from the upstream *dnaA* promoter, compromising our attempt to not interfere with the native regulation of *dnaN* and *recF*. The AC-constructions will therefore express a truncated version of DnaA from the native promoter and the full-length, functional DnaA protein from the g6p-regulated promoter. The truncated protein, comprising the DNA-binding domain of DnaA, is not functional for initiating DNA replication [15].

5.2.2 Design of the CL-strains

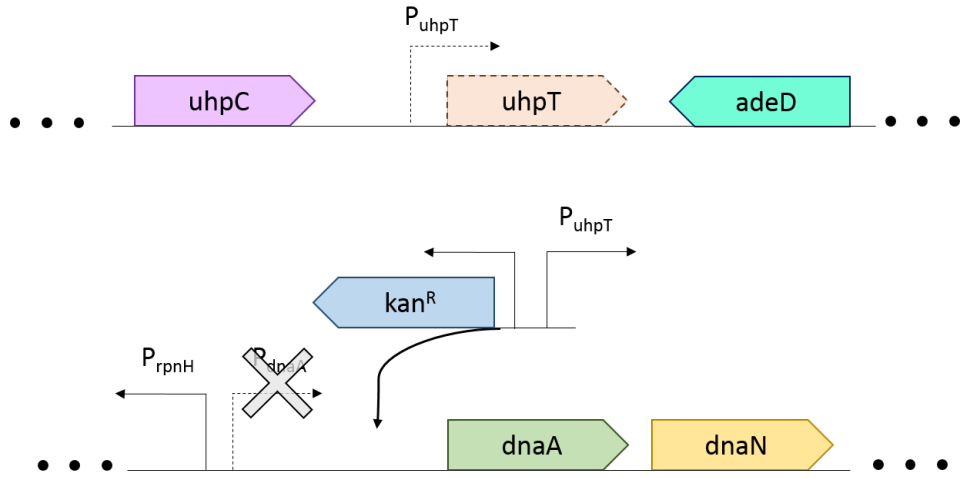

Figure 5.3: Design of the CL-strains. *The uhpT gene, along with its promoter, p_{uhpT}, were deleted from the chromosome. A copy of this promoter region was placed upstream of the dnaA operon, including the last 59 bp of the uhpC gene, thereby retaining a CRP binding site that regulates the* uhpT *promoter. The construction replaces the entire region upstream of the start codon of dnaA by the corresponding sequences of p_{uhpT}.*

The second type of design, producing the CL-strains, maintains the integrity of the *dnaA* operon, but puts the entire operon under the control of the inducible *uhpT* promoter. In order for g6p not to be metabolized, we have to delete *uhpT* (Figure 5.3). The motivation for this design was to remove any auto-regulatory controls from the expression of *dnaN* and *recF*. In the AC-constructions, since DnaA expression is controlled independently, we create an interference with the transcription of *dnaN* and *recF* because the *dnaA* promoter (also transcribing *dnaN* and *recF*) is negatively auto-regulated by DnaA. Thus, when we decrease the DnaA concentration in the AC-strains, we automatically increase the concentrations of DnaN and RecF. The simultaneous over-expression of DnaN and RecF is lethal [30], although we do not know if the *dnaA* promoters, even fully induced, produce sufficient amounts of DnaN and RecF to cause lethality.

The CL-strains avoid this potential problem, but create another difficulty: when modulating the expression of DnaA, we modulate the expression of the entire operon. In other words, we

still do not completely uncouple the expression of DnaA from the expression of the rest of the genome. In theory, physiological of growth effects of different DnaA concentrations could, in part, be mediated by DnaN or RecF.

5.2.3 Strain AC8 is used for most experiments

A big difference between the two constructions is that AC keeps P_{uhpT} in its native location while the CL-strain does not. Even though the chromosomal locations are close, the chromosomal context affects the overall promoter strength [98, 99]. Thus, we can be certain that the promoter strength of the AC construction is identical to the promoter strength of the native *uhpT* promoter. This will allow us to make direct comparisons with a previous characterization of the promoter [1].

In light of this, we decided to focus our experiments on the AC construction, while developing the CL strains in parallel. In other words, most of our experiments used the AC-strain, and in particular AC8, where the expression of *dnaA* is under the control of P_{uhpT} at the *uhpT* locus.

5.3 Construction of the AC- and CL-strains

Both constructions are rather difficult because dnaA is an essential gene. Both constructions were performed by homologous recombination, utilizing the selection-counterselection cassette described in subsection 1.2.

Both constructions begin by transforming the plasmid carrying the λ_{red} recombination functions, pSIM5, into PP1 and PP2. A first recombination reaction removes *uhpT*. The cassette was PCR amplified with primers p1f and p1r that add the homologous sequences flanking *uhpT* to the cassette (see Table 5.2). Integration of the cassette therefore removes *uhpT*. This constitutes the starting strain for the following manipulations.

5.3.1 AC construction

The second step in the construction of the AC-series consists in replacing the cassette with a second copy of *dnaA* (Figure 5.2). The dnaA gene was PCR amplified using primers pac1f and pac1r (Table 5.3). The recombinants were selected on LB-arabinose plates (see subsection 3.1.4) and selected clones were verified by sequencing using primers ps1f and ps1r (Table 5.3).

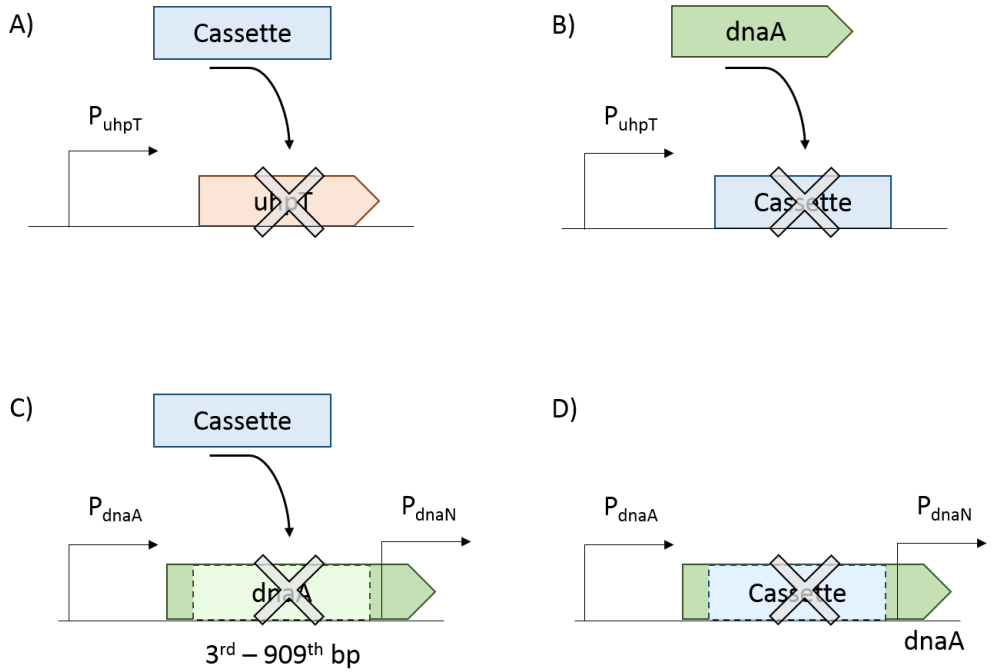

Figure 5.4: Construction of the AC-strains. ***A)*** *The first step of the construction consists in inserting the cassette into the uhpT locus, thereby removing uhpT.* ***B)*** *In a second step, the cassette is replaced by a copy of dnaA.* ***C)*** *The cassette is inserted into the native copy of dnaA, removing codons 2 to 303. The promoters of dnaN remain untouched.* ***D)*** *The cassette is removed by a PCR assembled sequence homologous to the regions flanking the cassette. This leads to a "clean" deletion of the cassette.*

We then removed the native copy of *dnaA* by inserting the cassette between nucleotides 3 and 909 of the *dnaA* coding sequence. The cassette had been amplified with primers pac2f and pac2r (Table 5.3). All media were, of course, supplemented with g6p in order to ensure the transcription of *dnaA* by p_{uhpT}. Finally, we removed the cassette using recombination with a PCR synthesized homologous to the sequences flanking the cassette. This DNA fragment was amplified using primers pac3f and psac3r (Table 5.3).

The resulting construct deletes nucleotides 3 to 909 of *dnaA* ($\Delta dnaA_{3-909}$) and puts the only remaining copy of dnaA under the control of p_{uhpT}. For all recombination reactions, the bacteria were grown at 30°C in order to maintain the recombination plasmid, pSIM5, which contains a temperature-sensitive origin of replication. We grew the final construct at 37°C for at least twenty generations in order to eliminate the plasmid. Chloramphenicol sensitivity of the final clone ensures the loss of pSIM5. Sequencing of the clones, using the primers ps2f and psac3r (Tables 5.2 and 5.3), confirmed the success of the constructions.

We thus constructed the strains AC7 and AC8, respectively derived from the parent strains PP1 and PP2. As expected, AC7 grows only in the presence of g6p, whereas AC8 requires both g6p and IPTG for growth.

5.3.2 Construction of the CL-strains

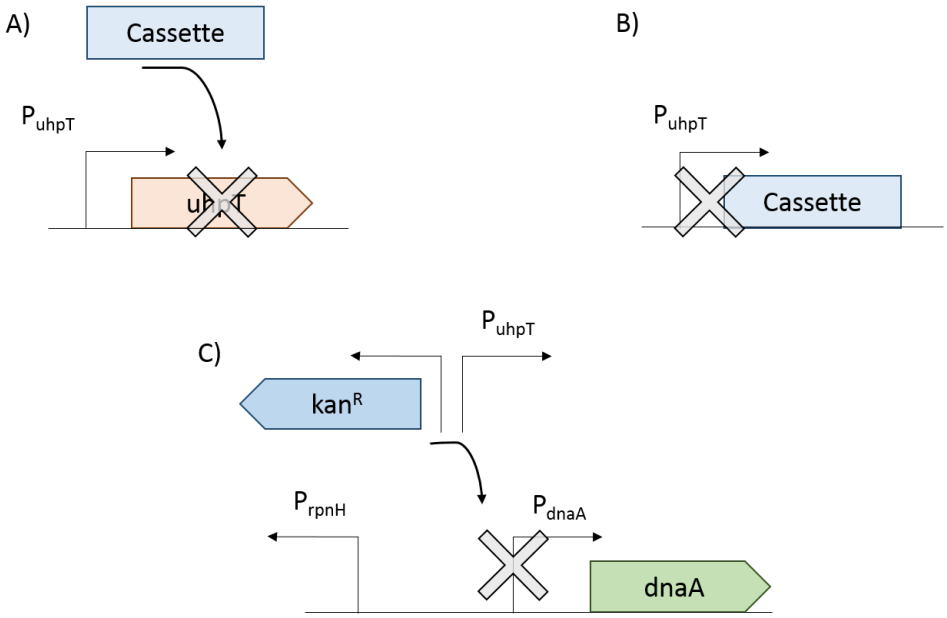

Figure 5.5: Construction of the CL strains. A) *The uhpT region, including the promoter and the coding sequence, was removed by inserting the cassette.* **B)** *The cassette was removed, leaving the strain ΔuhpT.* **C)** *We assembled a fragment containing four function elements: (i) 400 bp of homology with the region upstream of p_{dnaA}, (ii) the gene coding for the resistance to kanamycin, (iii) the* uhpT *promoter region up to and not including the start codon of* uhpT*, and (iv) the first 400 bp of the* dnaA *coding region including the start codon. This fragment was recombined into the ΔuhpT strain in order to obtain a construct where the entire dnaA operon is under the control of p_{uhpT}.*

The single-stranded DNA-fragment used for removing the cassette inserted into uhpT (pcl1 Table 5.4) is homologous to 50 bp upstream of P_{uhpT} and 50 bp downstream of the *uhpT* gene. Removal of the cassette therefore also removes the promoter of uhpT, leaving us with a Δ*uhpT* derivative of PP1 or PP2. The next step of the construction consists in replacing the promoter upstream of *dnaA* with the promoter of *uhpT*. For lack of time, we decided to use a single recombination reaction for this replacement, with the drawback that the strain will carry an antibiotic resistance gene. The construction is schematically summarized in Figure 5.3.

We therefore assembled a recombination fragment comprising the following functional elements which were synthesized by PCR: region upstream of P_{dnaA} - antibiotic resistance gene - p_{uhpT} and first part of coding region of *dnaA*. The first and third part of the fragment are homologous to the *dnaA* region of the chromosome, the second part provides a selectable marker and the third part places the *uhpT* promoter upstream of the *dnaA* operon. The recombination fragment was constructed by Gibson assembly as described in subsection 4.3.2. The assembly of the four functional elements required the concatenation of three DNA fragments: the region upstream of p_{dnaA} assembled using primers pcl2af and pcl2ar (Table 5.4), the gene coding for kanamycin resistance including its promoter amplified using primers pcl2bf and pcl2br (Table 5.4), and the *uhpT* promoter upstream of *dnaA*, amplified using primers pcl2cf and pcl2cr (Table 5.4) and the AC8 strain as the template. All three fragments had an overhang that allowed them to be

concatenated by PCR. During the Gibson assembly, we used two more primers, pcl3f and pcl3r, in order to amplify the full fragment. The finished DNA construct was extracted from an agarose gel after verifying its size.We then proceeded to insert this fragment into the strain lacking *uhpT*.

The successful recombination was selected by kanamycine resistance. pSIM5 was removed by growth at 37°C and the construction was verified by sequencing, using the primers pcl2af and psac2r (Tables 5.3 and 5.4). As expected, the strains need g6p for growth. The strains derived from PP1 and PP2 were called, respectively, CL1 and CL2.

Name	Type	Description	Sequence
p1f	Forward primer	Amplification of cassette for *uhpT* deletion. WT template	TGATTTTTA CAATGCATG CCTCACGCA GGTATTCAT TTCAGGAGT AACCCTTAT ATTCCCCAG AACATCAGG TTAATGGCG
p1r	Reverse primer		GAATAATAA AAAAAGCCC GGCGTCATG CCGGGCAAA AGTCACCAG TTACGTTTA GAAGAACTC GTCAAGAAG GCGATAGA
ps1f	Forward primer	Sequencing upstream of *uhpT*	ACAATGCAT GCCTCACGC AG
ps1r	Reverse primer	Sequencing downstream, starting in *ade*	CACTACGCT GGAAGTCAC GG
ps2f	Forward primer	Sequencing upstream of *dnaA*	CGTCACCCT CAAGCAGGG TC

Table 5.2: General Oligonucleotides used for this study

Name	Type	Description	Sequence
pac1f	Forward primer	Amplification of *dnaA* gene for insertion in *uhp* region. Template: wt cells.	TGATTTTTA CAATGCATG CCTCACGCA GGTATTCAT TTCAGGAGT AACCCGTGT CACTTTCGC TTTGGCAG
pac1r	Reverse primer		GAATAATAA AAAAAGCCC GGCGTCATG CCGGGCAAA AGTCACCAG TTACGTTTA CGATGACAA TGTTCTGAT TAAATTTGA AAAATCTTC
pac2f	Forward primer	Amplification of cassette for deletion of *dnaA* gene . Cassette DNA fragment as template.	GCCTTAGTC ATTATCGAC TTTTGTTCG AGTGGAGTC CGCCGTGTT AGAAGAACT CGTCAAGAA GGCGATAGA AG
pac2r	Reverse primer		CGAAGCGGG ATTTCAAAC GATCCTCAA CGCCGTTGA TCTCTTATA TTCCCCAGA ACATCAGGT TAATGGCG
pac3f	Forward primer	Amplification of sequence between *dnaN* and P_{dnaN} in order to remove the cassette. template: WT.	TGTTTCAGC CTTAGTCAT TATCGACTT TTGTTCGAG TGGAGTCCG CCGTGATCC TGATGAAAA AGGCCGACG AAAACG
psac3r	Reverse primer	Amplification of sequence between *dnaN* and P_{dnaN} in order to remove the cassette. Used for sequencing downstream of *dnaA*.	CTGGCTCGT GTGGCTGAA CC

Table 5.3: Oligonucleotides specific to AC construction

Name	Type	Description	Sequence
pcl1	primer	Removal of cassette in *uhp* region for CL construction	TAAAAAAAG CCCGGCGTC ATGCCGGGC AAAAGTCAC CAGTTACGT TCACGCTTC GCGCGGTGT CTGGGCGTT CAAAAAGGG CAGTAACAG
pc12af	Forward primer	Amplification of fragment downstream of P_{dnaA} for insertion of $kan^R - P_{uhpT}$ fragment. Template: WT strain.	GTTGCGCTT CAGTACAGA CG
pcl2ar	Reverse primer		TGACGAGTT CTTCTAAAA TTGGCTTAA GGCGT
pcl2bf	Forward primer	Amplification of kan^R gene for insertion of $kan^R - P_{uhpT}$ fragment. Cassette used as template.	ATCGATTAA GCCAATTTT AGAAGAACT CGTCAAGAA GGC
pcl2br	Reverse primer		AGCAGTGCG GAAATCCGG CGCCCTCTG GTAA
pcl2cf	Forward primer	Amplification of P_{uhpT} sequence and a portion of *dnaA* for insertion of $kan^R - P_{uhpT}$ fragment. Template: AC8-strain.	CTTACCAGA GGGCGCCGG ATTTCCGCA CTGCT
pcl2cr	Reverse primer		GACGTTATC CCAACCTGA GC
pcl3f	Forward primer	Amplification of $kan^R - P_{uhpT}$ assembled fragment during Gibson assembly	GTTGCGCTTCAGTACAGACG
pcl3r	Reverse primer		GACGTTATCCCAACCTGAGC

Table 5.4: Oligonucleotides used for the construction of the CL strains

Chapter 6

Population Analysis

With our engineered strains, we can now observe how different concentrations of DnaA affect growth. Because of the small numbers of molecules involved, single cell analysis is necessary to fully understand our system. However, to properly analyze single cell data, we need to first acquire a basic understanding of the average behavior of our system. In other words, we will start by characterizing the system at the population level.

Population experiments are simpler to analyze and will allow us to study the average behavior of our strains. In particular, we want to investigate how the cells behave at low concentrations of DnaA. In order to correctly estimate the quantity of DnaA proteins in the cell, we have to quantitatively analyze the activity of the promoter of *uhpT*, P_{uhpT}, and compare our results to previously published reports [1]. We provide quantitative interpretations of the observed behaviors and develop a mathematical model describing the activity of the *uhpT* promoter, and therefore the quantity of DnaA in the cells. We start by assessing the concentration of DnaA during exponential growth, i.e., in steady state conditions, which greatly simplifies the mathematical analysis.

6.1 All or none response of growth rate to varying concentrations of DnaA

We are most interested in the behavior of the engineered strains in conditions where they can grow without any nutrient limitations. In other words, we focus here on exponential growth, where the only limitations are due to the variation of the concentration of DnaA, which we control by varying the activity of P_{uhpT}. Under these conditions, we can formulate the following equation describing the the total volume, V, of cells in the culture as a function of time:

$$\frac{dV}{dt} = \gamma(c)\, V \tag{6.1}$$

V is the total biomass in the culture (proportional to the volume of the bacteria), c is the intracellular concentration of DnaA, and γ is the growth rate, which depends only on the concentration of DnaA.

We use the following assumptions for our model: at a given concentration of DnaA (c), the growth rate is constant as long as the carbon source has not been depleted. If the concentration

of DnaA falls below a certain threshold, growth stops: $\gamma(c) = 0$. If the DnaA concentration is not much greater than the nominal value, $\gamma(c) = \gamma_{max}$. Only if the concentration of DnaA increases much further, $\gamma(c)$ will decrease. This behavior is schematically shown in Figure 6.1.

These assumptions agree with the reported response of *E. coli* to variations in the concentration of DnaA. Growth stops if the quantity of DnaA proteins goes below a critical value [14] and a strong over-expression of DnaA leads to double stranded breaks, which reduce cell viability [32]. However, at a moderately increased concentration of DnaA ($\lesssim$ 10 times the nominal concentration), the growth rate remains at its normal value [100]. The experimental challenge is to recreate this behavior with our strains using the external control over the intracellular concentration of DnaA. In particular, we want to study the transition from normal growth to growth arrest when the concentration of DnaA passes the lower threshold (Figure 6.1).

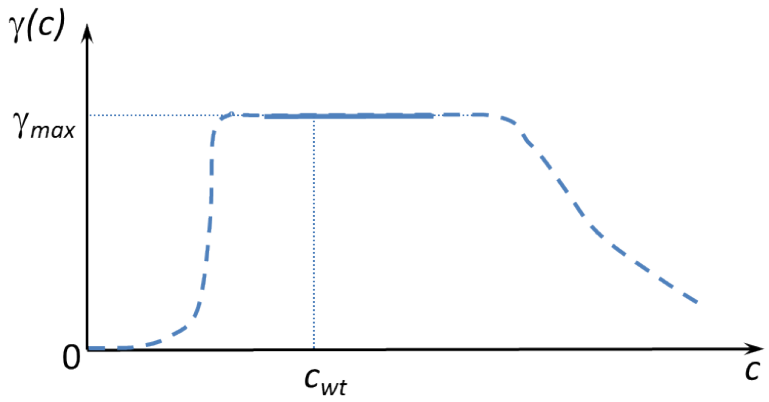

Figure 6.1: Diagram showing the growth rateγ as a function of the concentration of DnaA, c. *DNA replication, and therefore cell division, stops when the concentration of DnaA is below a critical threshold. For different reasons, a large excess of DnaA also leads to reduced growth rates or cell death. Over a roughly 10-fold range around the concentration of DnaA in wt cells, called "nominal" concentration (c_{wt}), the growth rate remains constant at the maximal value. The growth rate dependence is shown as a dashed line and the region of the constant, maximal growth rate is indicated by a solid line.*

These assumptions are true for a given set of growth conditions. If the type of growth medium is changed, it is to be expected that DnaA concentrations and growth rates will change. For this reason, we will not compare two different types of growth media and carry out all of our experiments in well-defined, minimal media.

6.1.1 Mathematical description of the concentration of DnaA

In order to experimentally obtain $\gamma(c)$, we need to measure, or infer, the growth rate and the intracellular concentration of DnaA. Growth rate is directly obtained by measuring the time-course of the absorbance or fluorescence of the culture [101]. To a first approximation, these spectroscopic properties are proportional to biomass. Since we can not measure the concentration of DnaA directly, we have to devise indirect methods for inferring this quantity.

In exponential growing cells, the concentration of DnaA changes mainly by three mechanisms. The first is the production of DnaA. In the AC strains, the synthesis of DnaA is dependent on

the activity of P_{uhpT}, and therefore externally controlled by the experimenter. Two mechanisms lead to the degradation of DnaA: proteolysis of DnaA and growth dilution. The proteolytic degradation (destruction of the protein by natural hydrolysis or active degradation by a protease) is slow. As most proteins in the cell, DnaA has a half-life on the order of 20 hrs. This rate is therefore negligible to "degradation" due to dilution of DnaA proteins during growth. At each doubling of the cell volume, the concentration of DnaA is divided by two. Thus, the equation describing the variation of the concentration of DnaA in the cell as a function of time is as follows:

$$\frac{dc}{dt} = \alpha \cdot p_{uhpT}\left(g6p\right) - \left[\gamma\left(c\right) + \gamma_{int}\right] c \tag{6.2}$$

The production term is proportional to a constant α and the activity of P_{uhpT}, which itself is a function of the concentration of the externally added inducer, $g6p$. The constant α encompasses all cellular processes necessary for the production of DnaA, such as the efficiency of ribosomes in translating proteins and the half-life of the mRNA. By definition, all these quantities are constant during steady-state growth. The degradation term is proportional to the concentration of DnaA and the sum of the rate constant for the intrinsic proteolysis of DnaA (γ_{int}) and for growth dilution ($\gamma\left(c\right)$). During growth, γ_{int} is negligible compared to $\gamma\left(c\right)$. Since the steady-state condition implies $\frac{dc}{dt} = 0$, we can express the steady state concentration of DnaA as:

$$c = \frac{\alpha \cdot p_{uhpT}\left(g6p\right)}{\gamma(c) + \gamma_{int}} \tag{6.3}$$

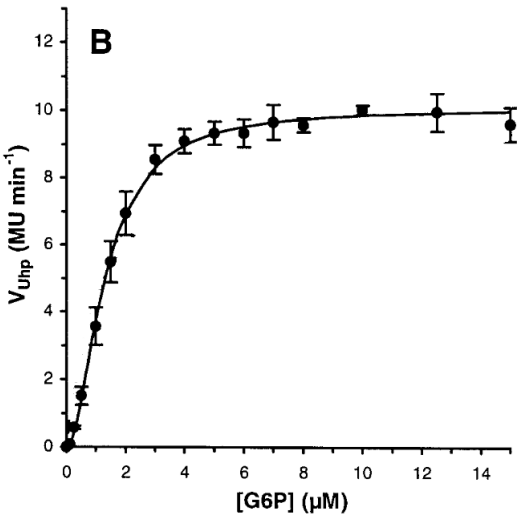

Figure 6.2: Activity of the uhpT promoter as a function of the concentration of g6p in *E. coli* RK5115. *Cells where grown in MOPS minimal medium supplemented with 10 mM glucose. The graph shows the promoter activity of P_{uhpT} (in Miller Units per minute) as a function of the concentration of g6p added to the growth medium [1]. A fit to the Hill function, $\frac{g6p^{n}H}{K_{0.5}^{g6p}+g6p^{n}H}$, yields: $K_{0.5}^{g6p} = 1.3\mu M$ and $n_H = 1.95 \pm 0.17$. g6p is the concentration of glucose-6-phosphate, $K_{0.5}^{g6p}$ is the Hill constant (related to an apparent dissociation constant of g6p for the activating complex of the promoter), and n is the Hill coefficient, measuring the cooperativity of the activation.*

Equation 6.3 relates the intracellular concentration of DnaA, c, to the activity of P_{uhpT}. Given the relationship between the concentration of $g6p$ in the medium and the activity of P_{uhpT}, we can estimate c (up to a multiplicative constant, α) from P_{uhpT}. The activity of P_{uhpT} is a function of the $g6p$ concentration and has been measured before (Figure 6.2).

We now possess all the ingredients for measuring $\gamma(c)$, the function that relates the concentration of DnaA in the cells to their growth rate. In Figure 6.2, we can see that in MOPS minimal medium, P_{uhpT} varies in the concentration range $0-10\mu M$ of g6p. We therefore decided to measure the growth rate of our cells in MOPS minimal media supplemented with g6p in this concentration range.

6.1.2 Population growth as a function of g6p

We grew AC8 cells, along with an equivalent strain containing the WT dnaA operon (IJ39), in the presence of $150uM$ IPTG. This concentration of IPTG produces roughly the same amount of RNA polymerase as in the WT strain. We used MOPS minimal medium containing 0.2% glucose as the sole carbon source, making these experiments comparable to published data measuring the activity of P_{uhpT}. The results are shown in Figure 6.3.

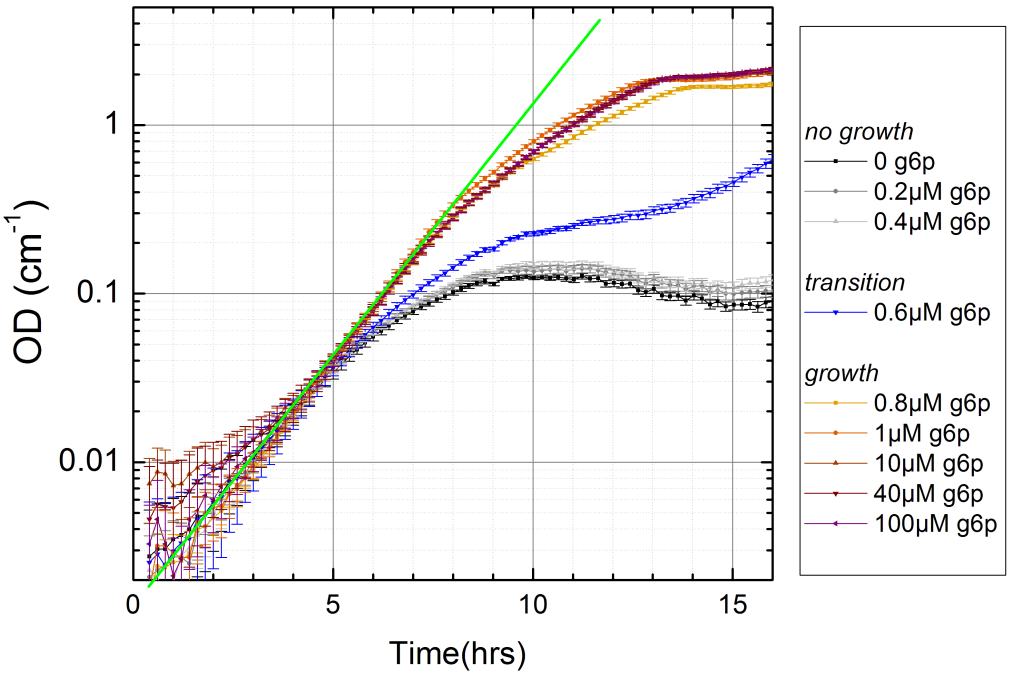

Figure 6.3: Growth of the AC8 strain in a 96 well plate with varying concentrations of g6p. *AC8 was grown in MOPS minimal medium supplemented with 0.2% glucose, 150 μM IPTG and different concentrations of g6p (see legend). The green continuous line depicts an exponential growth at γ_{max}(a doubling time of 60 min). Cells were grown as described in setups and cell growth (section 3.1.6).*

The results show that even when P_{uhpT} is fully induced, the growth rate remains at the

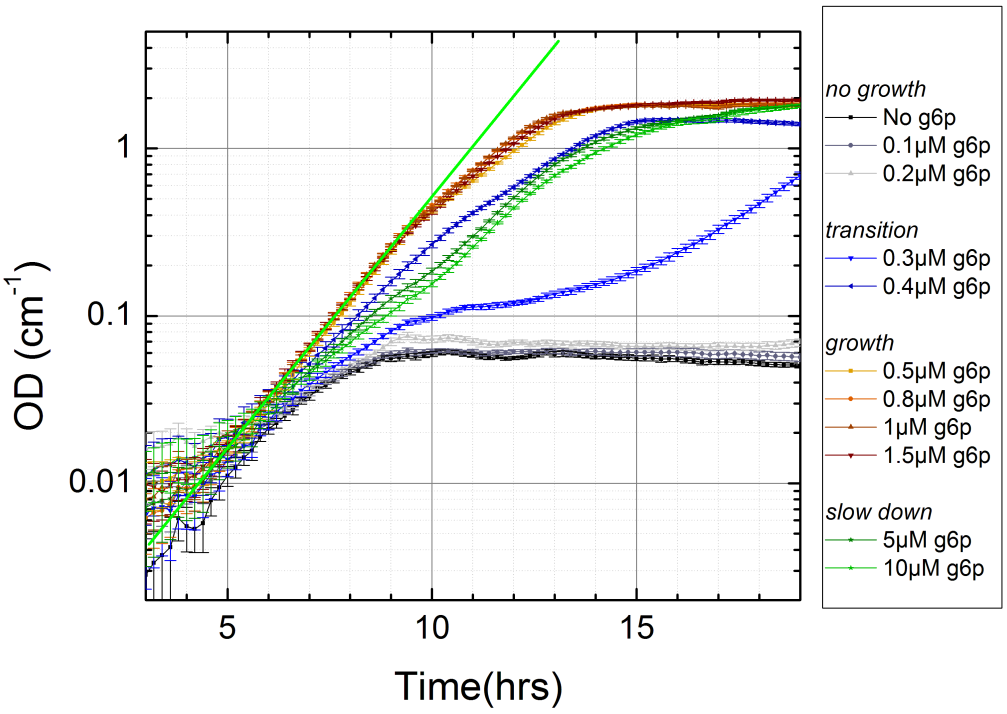

Figure 6.4: Growth of CL2 strain in a 96 well plate with varying concentrations of g6p. *CL2 was grown in MOPS minimal medium with 0.2% glucose , 150 μM IPTG and different g6p concentrations (see legend). The green continuous line depicts an exponential growth at γ_{max}(a 60 min doubling time).*

maximum value, which corresponds to a doubling time of ∼60 minutes, the same as in wild type bacteria. The fact that we can not decrease the growth rate by over-expressing DnaA simply means that the production of DnaA is not in the growth limiting range (see Figure 6.1). In other words, the strength of P_{uhpT} at maximal induction (∼10 MU) is still too weak to impede cell growth as has been observed for stronger over-expression of DnaA [30].

On the other hand, lowering the expression of DnaA produces an abrupt drop in growth rate at a concentration of g6p below $0.8\,\mu M$. At g6p concentration between 0.7-0.6 μM, cells behaved somewhat erratically. In certain samples, growth was arrested, while in others the bacteria showed an intermediate behavior between growth arrest and exponential growth as shown by the blue line in Figure 6.3. Intermediate growth rates could not be maintained throughout the experiment, meaning that the dependency of growth rate on the concentration of DnaA follows an all-or-none transition, a step-function. This shows that the dynamical system possesses two stable states, either maximal growth rate or no growth at all.

The CL strains grown in the same conditions also present an arrest of growth when the *g6p* concentration drops below a critical value, in this case $0.5\,\mu M$ (6.4). The minimal concentration at which the growth remain exponential is thus significantly lower than for the AC strains ($0.8\,\mu M$). At intermediate concentrations of g6p, between $0.5 - 1.5\,\mu M$, the CL-bacteria grow exponentially at the maximal growth rate (brown curves) and slow down at higher *g6p* concen-

trations, $5-10\,\mu M$ (green curves). A summary of the growth behavior of both type of cells (AC and CL) is given in Figure 6.5.

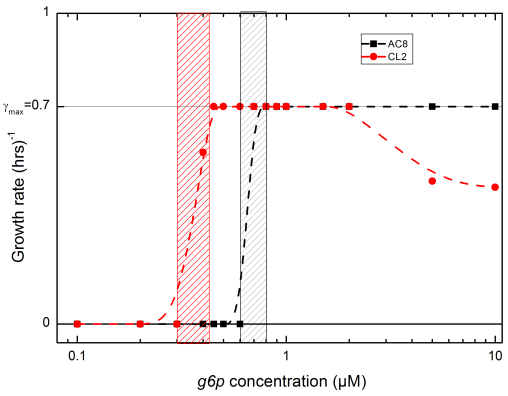

Figure 6.5: Growth rates of CL2 and AC8 as a function of the concentration of g6p. *The black squares and red dots are the estimated growth rates of AC8 and CL2 from the experiment shown above (Figures 6.3 and 6.4). The black and red dashed lines are a guide for an easy reading of the figure. The shaded portions of the figure, represent the transition regions between growth and no-growth.*

A comparison of the behavior of the AC and CL-strains shows a qualitatively identical behavior (Figure 6.5). However, the threshold value of g6p below which growth stops is different for the two strains. The difference has to be due to different strengths of the uhpT promoter at the same concentration of g6p. This could be explained by P_{uhpT} being in a different chromosomal location in the two strains. Such position effects can profoundly affect promoter activity [99, 98]. Another difference is the fact that in the CL-stain the entire *dnaA* operon is under the control of g6p. When fully inducing P_{uhpT}, we overexpress not only DnaA, but also RecF and DnaN. The joint overexpression of these two genes is known to be lethal to the cells [30]. Thus, the slower growth rate of the CL2 strain at high *g6p* concentrations may be attributed to either a much stronger expression of DnaA or to a slight mortality due to the higher concentrations of all three genes of the operon.

6.2 Growth arrest after removal of g6p yields an upper bound on the intracellular concentration of DnaA

We next explored the physiological consequences of arresting the production of DnaA by removing the inducer g6p from a culture of exponentially growing cells. We expect that the cells, depending on the conditions of the preculture, accumulate a reservoir of DnaA of variable size. When the production of new DnaA is stopped, the cellswill continue to divide if their "reservoir of DnaA" is higher than the critical threshold for initiation DNA replication. We can write this hypothesis mathematically, then test it experimentally.

For the *g6p* range where the AC8 and CL2 strains grow exponentially at the maximal growth rate $\gamma_{max} = 0.7\,hrs^{-1}$, equation 6.3 can be re-written as:

$$c = \frac{\alpha \cdot p_{uhpT}\,(g6p)}{[\gamma_{max} + \gamma_{int}]} \tag{6.4}$$

For the AC8 strain, at any inducer concentrations above $0.8\mu M$, the growth rate (as shown in Figure 6.5.) remains constant at γ_{max}. Therefore, we expect that the cells contain an excess of DnaA. The precise value of the excess concentration of DnaA is determined by the concentration of *g6p* in the growth medium according to equation 6.4. When *g6p* is removed from this exponentially growing culture, the bacteria will continue to divide for a number of divisions, dependent on the initial *g6p* concentration, until the concentration of DnaA drops below the threshold value. We can exploit this phenomenon to further estimate the pool of DnaA in the bacteria using an independent experimental strategy.

6.2.1 DnaA is expressed in excess of the minimal concentration needed for cell division

The experimental strategy consists in removing g6p from cultures growing at steady-state in the presence of different concentrations of g6p and measuring the increase in biomass before growth arrest (see Figure 6.6).

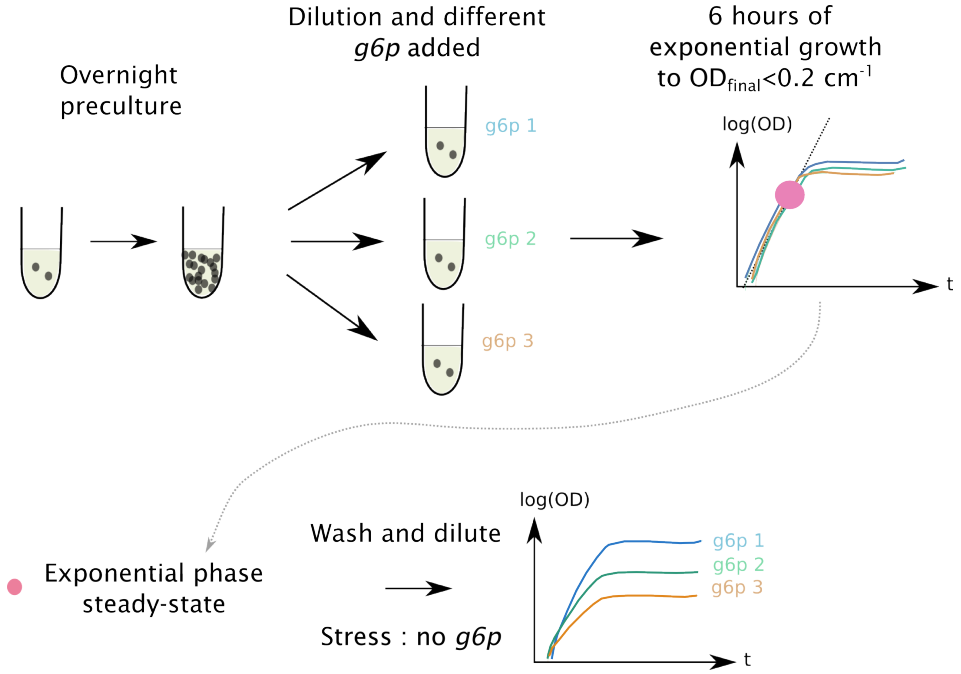

Figure 6.6: Experimental strategy for estimating the size of the excess pool of DnaA in exponentially growing cells. *The preculture of AC8 bacteria was grown in MOPS minimal medium supplemented with 0.2% glucose , 150 μM IPTG and 2 μM g6p. Cells were then strongly diluted and used to inoculat several cultures, each containing a different concentration of g6p. After 6 hours, the exponentially growing cultures were washed twice, resuspended in MOPS minimal medium without g6p, and transferred to wells of a microplate. We measured the growth of these cultures by monitoring the optical density (OD), the green fluorescence (flavins), and the red fluorescence (measuring the mCherry-tagged RNA polymerase). The values of the plateau of these growth curves in the "stress" condition (absence of g6p) are related to the reservoir of DnaA in the initial, exponentially growing cells.*

To ensure steady-state conditions before the removal of g6p, the cultures were inoculated at a very low cell density such that they reach steady-state well before attaining an OD of 0.2 cm^{-1}. At this point, the cultures were transferred to the identical minimal MOPS medium, but lacking g6p (see the protocol details in Section 3.1.5). As a control, and for the compensation of experimental errors, the cells were also grown in the presence of g6p. The media with and without g6p also contained 150 μM IPTG. The different cultures were diluted to the same OD and their growth was measured in a microplate reader. Slight difference in the initial concentrations of cells have been corrected as explained in Section 3.1.7. The results are shown in Figure 6.7

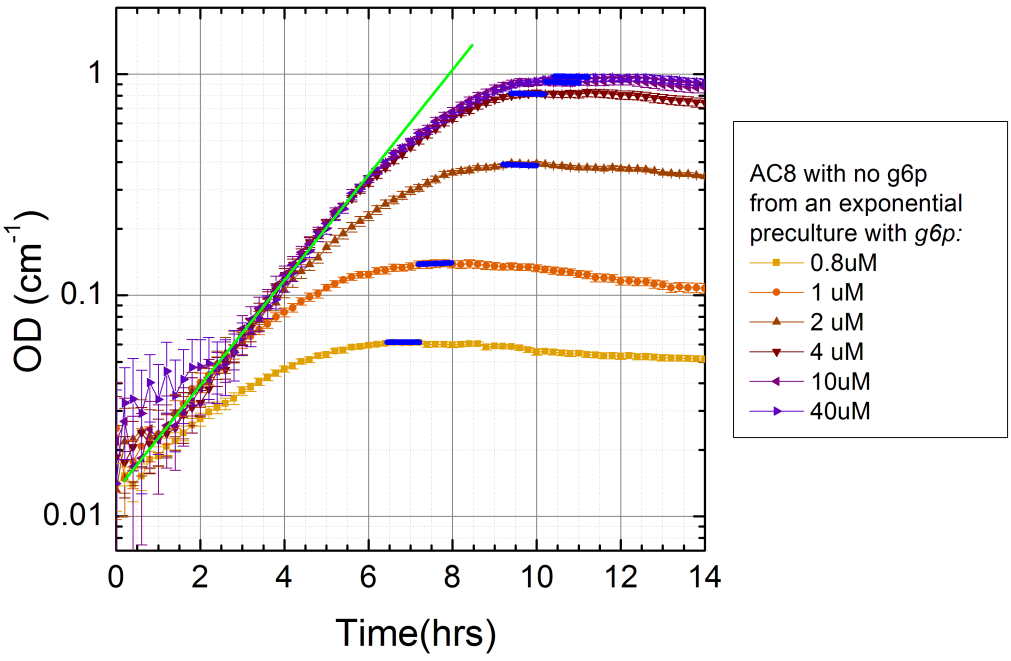

Figure 6.7: OD of AC8 cells growing without g6p after a steady state growth at different concentrations of g6p. *Cells were grown in MOPS supplemented with 0.2% glucose and 150 μM IPTG in a 96 well plate. Beforehand, cultures were grown for six hours in the same medium but with different concentrations of g6p (see legend), at low cell densities, ensuring exponential growth. The green continuous line depicts an exponential growth at γ_{max} (a doubling time of 60 min). The blue lines indicate the points used to calculate the average final OD plotted in Figure 6.10.*

We can see that the cells remain in exponential phase, with an unchanged, maximal growth rate, for a certain number of divisions after the arrest of DnaA production. Cells that were grown with the minimal concentration of 0.8μM g6p appear to slow down division almost immediately after transfer to the medium lacking g6p. Even though the OD continues to rise for up to 6 hrs, it does so in a non-exponential way. This signal increase most likely does not correspond to cell division, but rather to an elongation of the cells. Microscopy experiments (Chapter 7) will confirm this hypothesis. On the other side of the spectrum, at full induction of our promoter prior to DnaA arrest, exponential growth continues for up to 6 hrs before the slowing down.

For intermediate levels of g6p induction, we observe a continuous variation in biomass increase before growth arrest.

While for exponential growth and moderately low cell densities, the OD is an appropriate method to estimate biomass, at high concentration of cells, the OD becomes less quantitative. We have shown previously [101] that the green autofluorescene of the flavins, excreted by the bacteria into the medium, can provide a more reliable estimate of biomass. Furthermore, the excretion of flavins is directly related to the metabolic state of the cell. Therefore, by measuring the change in the green autofluorescence of the stressed bacterial population, we can establish a measurement independent of the OD to monitor cell growth.

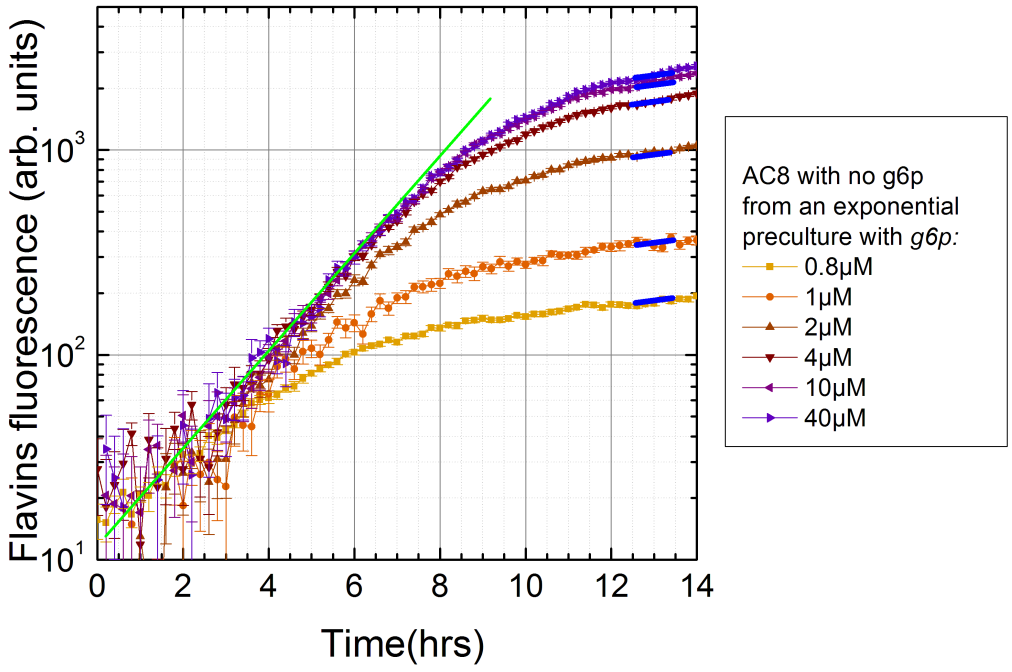

Figure 6.8: Green auto-fluorescence of AC8 cells growing after removal of g6p. *Cells were grown in MOPS minimal medium supplemented with 0.2% glucose and 150 μM IPTG in a 96 well microplate. Beforehand, cultures were grown for six hours in the same medium, but containing different concentrations of g6p (see legend), at low cell densities ensuring exponential growth. The green continuous line depicts an exponential growth at γ_{max}(doubling time of 60 min). The blue lines indicate the points used to calculate the average final green fluorescence plotted in Figure 6.10.*

Indeed, Figure 6.8 shows the same trend as the OD measurement, with the difference that here, the curves maintain a slight increase in slope even at the end of the experiment. This observation suggests that, even though the bacteria stop dividing, they retain a metabolic activity for at least 10 hours after the corresponding OD measurements have reached their maximum. Again, the bacteria induced with the lowest concentration of g6p, 0.8 μM,) decrease their rate of flavin production almost immediately after transfer to the medium lacking g6p. As for the OD measurements, there is a very little difference between the behavior of the cells pre-induced

with 10 and 40 μM g6p, the highest concentrations we used. This observation indicates that we have reached maximal pre-induction of P_{uhpT}, and therefore a a maximal value of the "DnaA reservoir".

We have yet a third signal that measures a quantity related to biomass. The AC-strains express the mCherry-tagged RNA polymerase. Since the concentration of RNA polymerase is constant in steady-state exponential growth, the red fluorescence should also be proportional to the total cellular volume of the culture. The measurements of the red fluorescence of mCherry after removal of g6p are shown in Figure 6.9.

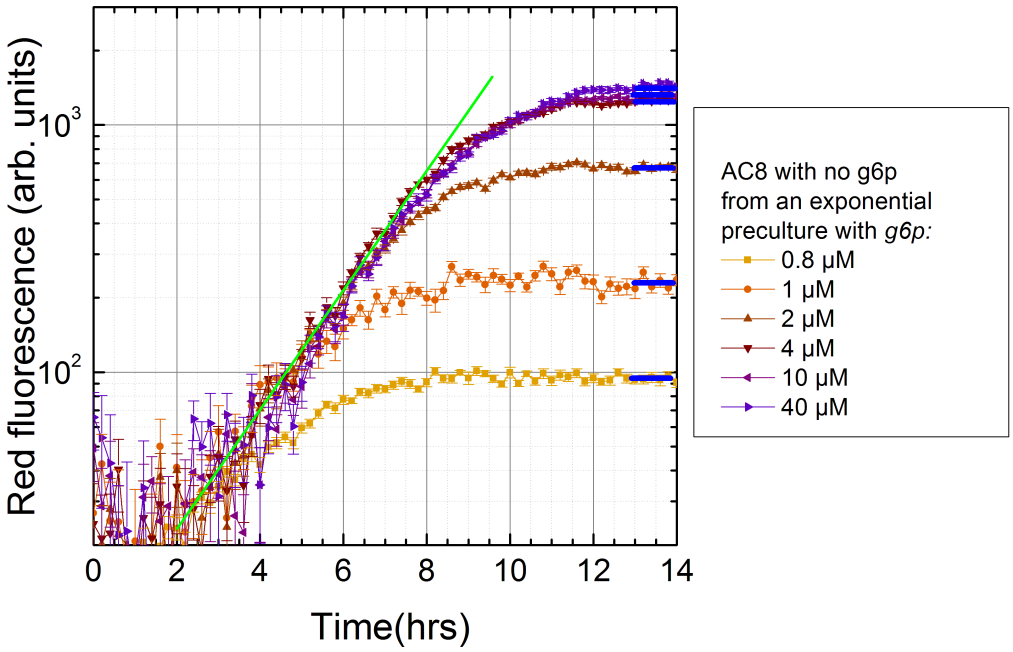

Figure 6.9: Red fluorescence of AC8 cells growing after removal of g6p. *The concentration of the mCherry-tagged RNA polymerase is constant during steady-state growth, making this signal a good indicator of the total number of cells (more precisely: the total cellular volume). Cells were grown in MOPS minimal medium containing 0.2% glucose and 150 μM IPTG, supplemented with different concentrations of g6p (see legends of the curves). At time zero, these cultures were diluted and transferred to a 96 well microplate in an identical medium, but lacking g6p. The green continuous line depicts an exponential growth at γ_{max}(doubling time of 60 min). The blue lines indicate the points used to calculate the average final red fluorescence plotted in Figure 6.10.*

The curves of red fluorescence reach a plateau $\simeq$3 hours after the OD-curves have reached their maximal value 6.7. Even though the cells have stopped dividing, RNA polymerase continues to be produced and/or the mCherry tag continues to mature. Eventually, the red fluorescence reaches a stable plateau value, which, as for the OD and green fluorescence measurements, is directly related to the initial pool of DnaA.

6.2.2 DnaA arrest curves follow the activity described for P_{uhpT}

From the three previous measurements (Figures 6.7, 6.8 and 6.9), we can extract for different g6p concentrations a final value of the OD, the fluorescence from flavins and the fluorescence of mCherry. The data obtained for the highest concentration of g6p in the preculture in each of the curves of Figures 6.7, 6.8, and 6.9 derive from bacteria growing with an initial concentration of DnaA equal to $c_{max} = \alpha \cdot p_{uhpT}(max) / (\gamma_{max} + \gamma_{int})$. The blue bars in these curves indicate the points that were used to obtain the averaged final signal for each type of measurement. According to our model, the number of cell divisions, or the increase in biomass, should be proportional to the initial pool of DnaA in the cells before removal of g6p. This initial pool of DnaA, in turn, should be proportional to the activity of P_{uhpT}.

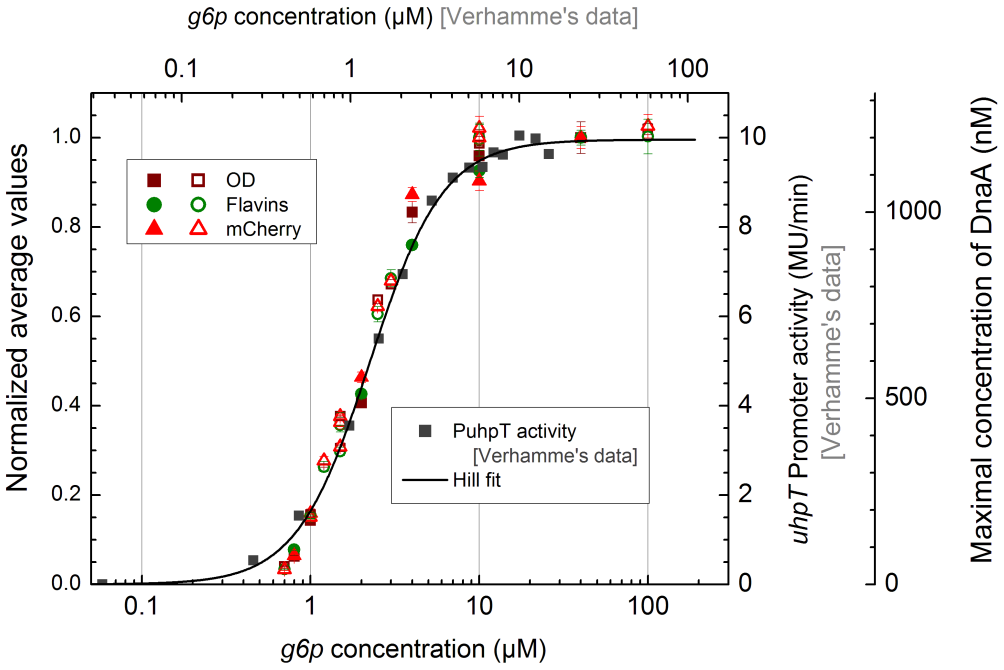

Figure 6.10: Normalized values of the initial reservoir of DnaA. *AC8 cells were grown in minimal MOPS medium supplemented with 0.2% glucose and varying concentrations of g6p. The cells were transferred to a medium lacking g6p and the increase in signals related to the biomass, OD, flavin and mCherry fluorescence (as shown in the legend), were measured. The data were normalized by their value at 40 μM g6p. Solid symbols correspond to the data extracted from Figures 6.7, 6.8, and 6.9, while the open symbols are derived from supplementary experiments (not shown). The first axis on the right shows the previously reported activity of P_{uhpT} (black squares) as measured* Verhamme *et al.* [1]*. The second right axis converts the Miller Units into intracellular concentration of the protein. Since the translation efficiency of dnaA is smaller than the corresponding translation efficiency of uhpT, these values represent an upper bound for the intracellular concentration of DnaA. The black line is a fit of the data obtained by Verhamme to the Hill equation with a Hill coefficient of 2 and dissociation constant for g6p of $K_D = 1.3\,\mu M$. The top axis is related to the bottom axis by a multiplicative factor of 1.7.*

Figure 6.10compares all three type of measurements with the plot of the activity of the *uhpT* promoter (Figure 6.2). We normalized all values by their maximum, in our case the one

obtained for a $g6p$ concentration of 40 μM. Note that in the context of equation 6.4, this maximum corresponds to $\alpha \cdot p_{uhpT}(max)$. On the same graph, we compare our data with the promoter activity previously measured by Verhamme *et al.* [1] and the fit of a Hill function, $P_{uhpT}(g6p) = V_{max} \cdot g6p^n/(K_D^n + g6p^n)$, to these data. The parameters of the fit are: the dissociation constant $K_D = 1.3\mu M$, the Hill coefficient $n = 2$, and the maximal promoter activity $V_{max} = 10\, MU\, min^{-1}$. As expected, our measurements closely match the shape of to P_{uhpT} activity. However, we had to divide our $g6p$ concentrations by a factor of 1.7 to match the data by Verhamme. It is possible that this lower sensibility to $g6p$ is due to a slightly higher phosphate concentration in our MOPS medium compared to the one used by Verhamme. It is known [1] that a high concentration of phosphate in the medium inhibits Uhp induction and a maximal induction is reached at higher values of $g6p$. The competitive inhibition by phosphate modifies the apparent dissociation constant as $K_D(1 + [P_i]/K_i^{P_i})$, with $K_i^{P_i}$ of the order of $1 - 3\, mM$ [1]. Furthermore, although unlikely, we can not exclude the possibility that our batch of g6p was not 100% pure.

The phosphate inhibition is very clearly visible when using the M9 minimal medium, where the phosphate concentration $[P_i]$ =64 mM , compared to 2 mM in MOPS. Since the growth rate is the same in M9 or MOPS minimal media, the steady-state concentration of proteins should also be roughly the same. Equivalent experiments carried out in M9 minimal medium show the same excellent fit of the amount of biomass accumulation after removal of g6p to the activity of P_{uhpT} (Figure 6.10). However, because of the inhibition by phosphate, we had to use much higher concentrations of g6p and we had to rescale the abscissa (concentration of g6p) accordingly. In order to obtain the same activity of P_{uhpT} in M9 medium compared to MOPS medium, we have to employ a 33-fold higher concentration of g6p as shown in Figure 6.11.

Taken together, these results are all consistent with our model of a "DnaA reservoir" (equation 6.4) and confirm that the activity of the $uhpT$ promoter is a good estimator of the steady-state concentration of DnaA in the bacteria. Cells containing concentrations of DnaA between the threshold value below which growth stops and 10-times that value possess identical growth rates. Apparently, a moderate excess of DnaA does not affect the growth characteristics of the bacteria. However, growth stops abruptly when the concentration of DnaA falls below the threshold value obtained by inducing the $uhpT$ promoter with $0.8\,\mu M$ of g6p. This threshold concentration of g6p can be converted to the intracellular concentration of DnaA.

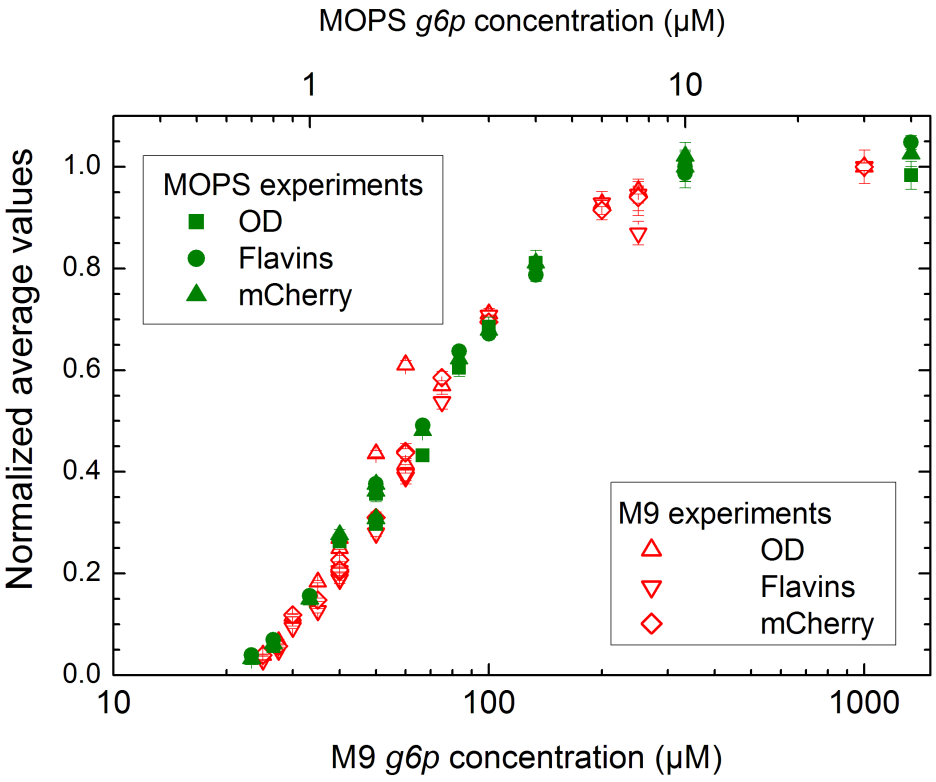

Figure 6.11: Comparison of the normalized values of the pool of DnaA obtained from growth after removal of g6p in two different growth media. *AC8 cells were grown in M9 minimal medium or MOPS minimal medium containing 0.2% glucose and variable concentrations of g6p. The growth curves obtained after removal of g6p were analyzed and plotted as in Figure 6.10. Green solid symbols and red open symbols are derived, respectively, for experiments in MOPS and M9 as shown in legend. The values were normalized, respectively, by their values at 1 mM g6p (in M9) or 40 μM g6p in MOPS. The top axis shows the concentrations of g6p uses in the MOPS experiments, the bottom axis the concentrations of g6p in the M9 experiments. The two axes are related to each other by a multiplicative scaling factor of 33.*

Published experiments have provided more or less coherent estimates of the average number of DnaA molecules per cell during normal growth. One recent report showed that in exponential growth phase, in MOPS minimal medium, the cells produce around 300 DnaA $molecules \cdot generation^{-1} \cdot chromosome^{-1}$ [48]. Taking into account that, for generation time of 60 min, a cell contains on average 1.8 choromosomes [102], we estimate the wild-type concentration of DnaA in exponentially growing cells to be 520 nM. To estimate the absolute concentration of DnaA in our strains, we will convert the know P_{uhpT} activity to numbers of molecules per cell. Such relationship has been established for the β-galactosidase reporter used by [1]. The promoter activity plotted in Figure 6.10 was obtained by Verhamme *et al* by using exactly such a *lac* translational fusion [1].

The activity of P_{uhpT}, reported in Miller Units (MU) / minutes, can be can be converted to a concentration of proteins as follows: $1\,MU \simeq 2$ proteins [9], 1 protein / bacterium $\simeq 1nM$ [103]. Thus, the concentration of a stable protein expressed from the maximally induced *uhpT* promoter

in exponentially growing bacteria with a doubling time of 60 min will be 1.2 μM: $P_{uhpT}^{max} = 10\,MU \cdot min^{-1}$, and therefore $c_{uhpT} = 10 \cdot 2 \cdot 60 = 1200\,nM$. This latter value has been used to convert the activity of P_{uhpT} to protein concentrationon the second right axis in Figure 6.10.

However, since the construction used by Verhamme *et al.* uses a different ribosome binding site (RBS) than the one used for expressing DnaA, we still have to scale the protein concentrations by an unknown, multiplicative correction factor. We can estimate the necessary correction by using a tool developed in the Salis laboratory [104, 105]. Their RBS-calculator allows to reliably predict the strength of a RBS solely based on the sequence. Using this tool, the RBS of our construction is predicted to be 5-fold weaker than the site used by Verhamme *et al.*. While the exact numerical value of the prediction remains debatable, we are confident about the fact that our RBS has a lower translation efficiency. Therefore, the equivalent number of proteins that we infer by the intermediary of the calibration by Verhamme *et al.* serve as an upper bound for the DnaA concentration in *E. coli*. The second right axis in Figure 6.10 therefore represents the upper bound of the intracellular concentration of DnaA.

6.3 Physiological condition of cells after arrest of DnaA production

One of the objectives we fixed ourselves was to study the possibility of keeping the cells from dividing and then allowing them to grow back by increasing their DnaA concentrations. The subsidiary question is: how long can cells remain viable without replicating their chromosome due to the lack of DnaA?

6.3.1 Re-growth after arresting the production of DnaA

To answer this question, we carried out a preliminary study at the level of the population. We grew our cells in M9 minimal medium with 40 μM of g6p and 150 μM of IPTG (see subsection 3.1.5). We then washed the cultures and diluted them into fresh M9 medium supplemented with 150 μM IPTG and continued growth in an Erlenmeyer flask. Eight replicates of this sample were transferred to a microplate for monitoring optical density (OD), red fluorescence (from the fusion of mCherry to RNA polymerase), and the green auto-fluorescence of the cells. These samples monitored in the microplate show the same growth characteristics as the main culture in the Erlenmeyer flask (shown as the black curve in Figure 6.12).

After eight hours of growth in the flask, DnaA becomes exhausted in the bacteria and the growth rate decreased. Starting at this time, we transferred samples from the Erlenmeyer flask to the microplate at regular intervals, supplementing the samples on the microplate with 40 μM g6p. In other words, we re-started the expression of DnaA after a variable time of growth without new production of DnaA. This procedure assesses the capacity of a growth arrested culture to resume normal growth. The results are shown in Figure 6.12. In this experiment, the inoculum was derived from a stationary-phase culture. The size of the initial reservoir of DnaA in these cells is not as well determined as in the experiments described in the previous section, where g6p was removed from exponentially growing cultures. The initial pool of DnaA is probably bigger than the pool of exponentially growing cells. Growth arrest occurs therefore later than in the

previous experiments.

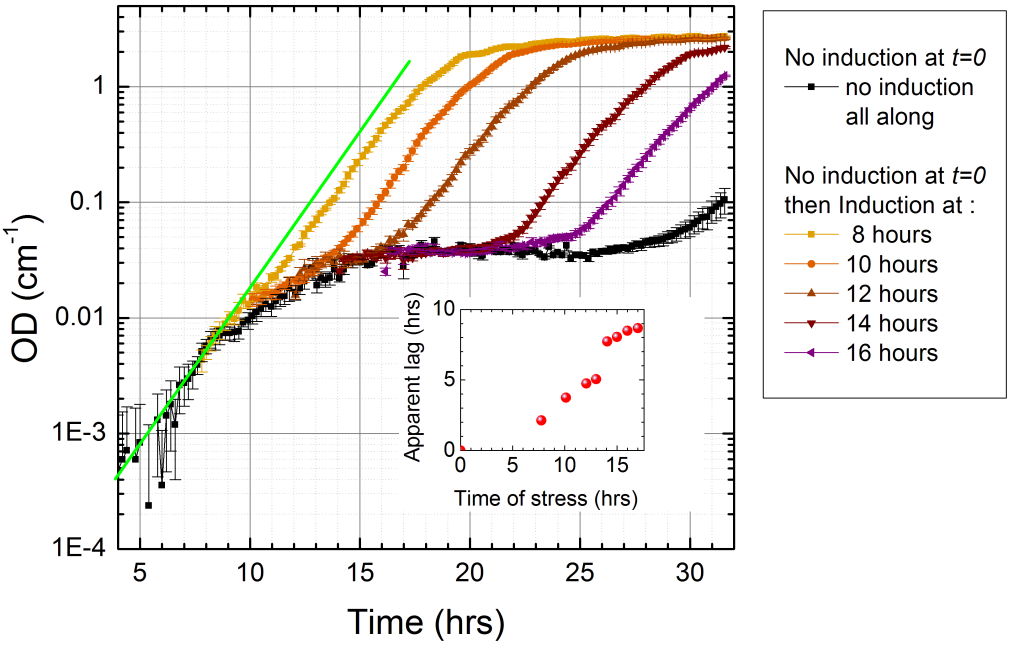

Figure 6.12: Regrowth after prolonged arrest of expression of DnaA. *A pre-culture of AC8 bacteria was grown in M9 minimal medium with 150 μM IPTG and 40 μM g6p. The culture was washed and transferred to a flask containing fresh medium, but lacking g6p. A sample from the flask was added to a microplate in order to monitor growth (black curve). Growth of this sample mirrors the growth of the culture in the Erlenmeyer flask. At regular intervals, a sample was taken from the flask, loaded onto the 96 well plate, and complemented with g6p to a final concentration of 40 μM (colored curves). These samples measure the capacity of the growth-arrested culture to resume normal growth. The green continuous line depicts an exponential growth at γ_{max} (doubling time of 60 min). Insert: a measure of the apparent growth lag as a function of the time after removal of g6p from the culture medium.*

The longer the bacteria remain without producing DnaA, the greater the lag before re-growth after the addition of g6p. Samples taken after eight hours without g6p resumed growth immediately after the addition of 40 μM g6p. At this stage, the growth rate of the original culture had hardly decreased. However, after nine or more hours without production of DnaA, growth rate decreases considerably (black curve in Figure 6.12) and we observe a "lag" before the culture resumes growth. The lag increases for samples taken up to 16 hours after g6p removal, then remains constant (not shown in the figure). The apparent lag stabilizes at about eight hours (see Insert in Figure 6.12). Two extreme interpretation of the observed lag are possible:

1. All cells of the population enter a physiological state that renders growth difficult. The lag-time corresponds to the time needed to reverse this physiological state

2. Only a small portion of the cells remains viable and these cells resume normal growth as soon as g6p is added to the medium.

6.3.2 Viability of cells after a prolonged arrest of production of DnaA

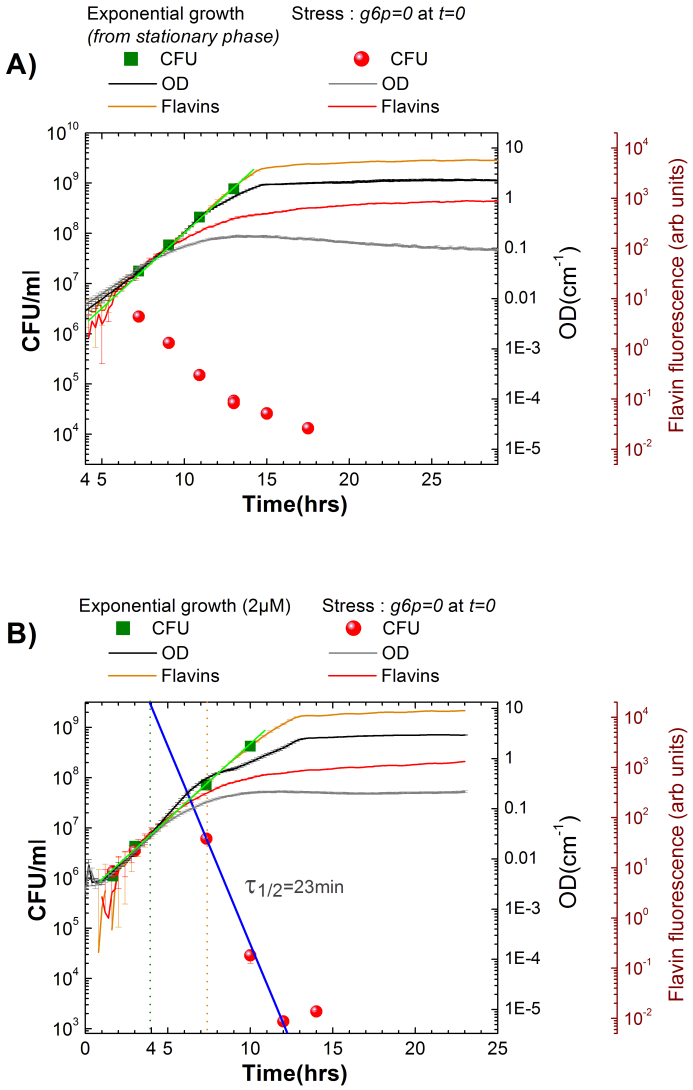

Figure 6.13: CFU of AC8 bacteria growing in the presence or absence of g6p. A) M9 minimal medium. *A stationary phase culture of AC8-cells in M9 medium containing* $40\,\mu M$ *g6p and* $150\mu M$ *IPTG was washed and diluted into two flask containing the same M9 minimal medium supplemented with* $150\,\mu M$ *IPTG. One flask also contained* $40\,\mu M$ *g6p, the other did not. Samples were taken from both flasks at different times and spread onto M9-agar Petri dishes supplemented with* $150\mu M$ *IPTG and* $40\,\mu M$ *g6p. Colonies were counted after incubation at* $37°$ *for about one day (green squares: +g6p, red circles: -g6p). The symbols are at the size of the error bars from two replicas. The green line indicates exponential growth with doubling time of 60 min. At the same time, two samples from the flasks with and without g6p were also transferred to a microplate in order to monitor growth continually : respectively, in black and gray (for the OD, first right axis) and orange and red (for the flavins fluorescence, second right axis).* **B)** *Same as A), except that the growth medium was MOPS , the initial pre-culture was in exponential growth with* $g6p = 2\,\mu M$*. The continuous blue curve shows an exponential death rate with a 23 min half time. The first vertical dashed bar marks (in green) growth arrest at t=4hrs, the second vertical bar (in orange) marks the time at which the cells begin to die (∼7.5hrs).*

We can directly test the second hypothesis by measuring cell viability. The basic setup of the experiment is simple: remove samples at regular intervals from bacteria that were transferred to a medium lacking g6p at time zero. The samples are plated on Petri dishes containing g6p and all ingredients necessary for growth. The number of colonies formed in this type of experiment is a direct measure of the number of viable cells.

We did two sets of experiments: one in M9 minimal medium, starting from an overnight pre-culture in stationary phase and a second one, with the initial state was much well defined. The pre-culture was done in MOPS minimal medium in 2 μM g6p and inoculated at a high dilution, ensuring that the cells were in steady-state exponential growth at the time of removal of g6p. The results for both experiments are shown in Figure 6.13

For both growth conditions we observe an exponential growth with a doubling time of 60 minutes for all signals: CFU (green squares), flavin fluorescence (orange line), and (in a more limited range) OD (black line). For stress conditions, starting $t \sim 7hrs$, the CFU number (red dots) show a dramatic decrease, the OD (gray lines) and flavins fluorescence (red lines) increase more slowly: the cells are dying, faster in the second condition (Figure 6.13B) than in the first. It seems therefore that the second hypothesis, of cell becoming less and less viable while they are growth arrested, is the only plausible one.

For a more detailed description, of the different steps of how that might happen, we will only refer to the second experiment (Figure 6.13B). The cells grow exponentially for a couple of hours (CFU red dots) after removal of g6p, then stop dividing (at $t{\sim}4hrs$). Between 4-7.5hrs, the OD (gray line) keeps increasing, but not exponentially. The flavin fluorescence (red line) keeps growing exponentially, then, around $t \sim 6hrs$, the increase slows also. A relevant scenario could be that at around $t \sim 4hrs$ the cells stop dividing but still increase their volume exponentially for 2 hours. Starting $t \sim 7.5hrs$, they arrive in a state that becomes more and more difficult to recover from. The decrease in the number of viable cells is fast with a halftime of 23 minutes (blue line). It is interesting to note that in the end we still have a residual population of viable cells, with $\sim$0.04% of the population remaining viable 15 hours after the removal of g6p.

The loss of viability of the cells could be due to the increased expression of DnaN and RecF in strain AC8. When the concentration of DnaA decreases, the activity of the native DnaA promoters increases because of negative auto-regulation. These promoters transcribe the entire *dnaA* operon (see Figure 2.1), including *dnaN* and *recF*. The simultaneous over-expression of DnaN and RecF leads to cell mortality [30]. This effect could explain why the majority of our cells do not grow back.

To directly assess the effects of DnaN and RecF on viability, we limit their expression at the same time as the expression of DnaA. This amounts to comparing the behavior of strains CL2 (the entire *dnaA* operon, including *dnaN* and *recF*, is under the control of P_{uhpT}) and AC8 (only *dnaA* is expressed from P_{uhpT}, while the auto-regulation of the promoters upstream of *dnaN* and *recF* could potentially lead to an over-expression of these genes). We grew the bacterial cultures in MOPS minimal medium supplemented with 150 μM IPTG . The AC8 strain was grown in the presence of 0.8 μM g6p, while the CL2 strain was grown in the same medium supplemented with 0.5 μM g6p. This couple of inducer concentrations produced equivalent growth curves of the samples after removal of g6p (cyan and brown curves in Figure 6.14).

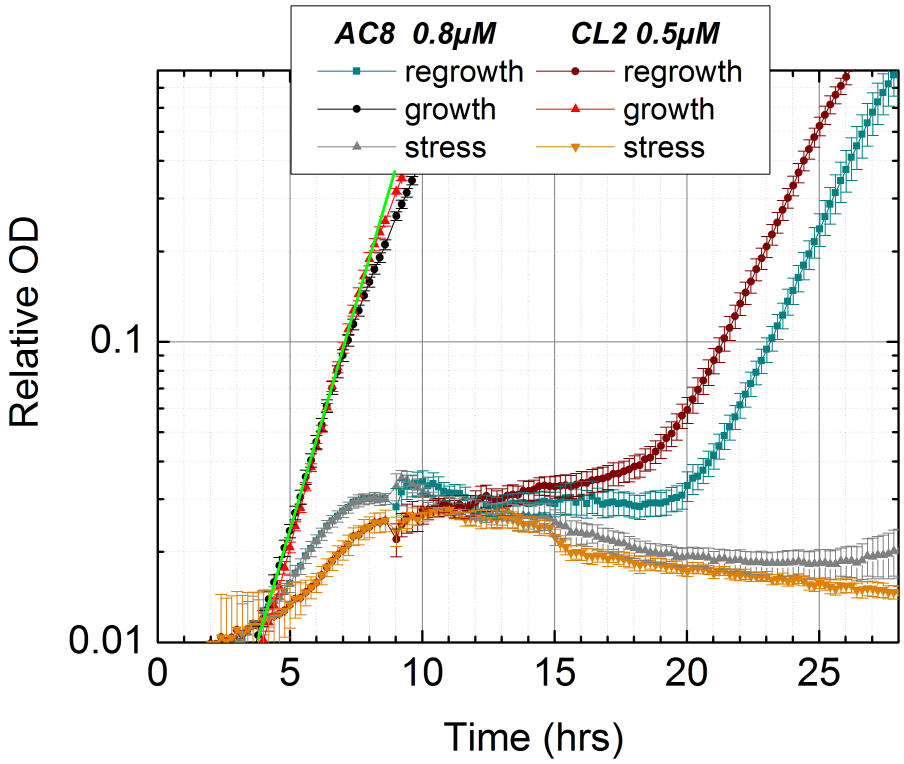

Figure 6.14: ***Regrowth of strains AC8 and CL2.*** *Both strains were grown in MOPS minimal medium supplemented with 150 μM IPTG, and respectively 0.8 and 0.5 μM g6p. After reaching an exponential steady-state, the cultures were washed and diluted into fresh medium containing* 150 *μM IPTG and no g6p. g6p was added to a final concentration of* 2 *μM after 9 hours. The green line shows a doubling time of 60 minutes.*

The growth curves of the two strains in the absence of g6p are very similar (figure 6.14). In particular, using the different initial concentrations of g6p makes them stop growing at the same OD. We conclude that their initial pool of DnaA was roughly identical. As discussed before, the difference in concentration of g6p needed to obtain the same growth profile for the two strainsis due to differences in the activity of P_{uhpT} which could depend on the site of integration into the chromosome.

When we added g6p nine hours later, both strains resumed growth after a long apparent lag, with the CL2 strain re-growing slightly earlier. Again, this behavior suggests that most of the cells have lost their viability, with the CL2 strain slightly less affected. However this improvement is small compared to the extended apparent lag in both strains. Hence, producing a much lower quantity of DnaN and RecF proteins, as is the case for the CL2 strain compared to AC8, did not improve viability significantly. We therefore conclude that an overabundance of DnaN and RecF proteins is not the main factor triggering the loss of viability in cells not replicating their chromosome.

To test the more general hypothesis that decreasing at the same time the production of DnaA

and the production of other cellular proteins might protect the cells from dying, we decided to limit global protein expression at the same time as shutting off the transcription of *dnaA*. In order to do so, we decreased the number of RNA polymerases in the cells by decreasing the concentration of IPTG after stopping the expression of DnaA. We transferred the pre-culture, grown in the presence of 150μM IPTG and 40μM to a flask containing fresh medium M9 without g6p. We grew the bacterial culture for two hours, then washed the cells again in order to remove IPTG from the medium. We then inoculated four different flasks containing M9 minimal medium with these bacteria. The media were supplemented with IPTG at final concentrations of 30, 40, 50 and 60 μM. As before, a sample from each flask was transferred to a microplate in order to monitor growth. The main culture in the flask was incubated at 37°. Seven and fifteen hours after the start of the experiment, we took a sample from each flask, added 100 μM of IPTG as well as 40 μM of g6p and transferred eight replicas to the microplate. While the samples harvested after 7 hours in the restrictive growth conditions grew back rapidly, the samples from 15 hours took a very long time to grow back. There was almost no difference in the lag before regrowth between the different flasks. In Figure 6.15 we show the results from the flasks containing 30 and 60 μM IPTG.

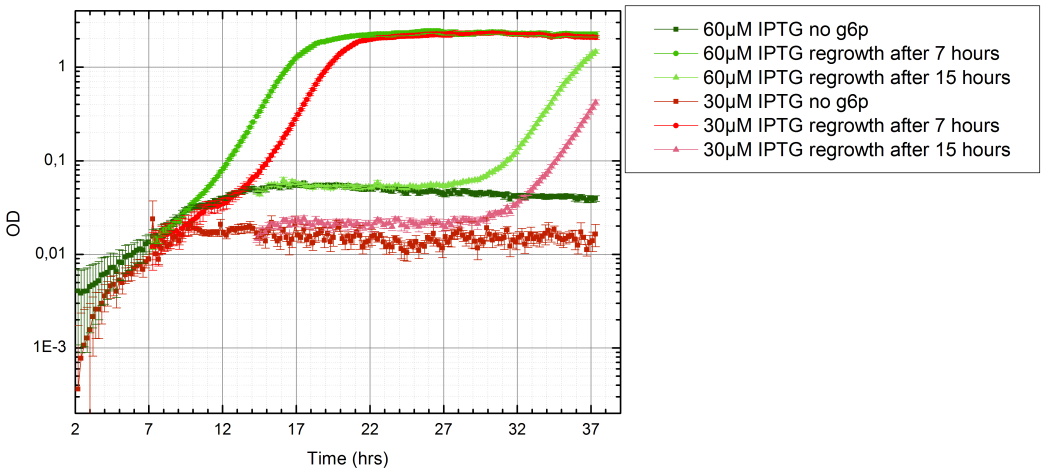

Figure 6.15: Regrowth of cells not expressing DnaA and limited expression of RNA polymerase. *We grew a pre-culture in M9 minimal medium supplemented with 150 μM IPTG and 40 μM g6p. Cells were washed and grown in fresh medium containing 150 μM IPTG for 2 hours, then washed again and transferred to flasks containing minimal media supplemented with different concentrations of IPTG. Samples were taken from the flask after 2, 7 and 15 hours since the beginning of the experiment. 100 μM IPTG and 40 μM g6p were added to the samples transferred to the microplate in order to monitor re-growth. The graph shows the growth of the original sample (no addition of g6p or IPTG, dark colors), and samples taken after 7 and 15 hours of growth without g6p.*

Decreasing the concentration of RNA polymerase in the cells did not decrease the lag before re-growth. We explored other timing schemes between reducing the expression of RNA poly-

merase and shutting of the transcription of *dnaA* (data not shown). For example, we limited RNA polymerase expression at the same time as shutting off *dnaA* transcription. None of these experiments diminished the lag for re-growth.

The absence of DnaA expression clearly affects cells within the population in different ways. It is very likely that when we grow them without g6p some cells stop dividing while others continue to grow for a little longer. In a similar way, we have sub-populations that appear when we grow back our cells. Some cells resume exponential growth while the majority is no longer viable, some cells retain their metabolic activity and excrete flavins into the medium (Figure 6.8) but others could possibly die. Thus we need to observe single cells within the population in order to give a complete description of our cells.

Chapter 7

Single cell Analysis

Protein expression in cells is a stochastic process. Therefore, the concentration of DnaA in individual cells is described by a random distribution, determined by the activity of P_{uhpT} and other cellular factors. When we stop the production of DnaA, cells divide for a certain number of times before ceasing to divide. The number of divisions depends on the initial reservoir of DnaA proteins in each individual cell. Thus, we expect cells to divide a variable number of times depending on the size of their initial DnaA reservoir. If, during growth, the average DnaA concentration of the population is close to the minimal value needed for DNA replication, some cells could stop dividing, even in presence of g6p. The number of cells that stop growing depends on the initial DnaA distribution in the population. This distribution can be traced by observing individual cell growth without DnaA expression. We used our AC8 strain for these experiments since we already characterized the average response of this strain to varying activities of P_{uhpT}.

7.1 Exponentially growing cells at low concentrations can stop dividing

We begin by observing how individual cells of a population behave during exponential growth. According to the basic model of the functioning of DnaA, chromosome replication should stop when the concentration of DnaA falls below a critical threshold. Depending on the shape of the distribution of the concentration of DnaA in individual cells of the population, a certain number of them would contain a sub-threshold concentration of DnaA. They should stop dividing even in conditions where the transcription of *dnaA* is still active. This phenomenon should become more prominent the closer we are to the minimal induction concentration, $0.8\,\mu M$ of g6p, that still enables exponential growth. We decided to investigate this type of growth arrest by observing cells growing in different concentrations of g6p for an extended period of time.

We used time lapse microscopy coupled with our micro-fluidics device in order to observe a large number of exponentially growing cells with different concentrations of g6p. We used MOPS minimal medium supplemented with $150\mu M$ of IPTG and $2\,\mu M$ of g6p to grow the cells for at least 16 hours before changing the concentration of g6p. In a first step, the concentration of g6p was lowered to $1\,\mu M$, then finally to $0.8\,\mu M$. We analyzed 20 different frames, each containing the image of about 15 channels, and took phase contrast and red fluorescence images every

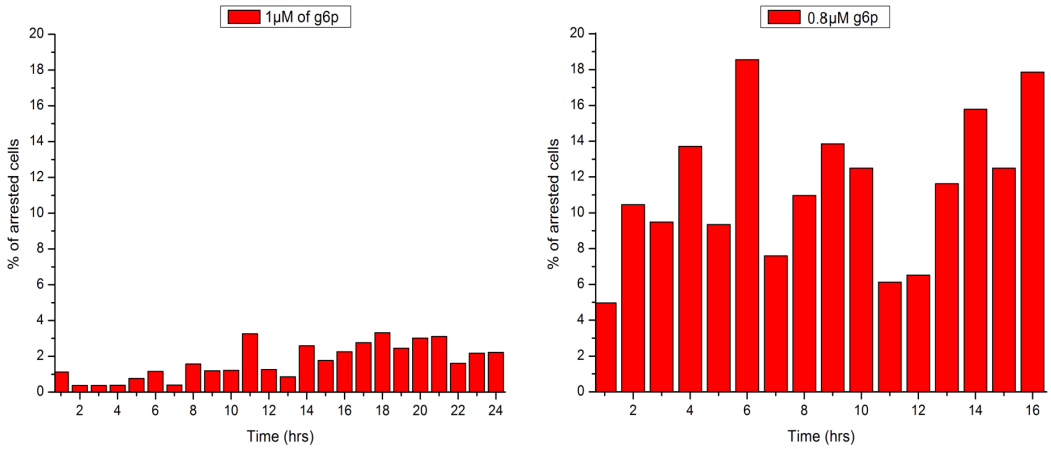

Figure 7.1: Probability of division arrest per hour during exponential growth with different concentrations of g6p. *Using time lapse microscopy coupled with microfluidics, we observed 294 cells growing in the presence of varying concentrations of g6p. We counted cells that permanently stopped dividing and recorded the time corresponding to the last division. Cells started growing in the presence of* $2\,\mu M$ *of g6p. Only 4 cells stopped growing after 24 hours. The same cells then grew with a concentration of* $1\,\mu M$ *of g6p. After 24 hours, only 71 cells kept growing. Finally, the concentration of g6p was lowered to* $0.8\,\mu M$ *of g6p. After 16 hours, only 15 cells kept growing in this condition. The histograms show the probability per hour, which corresponds to one generation of exponential growth, that a cell will stop dividing.*

20 minutes with the lowest illumination settings possible to avoid any photochemical stress. The entire experiment took three days.

Due to time constraints, we could not analyze the behavior of all the cells in the channels. It was also very difficult to differentiate between normal cells and cells that diminished their growth rate for some minutes because we only dispose of three images per cell cycle. We therefore only analyzed the cell at the bottom of each channel. After a cell had stopped dividing definitely, we recorded the time of its final division. All exponentially growing cells had a doubling time of 60 minutes. The statistics of "spontaneous" growth arrest is represented in the histograms of Figure 7.1. The histograms show the probability for a cell to stop dividing at different concentrations of g6p (and therefore production rates of DnaA) in the growth medium.

Cells growing in the presence of $2\,\mu M$ of g6p very rarely stopped dividing., In all our frames, we only identified four of them in a period of 24 hours. However, when the g6p concentration was lowered to $1\,\mu M$, some cells stopped dividing. A cell that had stopped dividing would either resume growth after a while or completely stop dividing for the rest of the experiment. The number of cells that stopped dividing was further increased when the g6p concentration was lowered to $0.8\,\mu M$ and once a cell stopped dividing, very rarely would it resume exponential growth. Our data show that a sizable portion of cells stop division completely. We estimate the probability for a cell to stop dividing when we induced P_{uhpT} with $1\,\mu M$ of g6p to be $2.00\pm0.05\%$. The data also show an apparent increase in the probability of cells stopping to divide during the 24 hours of the experiment (Figure 7.1). We will discuss possible causes for this observation in the Discussion section. The probability of division arrest increases significantly when the

concentration of g6p is lowered to $0.8\,\mu M$, attaining $10 \pm 1.5\%$ per hour.

Some cells resumed division after having stopped once. However, most of these cells only divided two or three more times before arresting division completely. This suggests that cells that had stopped dividing are more likely to return to the non-dividing state than the average cell in a given environmental condition. In molecular terms, we can interpret this behavior in terms of the threshold model of the concentration of DnaA. The concentration of DnaA drops below the threshold, leading to division arrest. However, the continued synthesis of DnaA can raise the cells above the threshold level again.

This "recovery" is limited in time. All cells that resumed normal exponential growth after an initial arrest of cell division, did so one or two hours after they stopped dividing. We observed no cells resuming growth after an extended period of arrest of cell division. This is compatible with our previous observations, cells that stop dividing for three or so hours rapidly loose their ability to divide again. It is still surprising in the context of the simple threshold model of DnaA functioning, as we expected cells would accumulate a sufficient quantity of DnaA and trigger new cell divisions before three hours have elapsed.

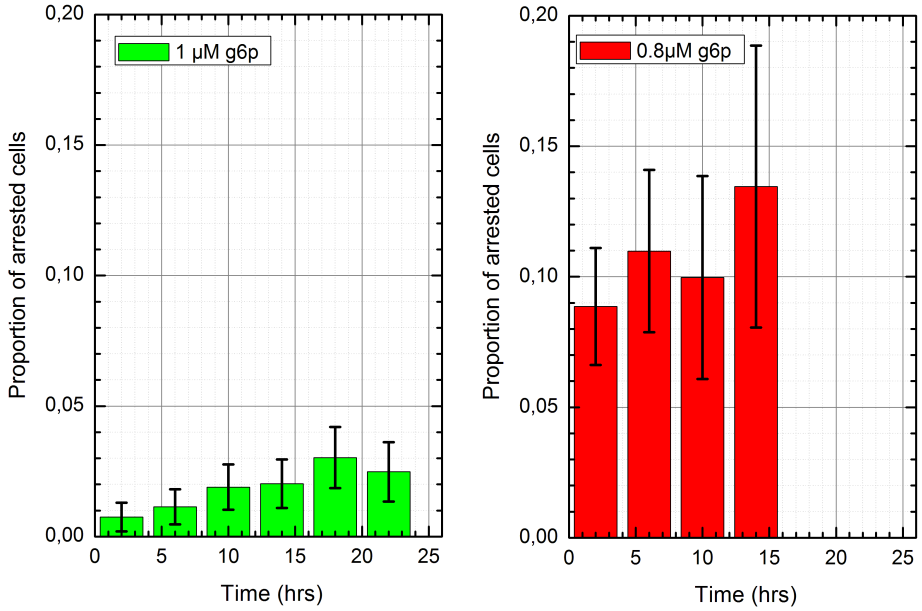

Figure 7.2: Grouped measurements from probability of division arrest in growing cells. *We grouped the measurements from Figure 7.1 in 4 hour bins in order to estimate a 95% confidence interval and evaluate weather cells stop dividing more frequently the longer they stayed in medium with* $1\,\mu M$ *and* $0.8\,\mu M$ *g6p.*

The percentage of cells that stop growing seems to increase as time goes on. We grouped our observations into four hour bins in order to estimate a confidence interval at 2σ and evaluate if the increase is an artifact or a real effect. The grouped measurements are shown in Figure 7.2, but weather there is a trend in our measurements is inconclusive. The confidence intervals from our measurements at $1\,\mu M$ overlap with each other, although it does seem that there might be

a small increase when comparing the first and last bin. The same can be said for the $0.8\,\mu M$ measurements. Stopping DNA replication could induce continuous damage to the cell, if the cell is damaged beyond a certain point it could loose its viability. Depending on how the damage is done, it could take some time for the first generations to accumulate enough defects thus creating a trend in the beginning of the measurements. When cells stop dividing, it is possible that they enter a dormant state, loosing their viability, or simply die after several hours without cell division. We will investigate this possibility below.

Within the resolution of our experiment, we never observed cells that grew slower than the canonical growth rate for more than one or two hours. In other words, the cells either grow at the maximal growth rate of one division per hour or they stop growth entirely. Intermediate values of the growth rate appear only transiently. This behavior resembles a bistable system, with two stable states: exponential growth at the maximal growth rate or no growth at all. Once a cell is locked into one of the two stable states, it tends to stay there.

7.2 Mapping the distribution of the intracellular concentration of DnaA

We observed that cells grown with $2\,\mu M$ of g6p grow just as well as the majority of cells grown with $1\,\mu M$ of g6p. However, the average DnaA concentration in the cells is higher at the higher concentration of g6p. We therefore hypothesize that at higher activities of P_{uhpT}, the cells accumulate a reservoir of DnaA. The size of this reservoir can be several times the minimal amount needed for cells start replicating their DNA. Can we estimate the size of this reservoir?

7.2.1 Cells not expressing DnaA in a population stop dividing at different times

By transferring exponentially growing cells to a medium lacking g6p, we should stop the production of further molecules of DnaA. In the context of our simple model of DNA replication, all remaining replication of the chromosome and cell division would have to rely on the accumulated reservoir of DnaA. Experimentally, we therefore need to observe individual cells after shutting off the activity of P_{uhpT}. Analyzing many such cells will yield statistics about the distribution of DnaA molecules in these cells.

We monitored AC8 in a microfluidics device using time lapse microscopy. Initially, the device was connected to a bottle with MOPS minimal medium supplemented with $150\,\mu M$ of IPTG and $30\,\mu M$ of g6p. This condition allowed exponential growth at the maximal growth rate of about one division per hour. We let cells grow for five hours in the medium containing $30\mu M$ of g6p, then we switched to an identical growth medium, but lacking g6p and continued our observations for another 18 hours. The temperature was maintained at 37° throughout the experiment. We took both phase contrast and red fluorescence images of twenty different frames every ten minutes. Between measurements, the cells were not illuminated to avoid inducing a potential stress due to photoreactions.

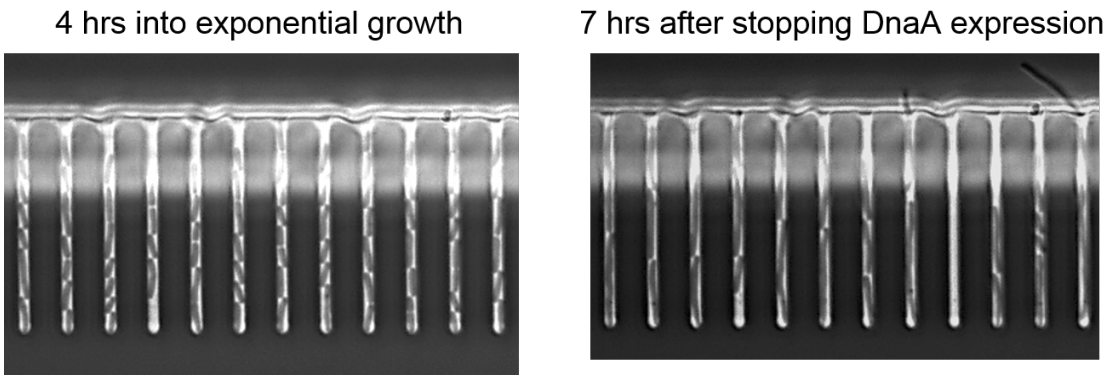

Figure 7.3: Time lapse microscopy image of AC8 growing in a microfluidics device. *Typical phase-contrast images of one particular field of the microfluidics device containing cells during exponential growth (left) and seven hours after the removal of g6p (right). The cells are more heterogeneous in size and, on average, much longer after growth arrest (right). The bacteria were grown in MOPS minimal medium supplemented with* $150\,\mu M$ *of IPTG and* $30\,\mu M$ *of g6p for the first 5 hours of growth, and* $0\,\mu M$ *g6p for the rest of the experiment. Cells were at* $37°$ *for the entire duration of the experiment. A phase-contrast image and a fluorescence image (not shown) were taken every ten minutes.*

Four to seven divisions after the removal of g6p, the cells stopped dividing and started to become filamentous, i.e., even though division no longer occurred, the length of the individual cells increased. Later in the experiment, some cells stopped elongating and seemingly "exploded", while others were lost into the central channel before their elongation stopped. Figure 7.3 shows a frame with cells during exponential growth and seven hours after the removal of g6p.

This basic experiment gave us some key information. (i) There is a great variability in our population since cells divided between four and seven times before stopping. (ii) Even though cell division stops, our cells continue to elongate, increasing their biomass. This was also observed for cells that stopped dividing during exponential growth conditions in subsection 7.1 (iii) A portion of the arrested population dies after a very long stress. This complements population analysis where we determined that a large portion of cells are still metabolically active after several hours without DnaA expression.

7.2.2 Distribution of descendants from individuals not expressing DnaA

We want to quantify the variability of the number of divisions after removal of g6p. The number of divisions provides an indirect measure of the DnaA reservoir before stopping P_{uhpT}. Even though these measures do not determine the exact number of DnaA proteins in the cell, we can establish an upper bound on the average number of DnaA proteins per cell.

To determine the exact number of divisions an individual cell undergoes, we must keep track of all its descendants. During cell division, the proteins are not distributed in a perfect 50/50 ratio to the two daughter cells. Monitoring just one line of descendants is therefore not sufficient. Our mother machine is therefore not suitable for this type of measurement because many of the daughter cells get washed out into the central channel.

We therefore transferred cells grown in liquid MOPS minimal medium supplemented with $150\,\mu M$ of IPTG and varying concentrations of g6p to MOPS agar pads lacking g6p. Each indi-

vidual cell will form a micro-colony. The observation of the agar pads by time lapse microscopy yields directly the number of daughter cells of one particular mother cell. Initial growth in the liquid medium (supplemented with $150\,\mu M$ of IPTG and varying concentrations of g6p) lasted for six hours in order to establish steady-state conditions. When the OD_{600} of the culture reached ~0.1, we washed the cells with fresh medium lacking g6p as explained in subsection 3.1.5. Then we deposited the cells on the pads. The time lapse experiment was performed with the cells kept at 37° for 6-20 hours (depending on the g6p concentration during exponential growth). The cells underwent a defined number of divisions before stopping and the majority of the cells continued to elongate for more than 10 hours. At this point the cells were so long that the field of view of the microscope was overrun. Figure 7.4 shows an image of the cells during the experiment.

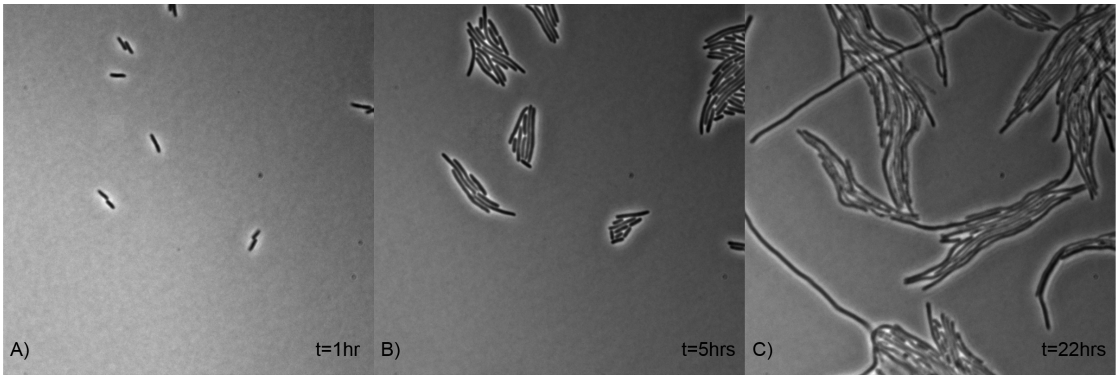

Figure 7.4: Time lapse microscopy of AC8-cells growing on pads without g6p. The pictures show phase-contrast images of bacteria grown on agar pads at 37° for the indicated amount of time. Pre-cultures were done *in MOPS minimal medium supplemented with* $150\,\mu M$ *IPTG and* $2\,\mu M$ *g6p. At time zero, the liquid cultures were transferred to the agar pad lacking g6p. Before being loaded onto the pads, cells were washed with growth medium lacking g6p. Time lapse images were taken* one hour after being loaded onto the pad (A), five hours after loading (B), or 22 hours after loading (C).

We analyzed 40 frames for three different concentrations of g6p during the exponentially growing pre-culture: $0.8\,\mu M$, $1\,\mu M$ and $2\,\mu M$. At concentrations above $2\,\mu M$ g6p, the cells continued to divide too many times for a proper analysis on an agar pad. The cells in a micro-colony started to overlap, preventing us from reliably counting the number of divisions.

Arrested cells kept growing on the pads for many hours. Even though they stopped dividing, they kept elongating. A small portion, about 5% of the cells, disintegrated for unknown reasons during this extended incubation without g6p. Almost all cells in Figure 7.4 kept elongating for at least 12 hours after having stopped the production of DnaA. After 22 hours, the the majority of cells stopped elongating. However, a small portion kept growing. At all three concentrations of g6p, some individuals divided more often than others. We observed filamentous cells already in the starting population derived from the pre-culture grown in the presence of $1\,\mu M$ and $0.8\,\mu M$ of g6p. There were few such cases in the $1\,\mu M$ pad, and many in the $0.8\,\mu M$ pad. This observation confirms that $0.8\,\mu M$ of g6p leads to an activity of P_{uhpT} that produces a number of DnaA molecules just above the threshold necessary for initiating DNA replication.

In order to quantify the distribution of the number of divisions of individual cells, we plotted a histogram of the number of descendants of an initial mother cell in a micro-colony. In other

words, we counted the total number of daughter cells derived from one single cell at the time when all of the cells in the micro-colony had stopped dividing. We showed previously that at each generation a small proportion of the population stop dividing even in exponential growth conditions when the concentration of g6p is below $2\,\mu M$. Since we want to map the distribution of the DnaA reservoir in the cells by counting divisions, we need to remove from our initial count the cells that had already stopped dividing prior to the arrest of P_{uhpT}. Otherwise, the first point of the distribution would include not only cells that stop dividing after one generation, but also the cells that had already stopped dividing in the preculture. Figure 7.5 shows the histograms of the counts. Precultures containing three different concentrations of g6p were transferred to the agar pads when the OD had reached a value ~ 0.1.

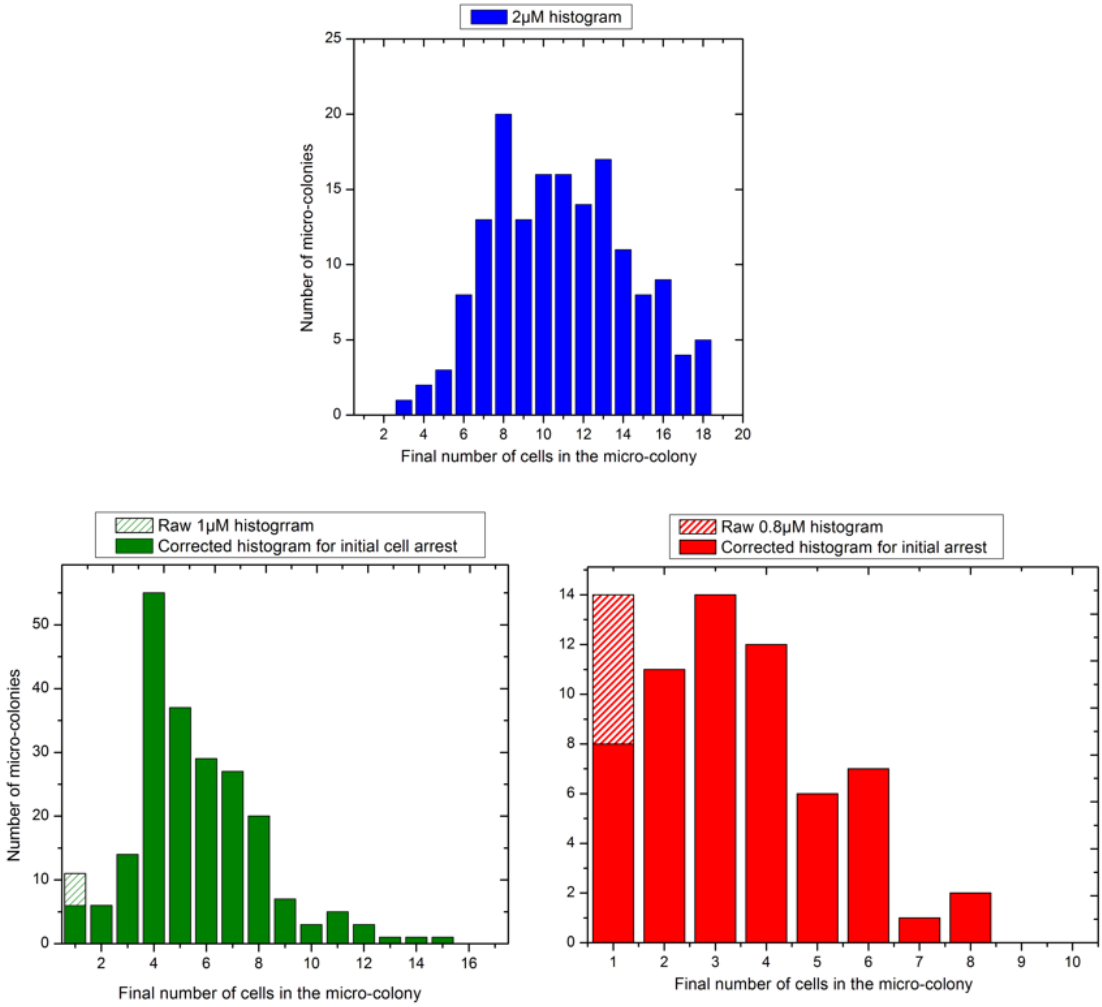

Figure 7.5: Number of cells in a micro-colony without production of DnaA. *Cells were grown in liquid MOPS minimal medium, supplemented with* $150\,\mu M$ *IPTG and the indicated concentrations of g6p for six hours. At time zero, the cells were transferred to an agar pad lacking g6p and the number of cells in a micro-colony initiated by a single mother cell were counted when all cell division has stopped. We counted isolated individuals in 40 different frames. This corresponds to 160 cells for the* $2\,\mu M$ *pad, 220 cells for the* $1\,\mu M$ *pad and 64 cells for the* $0.8\,\mu M$ *pad. Using the previously determined probability that a cell stop dividing during exponential growth, we corrected the first point of the histograms in order to include only the cells that stopped dividing after* P_{uhpT} *as arrested. The correction is shown as the shaded part of the bar.*

The histograms show a great variability of the number of offspring of an individual cell. Since cell division depends on DNA replication, which depends on a sufficient amount of DnaA, the variability in the number of offspring reflects the variability in the initial pool of DnaA proteins in the cell. Thisvariability is present in cells during exponential growth. The distributions of the division counts of cells grown with $1\,\mu M$ or $0.8\,\mu M$ of g6p are clearly skewed to the left. This observation agrees with the conclusion of the previous section: some individuals do not contain enough DnaA for replicating their chromosome.

7.3 Analyzing the distributions

We interpret the histograms in Figure 7.5 based on the hypothesis that the observed variability of number of divisions reflects the distribution of the number of DnaA proteins reservoirs in the cell. DnaA is expressed by the *uhp* induction system during exponential grow and we assume furthermore that the expression of P_{uhpT} stops rapidly after removing the inductor molecule g6p. Nonetheless we can not compare the DnaA distribution in wild type cells with our constructions. Our histograms reflect the expression of P_{uhpT} during exponential growth, not P_{dnaA}. Almost all proteins distributions in *E.coli* [106]follow a Gamma distribution :

$$p(x) = \frac{1}{b\Gamma(a)} \left(\frac{x}{b}\right)^{a-1} e^{-x/b} \tag{7.1}$$

The gamma distribution has two parameters a and b . The first one, a is a number, it is independent of the measurement unit of x and establishes the shape of the distribution. For $a \leqq 1$ this distribution is monotonously decreasing , while for $a \gg 1$ it has a bell-shape. The other parameter b , is a scale parameter. It has the same measurement unit as x and defines the exponential tail of the distribution. These two parameters have been linked to the number of mRNA/cell cycle (for a) and to the number of proteins/mRNA lifetime (for b) when the gamma distribution is used to describe the protein distribution in bacteria.

We fitted our histograms (Figure 7.5) using this distribution, results are shown in Figure 7.6.

All three histograms are well described by a gamma distribution, with the fit parameters given in Figure 7.6. The values obtained for b for the $1\,\mu M$ and $0.8\mu M$ distributions are very similar, 1.12 and 1.20 respectively. The fit of the $1\,\mu M$ distribution yielded a lower estimation for the b parameter, although the error on the parameter was still compatible with values around 1.1. We decided to fit this distribution with a fixed b parameter of 1.12, the same as for cells growing with $0.8\mu M$ g6p. The fit quality of the estimated distribution with and without fixing the b parameter was almost the same.

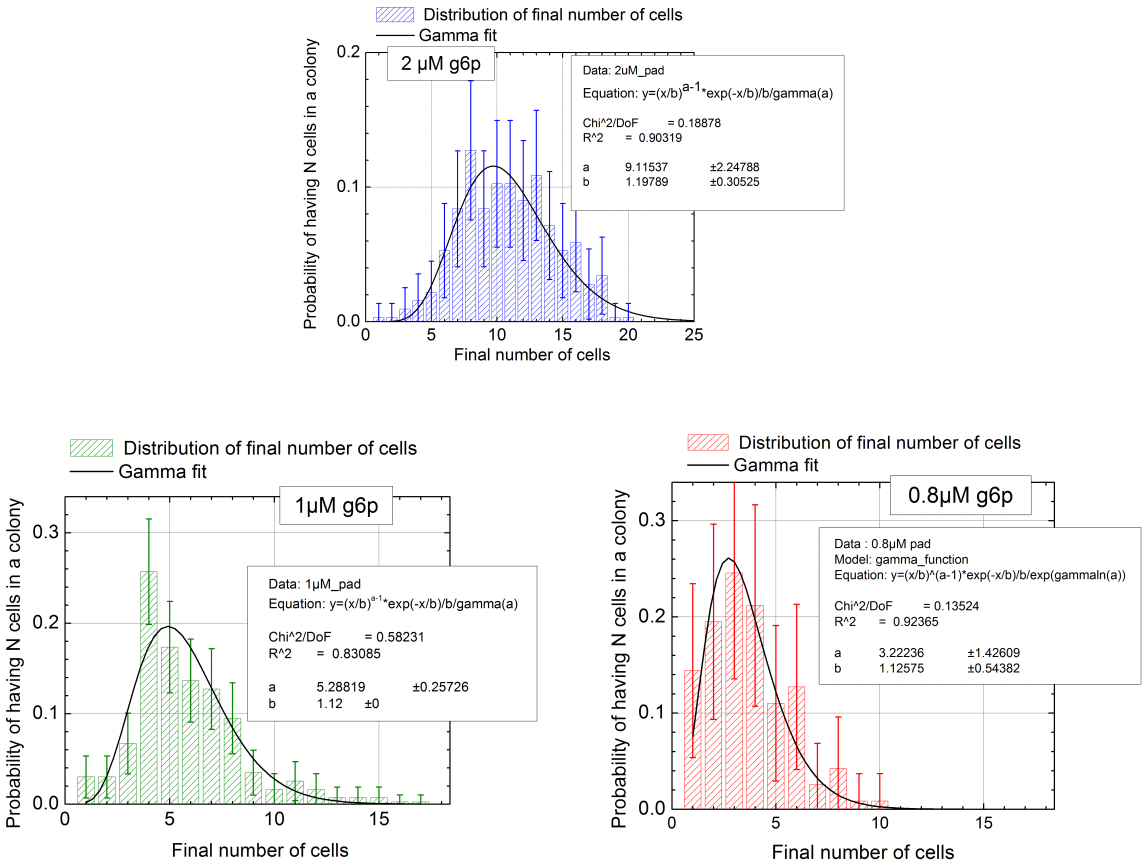

Figure 7.6: Gamma fit of the distribution of the final number of cells. *The fit of the distributions obtained from the histograms from Figure 7.5 by a gammma distribution* $p(x) = \frac{C}{b^a \Gamma(a)} x^{a-1} e^{-x/b}$. *We determined the parameters a and b for all fits*

By our hypotheses, the concentration of DnaA in the cells is a function of the activity of P_{uhpT}, given in equation **6.2**. At steady state of exponential growth, this equation becomes $c = \frac{\alpha \cdot p_{uhpT}(g6p)}{\gamma_{max}+\gamma_{int}}$. Supposing that p_{uhpT} follows a gamma distribution, the spread of the size of DnaA reservoir in the population, and hence the number of divisions after shut-off of P_{uhpT}, also follows this distribution. The parameter a therefore represent the number of mRNA/cell cycle and should be proportional to the activity of $p_{uhpT}(g6p)$. As the induction level increases, the concentration of UhpA∼P will increase too, which will modify the probability of having new mRNA transcribed per cell cycle. This variation is reflected by the parameter a. Comparing our estimation of a, which ranges from 3.2 to 9.1, with similar values in the literature [106], we estimate the number of mRNA transcribed ranges from 3 to 10 mRNA per cell cycle. However for any new mRNA produced everything else in the translation system (*i.e.* numbers of ribosomes, mRNA lifetime) remains constant. Therefore b, should be constant as indeed appears to be, which justifies us fixing the b parameter when fitting the $1\,\mu M$ histogram.

7.3.1 Minimum number of DnaA required for DNA replication

The average number of divisions depends on the average number of DnaA proteins expressed during growth for each histogram. From the data in Figure 6.10 we can estimate an upper

bound on the average number of DnaA proteins expressed by P_{uhpT} in exponential growing cells at different concentrations of g6p. We can then use these values to calculate the maximum number of DnaA proteins in the cells at the time of removal of g6p.

Using the measurements taken from the Hill fit of P_{uhpT} using our *g6p* ladder in Figure 6.10, we estimated the upper bound on the average number of proteins in growing cells at $2\,\mu M$ $1\,\mu M$ and $0.8\,\mu M$ g6p. The respective numbers are 500 proteins per cell, 190 proteins per cell and 130 proteins per cell respectively. The average number of cells an individual pre-grown in the presence of 2, 1 and $0.8\,\mu M$ of g6p divided into are, respectively, 10.9 ± 0.3, 5.6 ± 0.3 and 3.9 ± 0.3. DnaA being a stable protein [107], intrinsic degradation will not significantly decrease the concentrations of the reservoirs during the four or five hours of the experiment. Cells stop dividing when they can no longer replicate their DNA.We can therefore estimate the minimal number of DnaA proteins required to replicate DNA using the following equation:

$$DnaA_{min} = \frac{DnaA_{reservoir}}{Nb\,cells_{final}} \tag{7.2}$$

Here $DnaA_{min}$ is the minimal number of DnaA proteins necessary for DNA replication, $DnaA_{reservoir}$ is the number of DnaA proteins in the cells before stopping the expression of P_{uhpT}, and $Nb\,cells_{final}$ is the number of descendants of the cell without novel synthesis of DnaA.

We plotted $DnaA_{reservoir}$ against $Nb\,cells_{final}$ and fitted linearly the data points forcing the fit to go though zero in Figure 7.7. With only three points it is hard to assess the quality of the fit, however results are satisfactory.

If we look at equation 7.2, our plot should have a slope equal to the value of $DnaA_{min}$, which is equal to 43 ± 4 proteins. However our quantification gives us an upper bound. We can confidently say with that if a cell has 85 ± 7 DnaA proteins, then it will replicate its DNA and divide once more. But we can not predict if cells that have between 43 and 85 DnaA proteins will or will not initiate DNA replication. In order to give a more precise quantification, we would use a DnaA reporter and measure the concentration of DnaA proteins inside cells that stop dividing. That being the case, we chose to give an upper bound of the number of DnaA proteins necessary to replicate the chromosome which is 85 ± 7 DnaA proteins.

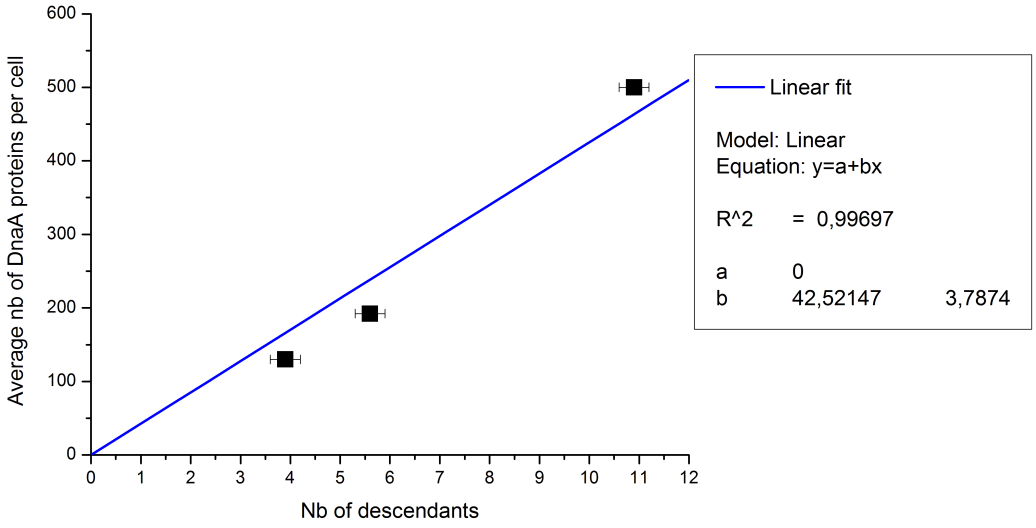

Figure 7.7: Average concentration of DnaA population as a function of the average number of divisions of individual cells. *We plotted equation 7.2 using the upper bound on the average DnaA concentration values we estimated via population analysis and the average number of divisions we calculated in pad analysis.*

Our assumptions did not take into account that DnaA protein interactions inside the cell are also stochastic. Even when DnaA concentration inside the cell is much lower than our estimated value, there is still a probability that DNA replication will be initiated. It is possible, although very improbable, for a very reduced number of DnaA proteins to bound to oriC and initiate replication.

The average number of DnaA proteins per genome in *E. coli* during exponential growth in minimal medium is around 290 proteins per genome [48]. Cells contain about 1.8 chromosomes [102] during growth in minimal medium. These numbers translate to ∼520 DnaA proteins per cell. This is over six times more than our estimate on the minimal number of DnaA proteins needed for cell division. However we only looked at cell division while DnaA is also regulates many genes within the chromosome. It is possible that due to the strength of the DnaA boxes inside oriC and the proximity between them and *dnaA* gene, the proteins will prioritize binding to oriC. We did not look at the different genes regulated by DnaA, it is certain that their expression is different in our cells when they have on average under 200 copies of DnaA. When our cells expressed a similar quantity of DnaA proteins as wild type cells, growing at $2\mu M$, it was exceedingly rare to see a cell stop dividing. On the contrary when cells grew with lower concentrations of DnaA an increasing number number of individuals stopped dividing.

7.3.2 Cells survive for four hours without dividing after DnaA arrest

The distributions we measured describe the initial behavior of cells after DnaA is no longer expressed. However, neither our pad or micro-fluidics experiments can be used to determine if cells remain viable once they stop dividing. They stay metabolically active since they keep elongating but they do not necessarily grow back after P_{uhpT} is activated again. Figure 6.13

shows that some hours after cells no longer divide cell viability drops sharply and abruptly.

The question we asked ourselves is can we estimate the amount of time cells retain their viability after they stop dividing? To extract this information we made a simple model that describes cellular behavior after DnaA expression was shut down. We used the $2\,\mu M$ distribution measured previously in order to establish the initial DnaA concentration in the population, as well as the following assumptions:

- Cells grown in an agar pad and in liquid medium behave the same way.
- Cells divide every 60 minutes.
- The number of divisions the initial cell sample undergoes is determined by a gamma distribution with parameters a = 9.11 and b = 1.20
- All cells are viable after X hours.

The simulation was performed as follows:

1. From the gamma distribution we determined the probability a cell divides a given amount of times.
2. We took a sample of n initial cells which will divide according to the previous distribution.
3. Each hour, the population that hasn't stopped dividing is doubled. All viable cells are counted
4. After X+1 hours, individuals that stopped dividing at X hours are removed from the population.

We will illustrate this simulation with an example. Suppose that a cell had a total of six offspring and we fix the time X cells retain their viability to four hours. After one hour there are 2 cells, after two hours there are 4 cells and after three hours there are 6 cells. To produce a total of six offspring, two of the four cells present after two hours must have divided, thus after two hours only two cells have stopped. After three hours the two cells that have just divided gave two additional offspring, none of them will divide again, thus these four cells stopped dividing after three hours. After a total of six hours, all cells are considered viable, after seven hours the number of viable cells drops to four and after eight hours none of the six cells are considered viable anymore.

In order to compare the results of our simulation with the c.f.u count experiment, we took an initial number of cells which enabled us to have the same number of viable cells after three hours in both the experimental and simulated measurements. We best described the experimental values when X was fixed to four hours. Results are shown in Figure 7.8.

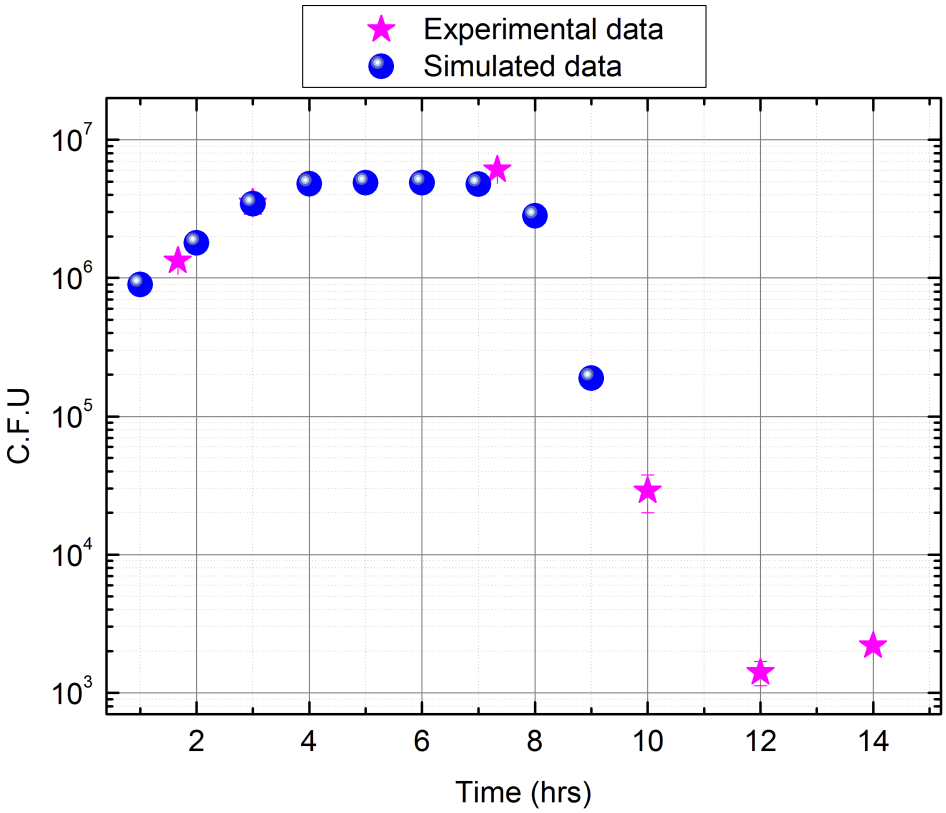

Figure 7.8: Experimental and simulated number of CFU **from cells transferred to a medium lacking g6p at time zero**. *The graph shows the superposition of the number of* CFU *experimentally (magenta stars) and the simulated behavior given the model explained in the main text. The simulation reproduces the observed rapid drop in viability seven hours after the removal of g6p.*

Figure 7.8 shows that the simulation follows very closely the experimental data, including the initial drop in cell viability. It is reasonable to assume that cells maintain their viability for four hours after DnaA induced division arrest. Unfortunately the probability cells divide six times is too weak to be properly computed, thus we were forced to stop at 9 hours in our simulation.

Cell viability drop reaches a plateau after a sufficiently long time. Other mechanisms need to be invoked in order to explain this. The role of this simple model was to extract the amount of time cell remain viable after their last division which we estimated to four hours.

Part IV

Discussion and Conclusion

Chapter 8

Discussion and perspectives

The main focus of this thesis is to study the DnaA protein and the impact of DNA replication inhibition when the numbers of said protein are low inside cells. However, our experimental strategy required us to develop an inducible expression system with a dynamic range that covered the expression levels of DnaA in wild type cells. This led us to the *uhp* expression system that we characterized and used in the construction of the modified *E. coli* strains.

8.1 Construction of strains with external control of DnaA.

A novel inducible promoter

The P_{uhpT} promoter from the *uhp* expression system was converted into an inducible promoter by deleting the *uhpT* gene. Extracellular glucose-6-phosphate is sensed by UhpC proteins, this ultimately leads to the activation of P_{uhpT}. But by deleting the *uhpT* gene we ensure g6p is not transported into the cell and metabolized, thus keeping the activation level of the *uhp* expression system constant. The activity of P_{uhpT} goes up to $10\,MU \cdot min^{-1}$ in MOPS minimal medium [1] which correspond roughly to 1 200 proteins expressed per generation at a 60 minutes doubling time. This is more than twice the number of DnaA proteins of a wild type cell in this growth conditions, which was a reported 290 $proteins \cdot generation^{-1} \cdot genome^{-1}$ [48]. Which, taking into account the partially replicated genome at this growth rate, corresponds to ~ 520 $proteins \cdot generation^{-1}$.

Strains that allow the reversible arrest of DNA replication

We constructed two strains using P_{uhpT}. In the AC *dnaA* was moved and placed in the *uhpT* locus, replacing *uhpT*. In the CL strain , we replaced the promoters of *dnaA* gene by P_{uhpT}. There are quantitative differences in the induction levels of the promoter at similar g6p concentrations in both strains. Even though the promoter sequence in the chromosome is identical between the two. However they both show a similar behavior when we reduce induction levels below a certain threshold: DNA replication and cell division stop abruptly.

At the other end of the concentration scale, a moderate excess of DnaA does not affect the growth of our modified strains. Only at a very strong over-expression of DnaA does the growth rate slow down, which occurs when fully inducing P_{uhpT} in the CL strain. We did not observe

this in the AC strain, where full induction of the promoter resulted in normal growth. The threshold concentration of g6p below which growth stops is lower for the CL strain than for the AC strain. We interpret this difference as a higher global promoter strength of P_{uhpT} in the CL strain, probably due to the specific chromosomal localization of the promoter. Such a position effect has been observed previously in *E. coli* [98, 99].

We exclude the possibility that the differences in the behavior of P_{uhpT} be due to mutations. The sequence of the promoter is identical in both strains. It is unlikely that mutations could after having verified the strains via sequencing. The mutation rate of *E. coli* is estimated to be about 10^{-10} per nucleotide per generation [6] and increased promoter strength would not confer a selective advantage because the strains were always propagated in the presence of high concentrations of g6p.

8.2 Control of DNA replication by DnaA

The physiological consequences of over-expressing are well documented [31, 32]. We therefore focused on the less studied effect of critically low DnaA concentration that lead to DNA replication and growth arrest. DnaA is key in initiating DNA replication and its inhibition stops cells from dividing [14, 15, 16]. Even though most of our experiments measure cell division, we conclude that DNA replication was arrested when cell division has stopped. For certain samples of growth-arrested cells (having spent a long time in the absence of g6p), we have used DAPI-staining of the chromosome to confirm that these cells only contain a single chromosome (data not shown).

Growth rate depends on the rate of production of DnaA in an all-or-none fashion

According to the site-titration model of DnaA functioning, we expected that reducing the DnaA expression would gradually decrease growth rate. The simplest version of the model stipulates that DNA replication is initiated as soon as the number of DnaA molecules per chromosome exceeds a threshold value. By slowing the production rate of DnaA, we expected to simply increase the time necessary for accumulating this critical number of molecules. However we never observed slow growth rates when reducing the concentration of DnaA. Instead we found that decreasing DnaA expression below a threshold abruptly stopped cell division. Our cells respond in an **all-or-none** fashion: either they grow at the maximal growth rate supported by the medium or they stop growing altogether. We observed this phenomenon in both population and single cells. While monitoring cells under the microscope at concentrations of g6p not too far above the threshold concentration, we observed that individual cells that started to lag behind others in exponential growth conditions either definitely stopped their propagation after a couple of extra divisions, or resumed normal exponential growth. These observations suggest that our system behaves as a **bistable system**: cells either grow at a normal rate or they do not grow at all.

Quantification of the intracellular concentration of DnaA

For lack of a direct quantification, we have tried to estimate the absolute concentration of DnaA in the AC and CL strains by indirect methods. Based on the promoter activity and the known conversion of Miller Units to proteins per cell, we were able to set an upper bound on the concentration of DnaA in *E. coli*. The limitation of the "upper bound" stems from the fact that DnaA in our strains is expressed from a weaker ribosome binding site than the *uhpT* gene for which the promoter activity had been calibrated.

The principle of our assay consists in stopping DnaA expression and observing the extent of additional growth. These additional cell division are entirely based on the pool (or reservoir) of existing DnaA proteins. The increase in biomass is therefore directly related to the size of this reservoir of DnaA at the time of removal of g6p. We thus mapped the characteristics of P_{uhpT} to the number of molecules in the cell in order to estimate an upper bound of the number of DnaA proteins in our growth condition (Figure **6.10**).

According to this calibration, for example, cells grown in the presence of 1 and 0.8 μM g6p contain on average less than ~ 190 and ~ 130 molecules of DnaA respectively. Yet, these cells undergo additional cell divisions after removal of g6p before stopping.. We used these numbers to estimate the upper bound on the number of DnaA proteins that are required *in vivo* for initiating a new round of DNA replication. We estimate that the critical concentration of DnaA proteins required for normal cell division is less than 85 ± 7 proteins per cell. This number is more than six times lower than the average number of DnaA proteins in wild type cells. There are over 300 DnaA boxes on the chromosome [35] and many genes are regulated by DnaA. It is therefore possible that the large number of DnaA proteins present in wild type cells are needed for the proper regulation of gene expression, rather than for initiating DNA replication.

Variability of the expression of DnaA

Our single-cell analyses showed that in a growing population containing on average fewer than 190 DnaA proteins per cell, a little under 2% of the population permanently stopped division at each generation. We interpret this permanent growth arrest as a consequence of the variability of P_{uhpT}, combined with the all-or-none response of growth to the concentration of DnaA. In other words, in a population of cells containing on average 190 molecules of DnaA per cell, about 2% of the population only contains less than 85 DnaA proteins (i.e., a number below the threshold for growth). This interpretation agrees with the observed gamma distribution of the promoter activity of P_{uhpT} (Figure 7.6).

It is possible that, in order to avoid that a portion of the population would end up with too few DnaA proteins, the cell expresses a significantly higher number than what would be required for initiating DNA replication. However, contrary to our artificial constructs, the natural expression of DnaA is highly regulated in wild type cells. In particular, DnaA negatively and positively auto-regulates its promoters [25, 57]. which helps reduce transcription noise [108]. The variability of DnaA expression inwild type bacteria is therefore most likely lower than what we observe for P_{uhpT}. In this case, cells would not require a large average DnaA concentration to avoid that a significant portion of the population drops below the threshold for initiating DNA replication.

8.3 Physiology at sub-limiting concentrations of DnaA

Cells keep growing in the absence of DNA replication

Single cell observations on cells that no longer express DnaA showed that, after they stop dividing and their DnaA reservoir goes below the critical DnaA concentration, they continue to elongate. We observed the corresponding signals in both population and single cell experiments. Optical density (OD) curves of cells growing without expressing DnaA start by increasing exponentially, exhausting their DnaA reservoir, then OD signal continue to increase at a slower pace. This slow apparent growth is the result of cells unable to divide but continuing to increase their biomass. We observed this phenomenon visually in micro-fluidics experiments as well as in agar pads, where cells elongated many hours after their final division.

Large variability of filament length after growth arrest

There is a large variability in the maximum size of the filaments formed by cells that increase in size when they stop replicating their DNA. Some cells barely grow after they stop dividing, their size similar to a cell during exponential growth. The majority of the cells elongate to more than 10 times the size of an exponential growing cell. Their exact size is difficult to estimate since they span several fields of view under the microscope or get dragged into the central channel in the microfluidics device before they stop elongating. Nonetheless, we found some individuals that reached sizes 50-fold longer than exponentially growing cells.

Non-dividing cells remain metabolically active

The optical density of cell populations growing without g6p increases for about 10 hours, then stabilizes or starts to decrease (Figures 6.3 and 6.4). However, green fluorescence, indicative of the production of flavins by the cells, keeps increasing for at least 16 hours after removal of g6p (Figure 6.8). While observing individual cells through the microscope we identified a small number of elongated cells that continued their growth after 20 hours without expressing DnaA. Both filament elongation and the secretion of flavins are a consequence of metabolic reactions. Our observations prove that cells that are no longer viable remain metabolically active. Even though flavins are secreted several hours after the majority of cells stop elongating, we can not ascertain that all cells of a population remain metabolically active. We do not know what proportion of the cell population continues to have an active metabolism.

Prolonged arrest of the production of DnaA causes a loss of viability

Even though the bacteria produce metabolic signals for a long time after growth arrest, they have already lost their viability at this point. The number of cells capable of forming colonies drops sharply about four hours after the cells stop dividing (Figure 6.13). Coupling these measurements with a simple model and the observed distribution of DnaA within the population, we estimate that the majority of cells remain viable for only$\sim$ 4 hours after their last division. However, almost all cells at this stage continue to elongate. The viability of the cells is therefore not directly linked to metabolic activity.

Roles of DnaN and RecF in the loss of viability

We have tried to determine the mechanisms responsible for the loss of cell viability. Possible candidates are the two other genes of the *dnaA* operon, *recF* and *dnaN*. Previous experiments have shown that the over-expression of *recF* and *dnaN* leads to filamentation and cell mortality [30]. The AC construction left the *dnaA* promoter untouched and dnaN and recF retain their original regulation. DnaA represses transcription from its own promoter, thus a decrease in the concentration of DnaA will lead to an over-expression of DnaN and RecF in the AC-strain.

However, in the CL strain, all three genes of the *dnaA* operon are transcribed by P_{uhpT}. We therefore compared cell viability of AC and CL strains with similar DnaA concentrations within the respective populations after removal of g6p. Our results show that the lag before re-growth after prolonged arrest of the production of DnaA is comparable in both strains. Although the lag is slightly shorter in the CL cell population, . the improvement in viability is small. It remains unclear whether cell viability is improved in the CL strain, or whether a different distribution of DnaA proteins in the population is responsible for the difference in lag-times. We must stress that P_{uhpT} does not behave identically in the two strains and this could slightly change the cellular DnaA distribution in the CL strain, compared to the AC strain. In conclusion, cell viability still drops considerably in the CL strain even though the expression of *dnaN* and *recF* is reduced compared to the AC strain. Their transcription comes solely from their own promoters.. Thus DnaN and RecF the main reason for cell elongation or the loss of viability.

Other causes of cell mortality

Grigorian et al. [30] report that over-expressing DnaA, or simultaneous over-expression of DnaN and RecF, hinders DNA replication, thus impeding cell division. The report concludes that DnaA, DnaN or RecF somehow affect DNA replication and create double strand breaks, leading to loss on cell viability. In the AC and CL strains we do not over-express DnaA, but we stop DNA replication altogether. This replication arrest could possibly lead to DNA lesions which would then trigger an SOS response [109].But in the work of Grigorian [30] (strong over-expression of DnaA), the loss of viability does not trigger an SOS response and we suspect that is also the case with our strains.

We also measured the cellular internal pH after arresting the production of DnaA using the pFGPR01 plasmid [93] (see table 4.1) hoping it could reveal a misfunction within the cells. Ultimately we did not observe any significant difference in the intracellular pH of cells not expressing DnaA and wild type cells. Another possible cause of death would be oxidative stress. To test this hypothesis, we grew AC8 in media containing thiourea or gluthatione, powerful hydoxyl radical scavenger that are used to improve cell viability during oxidative stress (data not shown). We did not see any improvement in cell viability. We can not definitively exclude oxydative stress as a cause of mortality, but all our circumstantial evidence argues against this explanation.

We believe that the loss of viability and cell elongation is a consequence of the inability to replicate DNA. This default could also imply some form of damage to the chromosome and an increased mutation rate. In any case, the decrease in cell viability is not a direct consequence of an excess of DnaN or RecF, but these proteins could enhance the loss of viability.

A small proportion of cells remains "indefinitely" viable

Cell viability does not drop indefinitely as cells spend a long time without expressing DnaA. All experiments measuring CFU as a function of time in different pre-growth conditions show that cell viability stops decreasing about 12 hours after stopping DnaA production. We observe the equivalent phenomenon in population experiments. The apparent lag before re-growth increases with the time cells spent in the absence of production of DnaA. However, this lag reaches a plateau for cells exposed to the stress condition for more than 14 to 16 hours (Insert of Figure 6.12). The apparent lag never exceeds nine hours. Such a lag could be explained by assuming that $\sim$0.1% of the population remains viable and grows at nominal rate with a doubling time of 60 min ($2^{10} \approx 1\,000$). The direct count of CFU yields a similar value. In these experiments, cell viability never drops below 0.025% of the initial population (see Figure 6.13). This percentage is smaller than 0.1%. However the difference could be compensated if during re-growth, elongated cells divide into several small independent cells that would resume normal growth. This behavior was observed during re-growth experiments on cells that stopped dividing when the concentration of RNA polymerase was too low [19].

Going back to the issue of cell viability, apparently some cells of the population do not lose their viability, regardless of the time they spent without expressing DnaA. We took some cells from a cell population that grew back after more than 14 hours of not expressing DnaA and examined their growth characteristics. By repeating the growth arrest experiment (removal of g6p after an initial growth in the presence of g6p, data not shown) with these cells, we obtained an OD profile similar to the one of the original AC8-strain (Figure 6.7). We can therefore conclude that the survival of the small portion of the population is not due to a mutation of P_{uhpT} in these cells.

Potential applications in biotechnology

One of the motivations of this project was to test whether stopping DNA replication could improve the yield of an industrially interesting metabolite. The proof of principle of this idea has been reported by Izard *et al.* [19]. They succeeded in improving the yield of glycerol production in *E. coli* after stopping growth. Their strategy was to arrest transcription to stop growth. We have shown that metabolic activity persists after stopping growth by arresting DNA replication. The growth arrested AC8 cells may well constitute the sought for "bag of enzymes". We will test the glycerol yield of AC8 in different growth conditions using the same plasmid that provides the enzymes for glycerol production, pGLYC (see table 4.1). By adjusting the process conditions, for example, alternating phases of DnaA production with phases of removal of g6p, we may be able to circumvent the problem of loss of viability.

8.4 The control system, uhp, produces a wide distribution of promoter activities.

In order to construct the strains comprising a genetic control of the production of DnaA, we had to characterize a new, inducible promoter: P_{uhpT}. This promoter could be useful in many other

situations. P_{uhpT} is regulated by activation, has low basal activity, and its inductor molecule, g6p, is not used by any other inducible promoters. We characterized the system and started to build a simple, predictive model in order to better understand and better exploit P_{uhpT}. The model is not yet fully operational. We are currently making some minor modifications and adjusting parameters. We therefore only present an outline of the current state of the model.

Signal transduction in the *uhp* expression system starts with UhpC, a trans-membrane protein that senses glucose-6-phosphate (g6p). Once this molecule is detected in the growth medium, UhpC undergoes a conformational change and allows the auto-phosphoryaltion of UhpB. As in any classical two-component system, UhpB transfers the phosphate to UhpA, thereby increasing its affinity for P_{uhpT} and activating this promoter. In addition to acting as a phosphate donor for UhpA, UhpB also possesses a phosphatase activity, removing the phosphate of UhpA∼P. Thus, the *uhp* expression system can be described by three components:

- The interaction of P_{uhpT} with the phosphorylated and non phosphorylated UhpA
- The Interaction between UhpA and UhpB in their respective phosphorylated and non phosphorylated forms
- The autophosphorylation of UhpB in presence of extracellular g6p.

This model description leads to a system of coupled differential equations (not shown). We are currently calibrating the parameters of the model extracted from data found in the literature [1, 84]. Using our current set of parameters and equations in steady state conditions, we can match the published characteristics of P_{uhpT}. Furthermore, the model makes prediction about the proportion of the phosphorylated forms of UhpA and UhpB, with respect to their total concentration. The results are shown in Figure 8.1.

Our preliminary simulation makes a surprising prediction: less than 1/10 of the total number of UhpA proteins are phosphorylated under any conditions. The total number of UhpA proteins in cells growing in MOPS minimal medium is ∼150 proteins per genome [48], which corresponds to ∼270 proteins per cell. Therefore, the prediction suggests that there are never more than 30 UhpA∼P proteins in *E. coli*. This number is surprisingly low. However, such a limiting amount of UhpA∼P may explain a curious experimental result. Cloning P_{uhpT} on a multi-copy plasmid only resulted in a 4.6-fold increase of activity compared to the same promoter in single-copy on the chromosome [110], suggesting that UhpA∼P was a limiting factor. Although we are still working on the model, the low percentage of phosphotylated UhpAis a stable prediction of the model.

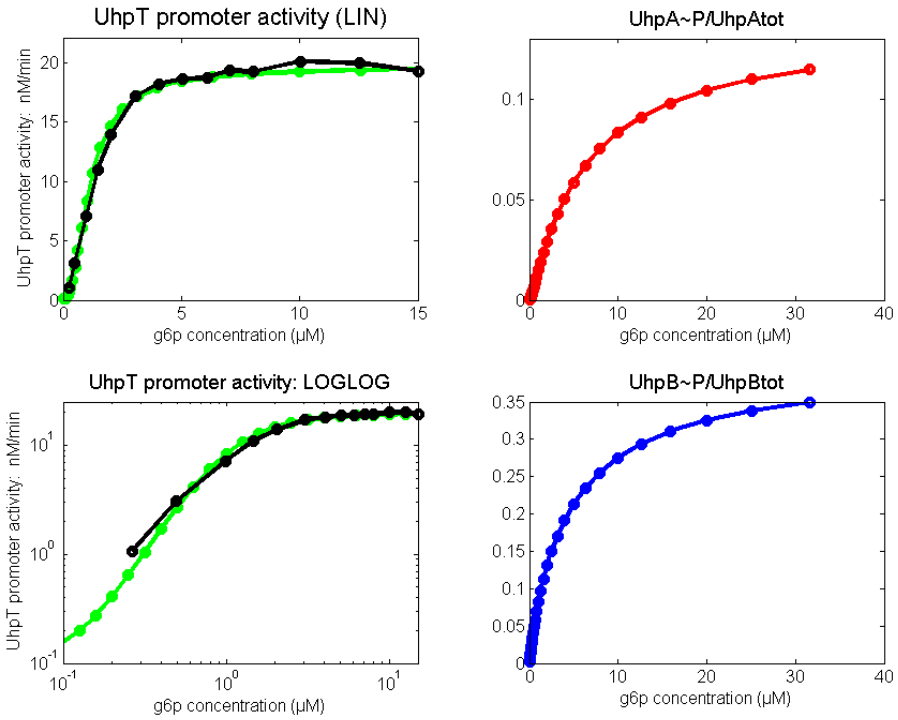

Figure 8.1: Preliminary simulation of the *uhp* expression system. *We plotted the simulated values of the activity of P_{uhpT} (green) against the experimental results from Verhamme (black). The plots are on a linear scale (top left) and on a loglog scale (bottom left). The simulated ratio of phosphorylated v/s total number of UhpA (top right) and UhpB (bottom right) proteins is shown for different concentrations of g6p.*

Another interesting feature of the uhp system is the particular distribution of the activity of P_{uhpT}. We estimated the parameters of the gamma distributions that best describe the number of offspring after stopping the production of DnaA. If the expression from P_{uhpT} stops rapidly after removal of g6p, these distributions reflect the distribution of DnaA proteins in exponential growth conditions. The gamma distribution can be used to extract mechanistic properties of promoters [111, 106]. The two parameters of the gamma distribution, a and b, correspond to the number of mRNA/cell cycle and to the number of proteins/mRNA lifetime.

We have found that during exponential growth with different concentrations of g6p b remains constant, while only a varies. As the g6p concentration increases, the concentration of UhpA∼P will increase too, which will modify the probability of having new mRNA transcribed during a cell cycle. This variation is reflected by the parameter a, whose value indicates that we transcribe 3 to 10 mRNA per cell cycle [106]. Increasing transcription does not change any component of the translation mechanism (*i.e.* numbers of ribosomes, mRNA lifetime). Therefore b should be constant, as indeed appears to be.

Of course our interpretation of the gamma distribution contains some rough assumptions, as the fact that the main stochastic event we considered is the production of new mRNA of DnaA. Allover we have many other sources of variability including the decision of stopping division when the DnaA threshold is reached. We considered for all 3 concentration of g6p studied in microscopy experiments that this does not dependent on g6p. This was also the basis of our

"reservoir model" in population experiments.

8.5 Conclusion and open questions

We have advanced our understanding of the mechanisms of the initiation of DNA replication by DnaA and we have constructed and characterized a new inducible promoter. Many open question remain for both topics.

A model of P_{uhpT}

Our current model of the *uhp* system correctly predicts the behavior of P_{uhpT}, and the prediction of the low number of UhpA∼P proteins is compatible with experimental observations. However, we are still adjusting the parameters in order to simulate the dynamics of the uhp system. A full model will allow us to better quantify the number of proteins involved in the activation of P_{uhpT} and could also help us simulate the shape of the distribution of DnaA in growing cells. The dynamics of the system would also indicate if DnaA expression stops rapidly after we remove g6p from the medium, which is one of our hypotheses. Our estimation of the minimal number of DnaA proteins required for DNA replication would change significantly if the expression of DnaA continued after removal of g6p.

Critical concentration of DnaA

Currently, we estimate that cells with at least 85 ± 7 DnaA proteins will replicate their DNA and divide. But we can not give an exact quantification, since we do not exactly know the translation efficiency in our strains. Therefore, we have planned to quantify the intracellular concentration of DnaA by an absolute mesurement using proteomics. With this information, we will be able to give the exact average DnaA concentrations in our cells, thus giving an exact number of DnaA proteins required for DNA replication.

Cell viability

We have shown that the viability in MOPS drops when cells are unable to replicate their DnaA for extended periods of time. Using a very simple model, we estimated that cells do not survive beyond four hours after they stop dividing. We have also hypothesized that accumulated damage to the cell is responsible for viability loss. However, our model does not take into account this possibility. We have planned to introduce additional parameters in order to model the accumulation of damage. Such an addition to our existing model will make it more likely for cells to loose viability the longer they remain unable to divide. In order to fine tune all new parameters, we also need to increase the number of experimental points we have at our disposal. Thus, a new Petri dish regrowth experiment is planned, with additional data points between four and nine hours of stress.

We also wish to test if cells that can not replicate their DNA suffer from accumulated damage. We would need to grow the cells and stop their DNA replication. In a second step, we would make them re-grow for two to three generations, and finally measure whether the duration during which they remain viable is shortened further.

Physiology of long-term survivors

The physiological state of cells that grow back after resuming DnaA expression is still not clear to us. We do not know if greatly elongated cells that do not stop growing after having stopped the activity of P_{uhpT} for many hours eventually divide and give new offspring, or if the long-term survivors are derived from the small cells that do not elongate at all. It would also be interesting to know if elongated cells can divide again, if they simultaneously divide into many small cells after having replicated many chromosomes, or if two genomes suffice for a first division.

A more in depth study of cells that grow back after very long periods without DnaA expression would give information not only about the functional roles of DnaA, but also on P_{uhpT}. We have discarded the possibility that these cells have a mutated P_{uhpT}. However, we can not say the same for the remaining three genes of the *uhp* expression system. It remains possible that mutations in other genes enable the cell to find other methods to initiate DNA replication.

Applications in biotechnology

The potential application of our construction in the industry is strongly limited by the loss of viability four hours after cell division stops. However, this can potentially be countered by an increased metabolic activity. It has been shown that by lowering the number of RNA polymerases in cells, additional resources are allocated to other cell mechanisms, including the production of molecules by enzymes [19]. Cells that do not replicate their DNA present a similar scenario, where the resources and energy required for DNA replication is free to be allocated elsewhere. If the efficiency of cells not expressing DnaA were much higher than wild type cells for producing molecules, this could compensate for the loss of viability. In order to verify this scenario, we will introduce the plasmid allowing *E. coli* to produce glycerol, pGLY (see Table 4.1), into our strains and compare the production of glycerol of our strains with wild type cells.

There are other systems that control cell division or DNA replication. Reducing the number of RNA polymerase stops cell division, although it does not stop DNA replication [19]. Recently, a publication showed that the CRISPR-Cas system could be used to block *oriC* and stop DNA replication [112] in a similar way as what we have done with DnaA. However, their genetic system is not easily reversible. Furthermore, neither this publication, nor the system controlling the expression of RNA polymerase have investigated the crucial property of cell viability for long periods of time.

French summaries for each chapter

La cellule

La deuxième partie du vingtième siècle a vu apparaître l'application de techniques typiquement utilisées en Physique pour l'étude de systèmes biologiques. Ceci a révolutionné le domaine de la biologie quantitative, par exemple avec l'utilisation de systèmes micro-fluidiques et la microscopie quantitative. Cette thèse vise à élucider certains aspects quantitatifs des mécanismes de l'initiation de la réplication du chromosome chez la bactérie *Escherichia coli* (*E. coli*).

Dans ce chapitre, nous décrivons la bactérie *E. coli* et nous détaillons les avantages de réaliser une étude en observant des bactéries uniques plutôt que des études en population.

E.coli est une bactérie Gram négative, possédant deux membranes qui entourent le cytoplasme où résident son ADN et la majorité des protéines. Elle possède un seul chromosome circulaire de 4,6 millions de paires de bases. Elle a un temps de division qui varie de 20 minutes à plusieurs heures suivant le milieu de culture. Il s'agit d'une des bactéries le plus étudiées. Les amples connaissances concernant cette bactérie, ainsi que les nombreux outils disponibles pour l'ingénierie génétique font de cet organisme le modèle idéal pour une étude biophysique quantitative.

L'expression des protéines se fait à partir de l'ADN. L'ARN polymérase fait une copie des gènes à exprimer dans un processus appelé transcription. L'ARN messager est à son tour lu par des ribosomes et traduit en protéines. Une modification de la concentration d'ARN polymérase affecte l'expression de toutes les protéines de la cellule. Or, l'ARN polymérase elle est elle-même un assemblage de protéines. Il en est de même pour les ribosomes. Ces boucles de rétro-actions sont typiques pour des systèmes vivants et rendent l'étude quantitative à la fois intéressante et complexe. De plus, l'expression génique est un système stochastique. Au niveau d'une population cellulaire, on observe uniquement des réponses moyennes, tandis que l'observation de cellules uniques révèle une grande diversité phénotypique au sein d'une population isogénique de bactéries.

L'expression génique est affectée par d'autres protéines ou ARN, appelés facteurs de transcription. Le nombre de copies de gènes par cellule modifie également l'expression génique. De nombreux facteurs contribuent donc à la variabilité entre cellules et seulement l'observation de cellules individuelles peut rendre compte de la complexité de la régulation des processus cellulaires.

La transcription d'un gène commence par la liaison de l'ARN polymérase au promoteur, un site de liaison spécifique en amont de la séquence codante. Chez les bactéries, un transcrit comprend généralement plusieurs gènes, appelé opéron. Un promoteur peut être actif de façon

constitutive ou être régulé par des facteurs de transcription : des activateurs ou des répresseurs. Souvent, les facteurs de transcription des promoteurs inductibles sont sensibles à des molécules dans le milieu ce qui nous permet de les contrôler. Juste en amont de la séquence codante se trouve le site de liaison des ribosomes (RBS). L'accès du ribosome à ce site est également régulé et la séquence du site (et d'éventuels régulateurs) détermine le nombre de protéines faites par transcrit.

Le code génétique est lu par triplets de bases que l'on appelle codons et qui codent pour les vingt acides aminés. Puisqu'il y a quatre bases différentes sur trois positions, on arrive à 64 codons différents. Le code génétique est donc redondant. Trois des codons sont appelés codons stop. Ils signalent l'arrêt de la traduction.

L'information génétique peut être altérée par des mutations. Par exemple, 20 des 64 codons peuvent être convertis en codon stop par une seule mutation. Ceci est important pour les promoteurs inductibles. La régulation par l'activation requiert la liaison de l'activateur pour la transcription, tandis que la régulation par répression est basée sur la dissociation d'un répresseur en réponse à un signal. En conséquence, la probabilité qu'une mutation aléatoire rende un gène régulé constitutif est bien plus élevé pour une régulation par répression. Il est beaucoup plus facile de rendre un répresseur inactif, par exemple par un codon stop, que de rendre un activateur constitutif, par exemple par une mutation qui mime la présence d'un inducteur.

Il est important de bien caractériser un promoteur pour une étude quantitative. Expérimentalement, les méthodes les plus utilisées consistent à utiliser des molécules fluorescentes ou de mesurer l'activité enzymatique d'un gène rapporteur (Miller Units pour la bêta-galactosidase). Cette dernière méthode a été calibrée et une unité Miller peut être convertie à environ deux protéines produites par minute.

La plupart des mesures et expériences quantitatives nécessitent la modification du génome des bactéries. Par exemple, la modification d'un réseau ou la construction d'un gène rapporteur font tous appel à ces méthodes de la génétique moléculaire. Chez *E. coli*, ils existent de nombreux outils pour efficacement changer le génome, essentiellement basé sur la recombinaison homologue et des systèmes de sélection positive et négative.

Il ne suffit pas de modifier les bactéries pour les observer, une étude quantitative nécessite un bon nombre d'individus et des observations prolongées. Des systèmes de micro-fluidique ont été développés dans lesquels les bactéries sont piégées dans des micro-canaux. Ces canaux sont alimentés par un flux de milieu qui permet aux bactéries de se diviser, mais impose une localisation précise dans les canaux. Cette méthode est un très bon complément de l'observation par microscopie en temps réel, qui consiste à prendre des images des bactéries à des intervalles fixes pour suivre l'évolution des cellules individuelles. Si les bactéries sont piégées dans les canaux de micro-fluidique, les bactéries restent au point focal et dans la fenêtre d'observation tout le long de l'expérience.

Dans cette thèse, nous avons étudié les mécanismes moléculaires de la réplication de l'ADN. Pour initier la réplication, il faut qu'une protéine appelée DnaA se fixe sur l'origine de réplication du chromosome. Une fois assemblé à l'origine de réplication, cette protéine recrute les autres acteurs de la réplication, tel l'ADN polymérase. Nous avons étudié de manière quantitative cette première étape clé du processus de réplication. Nos résultats apportent des informations

nouvelles et ouvrent la voie pour une application du contrôle de la réplication en biotechnologie.

Controller la réplication de l'ADN

Un but majeur de la thèse consiste à construire un système de régulation permettant d'arrêter et de re-démarrer la réplication de l'ADN par un signal externe à la cellule. L'arrêt de la réplication de l'ADN peut se faire de deux façons différentes. On peut utiliser des substances qui vont inhiber l'action de protéines nécessaires pour la réplication ou on peut modifier le génome. La première option n'affecte pas seulement la machinerie de réplication de l'ADN, mais tout le système cellulaire sera perturbé. En particulier, des bactéries traitées par ces molécules déclenchent une réaction de stress, appelée réponse SOS. La deuxième option est plus ciblée. On peut modifier un gène spécifique et minimiser l'impact sur les autres fonctions cellulaires.

Modifier le génome comporte deux choix : il faut définir le gène à modifier et le promoteur à utiliser pour l'exprimer. Dans ce chapitre, nous discutons nos choix et aboutissons aux buts et objectifs précis de cette thèse. Nous détaillons également notre stratégie expérimentale.

Nous étudions l'initiation de la réplication de l'ADN. Il y a essentiellement trois protéines qui sont impliquées pour le démarrage de ce processus, DnaA, DnaB et DnaN. Nous avons choisi de nous concentrer sur DnaA, car cette protéine est essentiel à la réplication et n'a pas été identifié comme un acteur important dans d'autres processus cellulaires.

Le gène *dnaA* est le premier gène d'un opéron qui comporte également les gènes *dnaN* et *recF*. L'expression de *dnaA* est sous le contrôle de deux promoteurs qui transcrivent tout l'opéron. DnaN est un gène essentiel, tout comme *dnaA*. DnaN est l'hélicase qui se fixe au chromosome lors de la réplication de l'ADN. Elle joue aussi un rôle dans la réparation du chromosome. Il y a quatre promoteurs de *dnaN* dans la dernière partie codante de *dnaA*. Le dérnier gène, *recF*, n'est pas essentiel.

Pour initier la réplication, DnaA se fixe à des séquences spécifiques du chromosome, appelées ''DnaA boxes". Il y en a plus de 300 de ces sites dans tout le chromosome, y compris sur l'origine de réplication. La liaison de DnaA aux sites de fixation à l'origine de réplication est la première étape de l'assemblage d'un complexe protéique qui aboutira à l'initiation de la réplication. DnaA peut se lier aux nucléotides adénosine-triphosphate (ATP) pour former la forme active de la protéine. Cette ATP peut être hydrolysée pour former de l'adénosine-diphosphate (ADP). La forme DnaA-ADP est la variante inactive de DnaA. La forme active possède une meilleure affinité pour ses sites de liaison spécifique. Pour se lier aux séquences sur oriC, il est estimé que 20 protéines de DnaA sont suffisantes. Or, en fonction du milieu de croissance, la bactérie possède entre 290 et 2000 molécules de DnaA par cellule.

L'expression de DnaA est très régulée chez *E. coli*. Il existe plusieurs mécanismes qui séquestrent ou inactivent DnaA afin d'empêcher une nouvelle réplication du chromosome de façon prématurée. Parmi ces mécanismes se trouvent : la séquestration de l'origine de réplication par SeqA, l'inactivation de DnaA par RIDA, le titrage de DnaA par le locus datA du chromosome et l'auto-régulation négative de DnaA. D'autres mécanismes, comme la réactivation de DnaA par le système DARS favorisent la conversion de DnaA dans sa forme active. En plus de son rôle pour l'initiation de la réplication, DnaA joue un rôle d'activateur ou de répresseur de

plusieurs autres gènes du chromosome. Parmi eux sont : nrdA/B, guaB, rpoH, mioC, uvrB et proS. Dans aucun de ces gènes, l'absence de DnaA provoquerait un effet létal dans la bactérie.

Nous avons choisi de construire un système génétique d'arrêt de la réplication de l'ADN en mettant la transcription de *dnaA* sous le contrôle d'un promoteur inductible. Afin d'éviter des problèmes avec la stabilité à long terme de la construction, nous devons utiliser un activateur. Aucun système habituel ne possède toutes les caractéristiques nécessaires. Nous avons donc décidé d'adapter le système de transduction du signal *uhp* pour nos besoins.

Le système *uhp* comprend un ensemble de gènes permettant l'import de sucres phosphorylés dans la bactérie. Le transporteur est codé par le gène *uhpT*. Ce dernier est exprimé seulement en présence de glucose-phosphate (g6p) dans le milieu. L'intensité du promoteur du gène *uhpT* est dépendante de la concentration de g6p. La délétion de ce gène transforme le promoteur d'*uhpT* en un promoteur inductible par g6p. La déletion d'*uhpT* assure que g6p n'entre pas dans la cellule et n'est donc pas métabolisé.

Le système *uhp* comporte en tout quatre gènes, *uhpA*, *uhpB*, *uhpC* et *uhpT*. UhpC et UhpB sont des protéines membranaires qui détectent la présence de l'inducteur, g6p, dans le milieu. Suite à la liaison de g6p, un changement de conformation d'UhpB active sa fonction kinase, et déclenche la phosphorylation d'UhpA. La forme phosphorylé d'UhpA se lie au promoteur d'*uhpT* et active la transcription.

UhpA est constamment dé-phosphorylée par UhpB et donc, en absence de g6p, P_{uhpT} est éteint. Les caractéristiques de ce promoteur font de lui un promoteur inductible qui répond à nos exigences. Par exemple, l'activité du promoteur est comparable à celle du promoteur de *dnaA*. De plus, sa position dans le chromosome est très proche de DnaA, ce qui facilite les interprétations. P_{uhpT} a déjà été étudiée auparavant. Son expression suit une loi de Hill avec des paramètres connus. Pour nos modèles, il est important de mettre en équation les phénomènes observés.

Ceci nous amène aux motivations de notre étude. La réplication de l'ADN et la caractérisation de DnaA sont des sujets qui ont été traités auparavant chez *E. coli*. Cependant, une étude quantitative en modulant à la baisse la concentration de DnaA n'a jamais été faite. Nos objectifs sont donc (i) de construire une souche d'*E. coli* dans laquelle nous pouvons ajuster la concentration de DnaA par un stimulus externe. (ii) A l'aide de cette souche, nous caractériserons les effets sur la réplication de l'ADN d'une baisse de la concentration de DnaA. Nous établirons, entre autres, la relation quantitative entre la concentration de DnaA et le taux de croissance des bactéries. (iii) En dessous d'une concentration seuil de DnaA, les bactéries arrêtent de se diviser. Nous déterminerons la réversibilité de l'arrêt de croissance et la viabilité des cellules arrêtées. (iv) Finalement, afin de réussir ce projet nous construirons et caractériserons un nouveau système de contrôle de la transcription, le système *uhp*. Les résultats attendus se situent dans la recherche fondamentale (compréhension des mécanismes moléculaires de l'initiation de la réplication), mais peuvent également mener à des applications en biotechnologie. Une souche dont on peut contrôler la division cellulaire peut être utilisée pour diriger les ressources cellulaires vers la production d'un métabolite d'intérêt industriel.

Les outils techniques mis en œuvre pour atteindre ces objectifs sont des techniques de la génétique moléculaire, la mesure en population et en cellules uniques de la croissance et de l'expression génique, ainsi que la modélisation mathématiques des phénomènes observés. L'observation en

cellules unique repose sur la construction d'un dispositif de microfluidique ou l'analyse d'images de micro-colonies se développant sur un milieu solide sous le microscope. Cette analyse en cellules uniques révélera les aspects stochastiques des processus étudiés.

Montages et techniques expérimentales

Ce chapitre est consacré à la description des différents outils expérimentaux que nous allons utiliser dans nos expériences. Dans cette description, nous abordons également des méthodes associées, qu'elles soient associées à la physique ou la biologie.

Un des points centraux de cette thèse est l'étude de cellules uniques. L'observation de cellules individuelles en temps réel nécessite la microscopie à lapse de temps. Nous combinons cette technique avec la micro-fluidique afin de faciliter le traitement d'images. En effet, l'utilisation d'un dispositif de micro-fluidique permet d'observer plus facilement les cellules puisqu'on les force à rester dans un plan particulier qui est confondu avec le point focal du microscope.

Le dispositif de micro-fluidique que nous avons utilisé s'appelle ''mother machine". Il est constitué d'un canal central duquel sortent 4000 canaux secondaires d'une taille permettant aux cellules d'y pénétrer sans pouvoir s'arranger côte à côte. Du milieu de culture passe par la ''mother machine" à partir d'une bouteille. Le flux permets à la fois de piéger les bactéries à l'intérieur des canaux et de leur donner des nutriments pour qu'ils puissent pousser. Les bactéries au bout des canaux secondaires sont expulsées dans le canal principal et emportées par le flux lors de la croissance. La ''mother machine" est fabriqué en PolyDiMethylSiloxane (PDMS). Nous décrivons brièvement le protocole de fabrication.

Une bouteille est connecté à la ''mother machine" qui apporte un flux constant de milieu de culture grâce à une différence de pression exercée par une pompe. Le flux sortant passe par un analyseur, qui mesure le débit. Le tout est contrôlé par ordinateur. Tout ce qui est en contact avec les cellules est passé par des filtres ou/et a été stérilisée.

Pour permettre aux bactéries de pousser normalement, notre microscope est maintenu à 37°, tout comme la bouteille de milieu. Le microscope et une caméra sont reliés à un ordinateur, ce qui nous permet de lancer des longues acquisitions (quelques heures à plusieurs jours) sans perturber les cellules.

Notre montage expérimental nous permet de faire de la microscopie à lapse de temps avec d'autres supports que la ''mother machine". Nous utilisons aussi des pads d'agar, qui sont des versions gélifiées des milieux de culture. L'avantage de ce support par rapport au précédent est que les cellules ne s'échappent pas. Toutes les cellules inoculées sur le support solide y restent pendant toute la durée de l'expérience. Les détails de fabrications sont décrits dans ce chapitre.

Il est important de bien contrôler le milieu de culture pour que les expériences soient répétables. Nous décrivons les trois milieux que nous utilisons dans cette thèse et donnons leur recette. Le premier est un milieu riche, le LB. Ce milieu est utilisé surtout pour faire pousser rapidement les cellules. Les deux autres sont des milieux minimum, c'est-à-dire qu'ils contiennent le stricte minimum pour permettre aux bactéries de se développer. Nous utilisons les milieux MOPS et M9. En particulier, nous contrôlons la seule source de carbone de ces milieux en les complétant avec 0.2% de glucose.

Nous réalisons aussi des études en population à l'aide d'un lecteur de microplaques. Ce lecteur nous permet de mesurer 96 puits contenant des cellules, ce qui est équivalent de 96 expériences simultanées. De plus, en répliquant les mêmes conditions à plusieurs puits, on peut augmenter le degré de confiance d'un résultat. Les bactéries poussent dans la plaque à la bonne température pendant que leur absorbance et fluorescence sont mesurées à des intervalles réguliers.

Comparer deux mesures de microplaques nécessite souvent un post-traitement. Des erreurs expérimentales, telles des concentrations initiales différentes, peuvent introduire une erreur supplémentaire aux courbes. Nous appliquons un facteur correctif aux courbes pour solutionner ce problème. Les détails de ce traitement sont expliqués plus en détail dans ce chapitre.

Méthodes de biologie moléculaire

Dans ce chapitre nous décrivons les méthodes de biologie moléculaire. L'étude du comportement d'un système génétique passe nécessairement par l'utilisation des techniques de biologie moléculaire, par exemple, pour modifier le génome de la bactérie. Ce chapitre est consacré à la description des techniques de génétique moléculaire utilisées dans cette thèse.

Pour modifier le génome d'une bactérie, ils existent deux options. La première est d'introduire un fragment circulaire d'ADN qui peut se répliquer de façon indépendante du chromosome, un plasmide. La deuxième est de modifier directement le chromosome grâce à la recombinaison homologue, par exemple. Les deux méthodes ont des avantages et des inconvénients.

Les plasmides contiennent trois éléments principaux. Le premier est l'origine de réplication qui définit leur nombre de copies de plasmides par cellule. Le deuxième est une résistance à un antibiotique qui permet de sélectionner les bactéries qui ont été ''transformées", c'est-à-dire qui ont accepté le plasmide. Le troisième est le système d'expression des gènes que nous voulons introduire dans la bactérie. La liste des plasmides que nous avons utilisés est détaillée dans ce chapitre. Les inconvénients des plasmides sont que, d'une part leur nombre varié dans la population, rajoutant du bruit aux études quantitatives. D'autre part, ils ne peuvent qu'introduire des gènes dans les cellules, mais pas les supprimer ou les altérer.

Pour cela nous utilisons la recombinaison homologue. Le principe de cette méthode repose sur la reconnaissance de deux séquences d'ADN. Si un fragment d'ADN comporte des homologies avec le chromosome à ses extrémités, l'intégralité du fragment peut être incorporée dans le chromosome à l'endroit de l'homologie. Même si la bactérie possède des mécanismes qui permettent la recombinaison, ce phénomène reste très peur fréquent. Pour cette raison, nous utilisons un plasmide pour exprimer le système lambda red qui favorise la recombinaison.

Les cellules qui recombinent représentent une très petite minorité dans la population. Pour les identifier, nous utilisons une double cassette de sélection-contre sélection qui est décrite plus en détail dans ce chapitre. Elle consiste en un gène kan^R, exprimé de façon constitutive, et du gène *ccdB* qui code pour une toxine létal à la cellule. *ccdB* est sous un promoteur inductible par l'arabinose, P_{BAD}.

Lors d'une recombinaison, nous introduisant dans un premier temps le plasmide qui contient le système lambda red. Puis, dans un second temps, nous introduisons la cassette de sélection à l'endroit où nous voulons modifier le génome. Puis, dans un troisième temps, nous la remplaçons la cassette avec notre construction cible. Qu'on utilise un plasmide ou qu'on modifie le génome, il

est nécessaire de synthétiser les fragments d'ADN qui portent les mutations ou gènes à exprimer. Pour ce faire, nous utilisons la PCR (polymerase chain reaction). En utilisant des oligonucléotides synthétiques d'une vingtaine de bases (les amorces), il est possible de copier et amplifier une séquence d'ADN à partir d'une séquence qui sert de modèle. Cette séquence est souvent un fragment d'ADN purifiée ou de l'ADN génomique. Dans une PCR, l'ADN qui sert de modèle est dénaturée le rendant ainsi simple brin. Ceci permet aux séquences synthétiques de s'apparier au fragment. Puis, une ADN polymérase réplique le brin à partir de l'amorce. En rajoutant deux amorces différentes bien choisies, on peut ainsi copier une séquence spécifique à partir d'un premier brin d'ADN.

La PCR permet aussi de relier plusieurs fragments d'ADN entre eux. En utilisant des séquences homologues apportées par les amorces entre différents fragments d'ADN, il est possible de les apparier entre eux et créer un fragment final qui est le concatamer des fragments séparées.

Il est aussi possible de construire des plasmides par PCR. Le protocole utilisé pour combiner plusieurs fragments d'ADN s'appelle « Gibson assembly ». Ce protocole utilise une exo-nucléase qui générer un fragment partiellement simple-brin aux extrémités, permettant ainsi l'appariement entre les différents fragments par homologie. Si tous les bouts sont homologues entre eux, le fragment final est circulaire. Cependant, si le fragment initial et final ne sont pas homologues entre eux, alors le fragment final sera linéaire. L'exo-nucléase ne peut pas dégrader l'ADN circulaire. Pour compenser la dégradation d'un fragment linéaire, nous rajoutons des amorces pour favoriser l'amplification du fragment final et compenser l'activité de l'exo-nucléase. Nos protocoles sont expliqués plus en détail dans ce chapitre.

Les fragments crées par PCR sont vérifiées par électrophorèse, une méthode qui sépare des fragments d'ADN en fonction de leur taille. La vérification finale des clonages se fait par séquençage. Le séquençage est industriel et nous permets de connaître la séquence exacte d'un fragment d'ADN. C'est un contrôle très important qui nous permets d'être sûr que nos constructions sont correctes et sans mutations.

Construction des souches

Dans ce chapitre, nous définissons les souches d'*E. coli* qui vont être utilisés pour l'étude du gène DnaA. Nous décrivons également les modifications que nous allons introduire dans leur génome.

Nous avons construit deux séries souches distinctes, nommées AC et CL. Toutes les souches utilisent le système d'expression *uhp* en tant que promoteur inductible, permettant d'exprimer la protéine DnaA via le promoteur $P_{\mathrm{uhp}T}$. Cependant, la position du promoteur $P_{\mathrm{uhp}T}$ et du gène *dnaA* sont différentes dans les deux souches.

Les deux souches AC et CL ont été construites à partir d'une souche d'*E. coli* BW25113, modifié par Jérôme Izard afin de contrôler l'expression des RNA polymérase via un promoteur inductible dérivé du PLAC. de plus, une fusion traductionnelle a été réalisée sur cette même souche afin de rajouter une étiquette, la protéine fluorescente mCherry, à une des sous-unités de l'ARN polymérase. En modifiant le nombre d'ARN polymérases, nous modifions l'expression génique, tandis qu'avec le gène *dnaA*, nous contrôlons la réplication de l'ADN. Avec la combinaison des deux mécanismes, nous pouvons moduler dans nos souches, de façon indépendante, deux des fonctions principales de la cellule : la transcription et la réplication.

La souche AC possède une seule copie de *dnaA* qui remplace le gène *uhpT* et est sous le contrôle de $P_{\mathrm{uhp}\,T}$. La copie originale du gène *dnaA* a été supprimée seulement partiellement afin de ne pas perturber l'expression de *dnaN* et *recF*, les deux autres gènes de l'operon *dnaA*. Néanmoins, le fragment de *dnaA* restant dans l'opéron original n'est pas fonctionnel. Le gène *dnaA* étant un gène essentiel, nous avons d'abord remplacé le gène *uhpT* par une copie de *dnaA*. Puis, dans une deuxième expérience de modification génomique, nous avons partiellement supprimé la copie originale de *dnaA*. Chaque étape de la construction implique deux recombinaisons homologues successives : la première recombinaison introduit une cassette de sélection-contre sélection à l'endroit cible sur le chromosome, la deuxième recombinaison remplace la cassette par la séquence finale souhaitée.

La souche CL possède également une seule copie de *dnaA* dans son emplacement original dans le chromosome. Ce sont les promoteurs du gène *dnaA* originales que nous avons remplacé par le promoteur $P_{\mathrm{uhp}\,T}$. Nous avons aussi supprimé le gène *uhpT* avec son promoteur afin de ne pas modifier radicalement l'expression de $P_{\mathrm{uhp}\,T}$. Nous avons commencé par supprimer le gène *uhpT* et son promoteur avec le même procédé qu'avec la souche AC. Ensuite, nous avons remplacé les promoteurs de *dnaA* en un seul événement de recombinaison homologue par $P_{\mathrm{uhp}\,T}$ et un gène de résistance à kanamycine.

Toutes les modifications que nous avons faites ont nécessité l'introduction préalable d'un plasmide afin de favoriser la recombinaison homologue. Ce plasmide ne se réplique pas si la température est au-dessus de 37°. Cette propriété nous a permis de l'éliminer des souches finales. Le détail des constructions et les séquences utilisées sont détaillés dans ce chapitre.

Analyse au niveau des populations

Nous avons commencé la caractérisation des souches modifiés en étudiant leur comportement moyen en population. Nous avons mesuré des propriétés comme la croissance ou l'auto-fluorescence des cellules en fonction du niveau d'expression de DnaA. Cette étude est complémentaire à une étude en cellule unique que nous décrirons dans le chapitre suivant.

On s'attend à que nos souches modifiées poussent normalement lorsque nous induisons le promoteur $P_{\mathrm{uhp}\,T}$ qui exprime *dnaA*. En baissant le niveau d'expression de *dnaA*, le taux d'initiation de la réplication des bactéries devrait également baisser. De plus, il est connu qu'une forte surexpression de *dnaA* mène également à un arrêt de croissance. Nous avons fait pousser nos cellules en présence de différentes concentrations d'inducteur (g6p) dans le milieu et, effectivement, si le promoteur $P_{\mathrm{uhp}\,T}$ est trop faiblement induit, les cellules s'arrêtent de pousser. De façon étonnante, la transition entre croissance et arrêt de croissance se produit très abruptement à une concentration critique de g6p dans le milieu. Ce seuil critique est légèrement différent pour les deux souches AC et CL. La souche AC nécessite des concentrations plus élevées pour croître, et une induction totale de $P_{\mathrm{uhp}\,T}$ ne modifie pas son taux de croissance, qui reste toujours maximal. La souche CL nécessite moins de g6p pour que les cellules poussent et l'induction maximale de $P_{\mathrm{uhp}\,T}$ réduit le taux de croissance légèrement. Bien que nos deux promoteurs aient exactement la même séquence, leur position sur le chromosome est différente. Cet effet de positionnement peut changer la force du promoteur. Il semblerait donc que dans la souche CL, $P_{\mathrm{uhp}\,T}$ exprime

plus de protéines de DnaA à concentrations égales de g6p. Cependant, la réponse des cellules par rapport à la quantité de DnaA produite est qualitativement la même. Dans aucune des deux souches, nous avons observé un ralentissement progressif du taux de croissance lorsque l'expression de DnaA est diminuée graduellement. Soit les bactéries poussent normalement à leur taux de croissance maximale, soit elles ne poussent pas du tout. La réponse à des faibles concentrations en g6p est du type tout-ou-rien.

Nous avons modélisé la variation de DnaA dans nos cellules en écrivant les équations correspondantes et nous avons trouvé que la concentration moyenne de DnaA dans des cellules qui poussent à un taux nominal ne dépend que de $P_{\mathrm{uhp}\,T}$. De plus, nous avons observé qu'enlever le g6p du milieu dans lequel des bactéries étaient en train de pousser n'arrête pas immédiatement leur croissance. En d'autres termes, quand on arrête l'expression de $P_{\mathrm{uhp}\,T}$ dans une population de bactéries qui auparavant poussaient normalement, les bactéries continuent de grandir exponentiellement pendant quelques heures. Plus l'induction de $P_{\mathrm{uhp}\,T}$ était forte pendant la croissance exponentielle, plus longtemps les bactéries vont continuer à pousser avant de s'arrêter.

Nous avons exploité ce comportement pour étudier l'activité de $P_{\mathrm{uhp}\,T}$. Nous avons utilisé des mesures de densité optique, d'auto-fluorescence issu de flavines et de fluorescence rouge (provenant de fusions traductionnelles de mCherry-RNA avec l'ARN polymérase) pour explorer le niveau d'activité de $P_{\mathrm{uhp}\,T}$. Par la suite, nous avons comparé nos mesures à une calibration publiée en unités Miller (MU) afin de quantifier le nombre moyen de protéines de DnaA dans les cellules en fonction de la concentration de g6p dans le milieu. Cependant, le site de liaison des ribosomes dans nos constructions est différent à celle des cellules étudiées pour la calibration de $P_{\mathrm{uhp}\,T}$ dans la littérature. En conséquence, nous pouvons pour le moment que donner une borne supérieur du nombre de protéines de DnaA dans la cellule. Néanmoins, ces mesures nous permettent de conclure que la bactérie sauvage contient bien plus de protéines de DnaA que les seules protéines de DnaA nécessaires pour initier la réplication de l'ADN.

Par la suite, nous nous sommes intéressés à savoir si des cellules, qui ont arrêtées leur croissance à cause d'une trop faible quantité de DnaA, peuvent reprendre une croissance normale si DnaA est de nouveau exprimée. Nous avons découvert que très rapidement, les cellules d'une population ne qui pousse plus de manière exponentielle ont du mal à reprendre la croissance. Lorsqu'on arrête la production de DnaA, les bactéries persistent d'abord en croissance exponentielle pendant quelques heures. Le taux de croissance diminue ensuite progressivement jusqu'à l'arrêt complet de la croissance. Les cellules qui restent ensuite plus longtemps sans croître avant d'exprimer DnaA de nouveau (en rajoutant du g6p dans le milieu), montrent également un délai plus grand avant de recommencer une croissance exponentielle. Même quand la ré-croissance tarde très longtemps, les cellules continuent à rejeter des flavines. Leur métabolisme continue donc à fonctionner. En étudiant le nombre des cellules viables en utilisant des ré-croissances sous forme de colonies dans des boites de Pétri, nous avons déterminé que les cellules qui ne répliquent plus leur chromosome perdent leur viabilité quelques heures après. La perte de viabilité est très brusque et rapide. Cependant, autour de 0.5% de la population reste toujours viable, peu importe combien de temps les cellules passent sans répliquer leur chromosome.

Nous avons étudiée des possibles facteurs qui pourraient expliquer cette perte de viabilité. Le premier est la surexpression des protéines DnaN et RecF par la souche AC. Nous avons comparé

la ré-croissance de la souche AC avec celle de la souche CL. Même si les cellules de la souche CL ont repoussé légèrement plus vite, la plus part d'entre elles ont aussi perdu leur viabilité. Donc DnaN at RecF ne sont pas les principaux responsables de la perte de viabilité. Nous avons aussi diminué de manière générale l'expression des protéines dans la souche AC en réduisant l'expression des RNA polymérases, mais cela n'a pas affecté la viabilité cellulaire. Nous ignorons pourquoi les cellules ne sont plus viables quand elles passent trop de temps sans pouvoir répliquer leur ADN.

Les phénomènes que nous avons observés montrent que dans une population certaines cellules ne se comportent pas comme les autres. Certaines bactéries restent viables tandis que d'autres perdent leur capacité de recroître. De plus, il est fort probable que toutes les bactéries d'une population ne s'arrêtent pas de pousser au même temps quand on enlève g6p du milieu. Pour étudier ces phénomènes, nous devons étudier les cellules individuelles et non seulement des populations de cellules.

Analyse en cellule unique

L'expression de protéines est un processus stochastique. Dans ce chapitre nous explorons en détail comment les bactéries individuelles réagissent à différents niveaux d'expression de DnaA et à son absence.

Nous avons établi dans le chapitre précédant, qu'une population des bactéries croit de manière exponentielle si la concentration intracellulaire de DnaA est au-dessus d'une concentration critique. Dans ce chapitre, nous montrons en utilisant des circuits de microfluidiques, que des individus au sein d'une population en conditions de croissance exponentielle, peuvent arrêter de se diviser. Nous avons mesuré le pourcentage des bactéries de la souche AC8 qui arrêtent de se diviser par génération en fonction du niveau d'induction par g6p. Ce niveau est très faible pour une concentration d'inducteur $1\mu M$ g6p à 2% alors que pour une induction à $0.8\mu M$ g6p ca augmente rapidement à 10%.

En utilisant le même circuit microfluidique, nous avons observé que lors d'un arrêt d'induction de $P_{\mathrm{uhp}\,T}$, les bactéries ne s'arrêtent pas toutes après un même nombre de divisions. Dans notre modèle cela se traduit par une distribution de « réservoirs de DnaA » initiaux, avant l'arrêt d'induction. Dans une série d'expériences dédiées, nous avons observé des bactéries posées sur des pads constitués des gels d'agar et milieu de culture sans g6p. Les bactéries individuelles, provenant préalablement d'une préculture en phase exponentielle, une fois posées sur le pad, se sont divisé un certain nombre de fois avant de s'arrêter. En comptant la descendance que chaque bactérie ait pu générer, on a établi l'histogramme du nombre de descendants dans une population de bactéries. Nous avons fait cela pour 3 concentrations de g6p dans la préculture. Les 3 distributions ainsi obtenues, sont bien décrites par de distributions gamma.

Il a été montré que la distribution de gamma décrit fidèlement la presque totalité des distributions de protéines dans la bactérie E.coli. La distribution gamma a deux paramètres, a et b, liés respectivement au nombre d'ARN/cycle cellulaire et au nombre de protéines/durée de vie d'ARNm. Nous avons trouvé que pendant la croissance exponentielle avec des concentrations différentes de g6p, b reste constant, alors que seulement a varie. Ceci est cohérent avec notre modèle : si la concentration g6p augmente, le taux de transcription du promoteur $P_{\mathrm{uhp}\,T}$ fera de

même, donc les nombres des transcrits par cycle cellulaire c'est-à-dire le paramètre a aussi. Par ailleurs si rien, ni dans le temps de vie des ARNm, ni dans la chaine de traduction ne change, il est donc normal que le paramètre b reste constant.

En combinant ces données avec nos estimations de la concentration DnaA au sein de la population de bactéries faite au chapitre précédant, nous avons estimé la borne supérieure de la concentration critique de DnaA nécessaires pour la réplication de l'ADN. En d'autres termes, nous avons estimé que si une bactérie a une concentration de DnaA de 85nM ou plus, elle peut répliquer son ADN.

Les distributions nous ont aussi permis de construire un modèle simple afin d'estimer combien de temps les cellules restent viables après leur dernière division. En comparant notre modèle avec les résultats du chapitre précédent, nous avons estimé que les bactéries restent viables 4 heures après leur dernière division. Or leur activité métabolique de certaines d'en elle reste active plus longtemps. Nous avons observé des cellules qui après avoir arrêté leur division continuent à s'allonger pendant plusieurs heures.

Discussion et perspectives

Le but principal de cette thèse est d'étudier le rôle de la protéine DnaA dans l'initiation de la réplication de l'ADN. En particulier, nous examinons les conséquences sur la réplication lorsque le nombre de protéines de DnaA est faible. Afin de mener à bien cette étude, nous avons dû développer et caractériser un nouveau système d'expression inductible, basé sur le promoteur $P_{\mathrm{uhp}T}$. Nous avons utilisé ce promoteur pour nos constructions permettant une expression contrôlée de *dnaA*.

Nous avons construit deux souches distinctes pour faire notre étude. Les deux utilisent le promoteur $P_{\mathrm{uhp}T}$ afin d'exprimer DnaA via l'induction par glucose-6-phosphate (g6p). Le promoteur $P_{\mathrm{uhp}T}$ peut exprimer entre 0 et 1200 protéines par génération dans du milieu minimum MOPS, ce qui est dans la gamme de concentration de DnaA chez *E. coli*. Bien que les promoteurs dans les deux souches AC et CL soient identiques, ils ne sont pas intégrés au même endroit dans le chromosome. Dans la souche CL, le promoteur $P_{\mathrm{uhp}T}$ n'est pas dans sa position usuelle, contrairement à la souche AC. Ce changement de position conduit à une force plus importante du promoteur $P_{\mathrm{uhp}T}$ dans la souche CL, comparé à la souche AC.

Avec nos constructions, nous pouvons contrôler l'expression de DnaA dans les souches AC et CL. Lorsqu'on induit l'expression de *dnaA*, les bactéries poussent normalement. L'arrêt de l'expression de *dnaA* conduit à un arrêt de la croissance. Les cellules qui ne produisaient pas assez de DnaA se divisent encore pendant quelques générations, puis s'arrêtent entièrement. Nous avons utilisé cette propriété et les informations sur $P_{\mathrm{uhp}T}$ afin d'estimer le nombre de protéines de DnaA dans les cellules en phase exponentielle. Puisque l'efficacité de traduction de nos constructions est différente de celle rapporté pour *uhp* dans la littérature, nous ne pouvons seulement donner une limite supérieure aux quantifications de protéines de DnaA chez *E. coli*, plutôt qu'un nombre exact. En combinant nos mesures sur la population et en cellule unique, nous avons estimé que les bactéries sont capables d'initier une réplication lorsqu'elles contiennent plus de 85 protéines de DnaA. Ce nombre est bien inférieur aux 520 protéines de DnaA présentes

en moyenne dans les bactéries. En plus d'être indispensable pour la réplication, DnaA régule également de nombreux gènes. Le surplus de molécules pourrait être nécessaire à ces fonctions.

Même après l'arrêt de la division (à cause de l'épuisement du réservoir de DnaA), les cellules continuent de s'allonger pendant plusieurs heures. Certaines cellules arrêtent leur croissance après quelques heures tandis que d'autres continuent de s'allongent même 20 heures après leur dernière division. Le métabolisme de ces cellules continue de fonctionner. Des mesures d'auto fluorescence, due à la production de flavines, confirment cette hypothèse. Cependant, la viabilité des cellules chute drastiquement si elles ne redémarrent pas rapidement la réplication de l'ADN après leur dernière division. A l'aide d'un modèle simple, nous avons estimé que quatre heures après leur dernière division, les cellules perdent leur capacité à reprendre la croissance par une nouvelle induction de l'expression de DnaA. Ceci pourrait être attribué à une surexpression de DnaN et RecF, induite par l'absence de DnaA dans la cellule. Cependant, la souche CL, qui ne surexprimé pas DnaN et RecF, ne supporte pas beaucoup mieux l'absence de réplication de son ADN. La perte de viabilité des cellules n'est donc pas dû a DnaN ni RecF, même si leur surexpression est pénalisant pour la cellule. Il est fortement probable que l'ADN des cellules soit endommagé pendant qu'elles ne peuvent pas le répliquer et que ceci conduit à la perte de viabilité. Néanmoins, nos résultats, ainsi que les expériences publiées, laissent penser que les lésions ne provoquent pas une réponse SOS dans les cellules.

L'étude de la croissance après l'arrêt de production de DnaA nous a permis de modéliser la quantité de DnaA dans les cellules pendant leur croissance avec une distribution gamma. Cette distribution dépend de deux paramètres. Le premier représente la fréquence avec laquelle la cellule exprime des ARN messager par cycle cellulaire. Le deuxième peut être interprété comme le nombre de protéines traduites par ARN messager. Ce deuxième paramètre ne change pas pour des concentrations différentes de g6p. Ceci est cohérent avec notre modèle : si la concentration g6p augmente, le taux de transcription du promoteur $P_{\mathrm{uhp}T}$ fera de même, donc les nombres des transcrits par cycle cellulaire aussi. Par ailleurs si rien, ni dans le temps de vie des ARNm, ni dans la chaine de traduction ne change, il est donc normal que le deuxième paramètre reste constant. $P_{\mathrm{uhp}T}$ est activé par la protéine UhpA phosphorylé. C'est la protéine UhpB qui transfert un phosphate sur UhpA lorsque la protéine UhpC détecte du g6p dans le milieu. Par son activité phosphatase, UhpB peut également enlever le phosphate d'UhpA-phosphate. Le résultat net du système de transduction du signal *uhp* est une activation de $P_{\mathrm{uhp}T}$ par le g6p dans le milieu. Nous avons commencé à développer un modèle de ce système d'induction. Des simulations préliminaires sont encourageantes et prédisent un nombre total de protéines UhpA phosphorylées qui est compatible avec des observations de la littérature.

Bien que nous ayons considérablement progressé dans la compréhension de l'initiation de la réplication par DnaA et du promoteur d'*uhpT*, ils nous restent encore plusieurs éléments à étudier. Par exemple, l'activité métabolique persiste dans une population de cellules dont la croissance a été arrêtée. Nous devons caractériser cette activité métabolique et déterminer quelle sous-population de cellules maintient cette activité. Nous voudrions aussi terminer le développement du modèle du système d'expression *uhp* et affiner le modèle qui nous a permis d'estimer le temps pendant que les cellules restent viables après leur dernière division. Nous comptons également travailler sur une quantification absolue de DnaA dans nos souches par des

méthodes de protéomique.

Nous avons réalisé les objectifs que nous nous étions fixés au début de la thèse. Nos souches permettent d'étudier la protéine DnaA et ses effets lorsque ils sont insuffisantes pour engager la réplication de l'ADN. Nous avons obtenu des résultats quantitatifs sur DnaA et le promoteur UHPT. Notre système est une possibilité parmi d'autres qui permet d'étudier l'absence de réplication de l'ADN sans arrêter des fourches de réplication et stresser la cellule. Le système CRISPR-Cas9 a récemment été utilisé pour bloquer l'origine de réplication. Néanmoins, ce système alternatif n'est pas facilement réversible et plus contraint dans les températures de croissance. Nous avons donc construit un outil qui a déjà fourni des données précieuses et qui pourra être exploité davantage pour répondre à des questions fondamentales ainsi que pour le développement de souches d'intérêt biotechnologiques.

Bibliography

[1] Daniël T. Verhamme, Pieter W. Postma, Wim Crielaard, and Klaas J. Hellingwerf. Cooperativity in Signal Transfer through the Uhp System of Escherichia coli. *Journal of Bacteriology*, 184(15):4205–4210, January 2002.

[2] James D. Watson. *Molecular Biology of the Gene, Fifth Edition.* Cold Spring Harbor Laboratory Press, San Francisco, 5th edition edition, December 2003.

[3] Alvaro Sanchez, Melisa L Osborne, Larry J Friedman, Jane Kondev, and Jeff Gelles. Mechanism of transcriptional repression at a bacterial promoter by analysis of single molecules. *The EMBO Journal*, 30(19):3940–3946, October 2011.

[4] L. B. Augustin, B. A. Jacobson, and J. A. Fuchs. Escherichia coli Fis and DnaA proteins bind specifically to the nrd promoter region and affect expression of an nrd-lac fusion. *Journal of Bacteriology*, 176(2):378–387, January 1994.

[5] Christoph Schaefer, Andreas Holz, and Walter Messer. DnaA Protein Mediated Transcription Termination in the GUA Operon of Escherichia coli. In Patrick Hughes, Ellen Fanning, and Masamichi Kohiyama, editors, *DNA Replication: The Regulatory Mechanisms*, pages 161–168. Springer Berlin Heidelberg, Berlin, Heidelberg, 1992.

[6] Sébastien Wielgoss, Jeffrey E. Barrick, Olivier Tenaillon, Stéphane Cruveiller, Béatrice Chane-Woon-Ming, Claudine Médigue, Richard E. Lenski, and Dominique Schneider. Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution Experiment With Escherichia coli. *G3: Genes, Genomes, Genetics*, 1(3):183–186, August 2011.

[7] Jeffrey H Miller. *Experiments in molecular genetics.* Cold Spring harbor Laboratory, 1992.

[8] X. Zhang and H. Bremer. Control of the Escherichia coli rrnB P1 promoter strength by ppGpp. *Journal of Biological Chemistry*, 270(19):11181–11189, January 1995.

[9] Hernan G. Garcia, Heun Jin Lee, James Q. Boedicker, and Rob Phillips. Comparison and Calibration of Different Reporters for Quantitative Analysis of Gene Expression. *Biophysical Journal*, 101(3):535–544, August 2011.

[10] Thomas Esquerré, Sandrine Laguerre, Catherine Turlan, Agamemnon J. Carpousis, Laurence Girbal, and Muriel Cocaign-Bousquet. Dual role of transcription and transcript stability in the regulation of gene expression in Escherichia coli cells cultured on glucose at different growth rates. *Nucleic Acids Research*, 42(4):2460–2472, February 2014.

[11] O. Shimomura, F. H. Johnson, and Y. Saiga. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *Journal of Cellular and Comparative Physiology*, 59:223–239, June 1962.

[12] M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher. Green fluorescent protein as a marker for gene expression. *Science (New York, N.Y.)*, 263(5148):802–805, February 1994.

[13] Dmitriy M Chudakov, Mikhail V Matz, Sergey Lukyanov, and Konstantin A Lukyanov. Fluorescent proteins and their applications in imaging living cells and tissues. *Physiological reviews*, 90(3):1103–1163, 2010.

[14] F G Hansen, T Atlung, R E Braun, A Wright, P Hughes, and M Kohiyama. Initiator (DnaA) protein concentration as a function of growth rate in Escherichia coli and Salmonella typhimurium. *J. Bacteriol.*, 173(16):5194–5199, August 1991.

[15] Mark D. Sutton, Kevin M. Carr, Matias Vicente, and Jon M. Kaguni. Escherichia coli DnaA Protein THE N-TERMINAL DOMAIN AND LOADING OF DnaB HELICASE AT THEE. COLI CHROMOSOMAL ORIGIN. *Journal of Biological Chemistry*, 273(51):34255–34262, December 1998.

[16] F Blaesing, C Weigel, M Welzeck, and W Messer. Analysis of the DNA-binding domain of Escherichia coli DnaA protein. *Molecular microbiology*, 36(3):557–569, 2000.

[17] Yukinori Hirota, José Mordoh, and François Jacob. On the process of cellular division in Escherichia coli. *Journal of Molecular Biology*, 53(3):369–387, November 1970.

[18] Melissa L. Mott and James M. Berger. DNA replication initiation: mechanisms and regulation in bacteria. *Nature Reviews Microbiology*, 5(5):343–354, January 2007.

[19] J. Izard, C. D. Gomez Balderas, D. Ropers, S. Lacour, X. Song, Y. Yang, A. B. Lindner, J. Geiselmann, and H. de Jong. A synthetic growth switch based on controlled expression of RNA polymerase. *Molecular Systems Biology*, 11(11):840–840, November 2015.

[20] Tobias Bollenbach, Selwyn Quan, Remy Chait, and Roy Kishony. Nonoptimal Microbial Response to Antibiotics Underlies Suppressive Drug Interactions. *Cell*, 139(4):707–718, 2009.

[21] Bénédicte Michel. After 30 years of study, the bacterial SOS response still surprises us. *PLoS Biology*, 3(7):1174–1176, 2005.

[22] B Michel, S D Ehrlich, and M Uzest. DNA double-strand breaks caused by replication arrest. *The EMBO journal*, 16(2):430–8, 1997.

[23] Philippe Quillardet, Marie-Ange Rouffaud, and Philippe Bouige. DNA array analysis of gene expression in response to UV irradiation in Escherichia coli. *Research in Microbiology*, 154(8):559–572, October 2003.

[24] Chiara Saggioro, Anne Olliver, and Bianca Sclavi. *Temperature dependence of DnaA-DNA interaction and its effect on the auto-regulation of dnaA expression.* 2012.

[25] Robert E. Braun, Kathy O'Day, and Andrew Wright. Autoregulation of the DNA replication gene dnaA in E. coli K-12. *Cell*, 40(1):159–169, January 1985.

[26] P M Burgers, A Kornberg, and Y Sakakibara. The dnaN gene codes for the beta subunit of DNA polymerase III holoenzyme of escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 78(9):5391–5395, September 1981.

[27] P. T. Stukenberg, P. S. Studwell-Vaughan, and M. O'Donnell. Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. *Journal of Biological Chemistry*, 266(17):11328–11334, June 1991.

[28] Mark D. Sutton. The Escherichia coli dnaN159 Mutant Displays Altered DNA Polymerase Usage and Chronic SOS Induction. *Journal of Bacteriology*, 186(20):6738–6748, October 2004.

[29] Kin-Hoe Chow and Justin Courcelle. RecO Acts with RecF and RecR to Protect and Maintain Replication Forks Blocked by UV-induced DNA Damage in Escherichia coli. *Journal of Biological Chemistry*, 279(5):3492–3496, January 2004.

[30] Aline V. Grigorian, Rachel B. Lustig, Elena C. Guzmán, Joseph M. Mahaffy, and Judith W. Zyskind. Escherichia coli Cells with Increased Levels of DnaA and Deficient in Recombinational Repair Have Decreased Viability. *Journal of Bacteriology*, 185(2):630–644, January 2003.

[31] Magdalena M. FELCZAK and Jon M. KAGUNI. DnaAcos hyperinitiates by circumventing regulatory pathways that control the frequency of initiation in Escherichia coli. *Molecular microbiology*, 72(6):1348–1363, 2009.

[32] Lyle A. Simmons, Adam M. Breier, Nicholas R. Cozzarelli, and Jon M. Kaguni. Hyperinitiation of DNA replication in Escherichia coli leads to replication fork collapse and inviability. *Molecular Microbiology*, 51(2):349–358, January 2004.

[33] M. E. Armengod, M. García-Sogo, and E. Lambíes. Transcriptional organization of the dnaN and recF genes of Escherichia coli K-12. *Journal of Biological Chemistry*, 263(24):12109–12114, August 1988.

[34] M. Villarroya. Stationary phase induction of dnaN and recF, two genes of Escherichia coli involved in DNA replication and repair. *The EMBO Journal*, 17(6):1829–1837, March 1998.

[35] Bjarke Bak Christensen, Tove Atlung, and Flemming G. Hansen. DnaA Boxes Are Important Elements in Setting the Initiation Mass of Escherichia coli. *J. Bacteriol.*, 181(9):2683–2688, May 1999.

[36] Alan C. Leonard and Julia E. Grimwade. Regulation of DnaA Assembly and Activity: Taking Directions From the Genome. *Annual review of microbiology*, 65:19–35, 2011.

[37] Qing Zhang and Hualin Shi. Coupling chromosomal replication to cell growth by the initiator protein DnaA in Escherichia coli. *Journal of Theoretical Biology*, 314:164–172, December 2012.

[38] Leise Riber, Jakob Frimodt-Møller, Godefroid Charbon, and Anders Løbner-Olesen. Multiple DNA Binding Proteins Contribute to Timing of Chromosome Replication in E. coli. *Frontiers in Molecular Biosciences*, 3, 2016.

[39] Takuma Ishida, Nobuyoshi Akimitsu, Tamami Kashioka, Masakazu Hatano, Toshio Kubota, Yasuyuki Ogata, Kazuhisa Sekimizu, and Tsutomu Katayama. DiaA, a Novel DnaA-binding Protein, Ensures the Timely Initiation of Escherichia coli Chromosome Replication. *Journal of Biological Chemistry*, 279(44):45546–45555, October 2004.

[40] Kenji Keyamura, Norie Fujikawa, Takuma Ishida, Shogo Ozaki, Masayuki Su'etsugu, Kazuyuki Fujimitsu, Wataru Kagawa, Shigeyuki Yokoyama, Hitoshi Kurumizaka, and Tsutomu Katayama. The interaction of DiaA and DnaA regulates the replication cycle in E. coli by directly promoting ATP–DnaA-specific initiation complexes. *Genes & Development*, 21(16):2083–2099, August 2007.

[41] D. S. Hwang and A. Kornberg. Opening of the replication origin of Escherichia coli by DnaA protein with protein HU or IHF. *Journal of Biological Chemistry*, 267(32):23083–23086, November 1992.

[42] Sundari Chodavarapu, Magdalena M. Felczak, Josette Rouvière Yaniv, and Jon M. Kaguni. Escherichia coli DnaA interacts with HU in initiation at the E. coli replication origin. *Molecular Microbiology*, 67(4):781–792, February 2008.

[43] Mark D. Sutton, Kevin M. Carr, Matias Vicente, and Jon M. Kaguni. Escherichia coli DnaA Protein THE N-TERMINAL DOMAIN AND LOADING OF DnaB HELICASE AT THEE. COLI CHROMOSOMAL ORIGIN. *Journal of Biological Chemistry*, 273(51):34255–34262, December 1998.

[44] Judith W. Zyskind and Douglas W. Smith. The bacterial origin of replication, oriC. *Cell*, 46(4):489–490, August 1986.

[45] Sukhyun Kang, Ho Lee, Joo Seok Han, and Deog Su Hwang. Interaction of SeqA and Dam Methylase on the Hemimethylated Origin of Escherichia coli Chromosomal DNA Replication. *Journal of Biological Chemistry*, 274(17):11463–11468, April 1999.

[46] G. Newman and E. Crooke. DnaA, the Initiator of Escherichia coli Chromosomal Replication, Is Located at the Cell Membrane. *Journal of Bacteriology*, 182(9):2604–2610, May 2000.

[47] K Sekimizu, BY Yung, and A Kornberg. The dnaA protein of Escherichia coli. Abundance, improved purification, and membrane binding. *J. Biol. Chem.*, 263(15):7136–7140, May 1988.

[48] Gene-Wei Li, David Burkhardt, Carol Gross, and Jonathan S. Weissman. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3):624–635, April 2014.

[49] Joseph A. Bogan and Charles E. Helmstetter. DNA sequestration and transcription in the oriC region of Escherichia coli. *Molecular Microbiology*, 26(5):889–896, December 1997.

[50] Risa Kitagawa, Toru Ozaki, Shigeki Moriya, and Tohru Ogawa. Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for Escherichia coli DnaA protein. *Genes & Development*, 12(19):3032–3043, January 1998.

[51] Masayuki Su'etsugu, Toh-ru Shimuta, Takuma Ishida, Hironori Kawakami, and Tsutomu Katayama. Protein Associations in DnaA-ATP Hydrolysis Mediated by the Hda-Replicase Clamp Complex. *Journal of Biological Chemistry*, 280(8):6528–6536, February 2005.

[52] Tsutomu Katayama, Shogo Ozaki, Kenji Keyamura, and Kazuyuki Fujimitsu. Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC. *Nature Reviews Microbiology*, 8(3):163–170, January 2010.

[53] Kazuyuki Fujimitsu, Takayuki Senriuchi, and Tsutomu Katayama. Specific genomic sequences of E. coli promote replicational initiation by directly reactivating ADP-DnaA. *Genes & Development*, 23(10):1221–1233, May 2009.

[54] Kazutoshi Kasho, Kazuyuki Fujimitsu, Toshihiro Matoba, Taku Oshima, and Tsutomu Katayama. Timely binding of IHF and Fis to DARS2 regulates ATP–DnaA production and replication initiation. *Nucleic Acids Research*, 42(21):13134–13149, January 2014.

[55] Tove Atlung, Erik S. Clausen, and Flemming G. Hansen. Autoregulation of the dnaA gene of Escherichia coli K12. *Molecular and General Genetics MGG*, 200(3):442–450, August 1985.

[56] C. Speck. ATP- and ADP-DnaA protein, a molecular switch in gene regulation. *The EMBO Journal*, 18(21):6169–6176, November 1999.

[57] Chiara Saggioro, Anne Olliver, and Bianca Sclavi. Temperature-dependence of the DnaA-DNA interaction and its effect on the autoregulation of dnaA expression. *The Biochemical Journal*, 449(2):333–341, January 2013.

[58] Yong Lee, Ho Lee, Jeongbin Yim, and Deog Hwang. The binding of two dimers of IciA protein to the dnaA promoter 1p element enhances the binding of RNA polymerase to the dnaA promoter 1p. *Nucleic Acids Research*, 25(17):3486–3489, January 1997.

[59] Lizbeth Hedstrom. IMP Dehydrogenase: Structure, Mechanism and Inhibition. *Chemical reviews*, 109(7):2903–2928, July 2009.

[60] Kuifu He, Yuechao Ma, Shanshan Du, Xixian Xie, Qingyang Xu, and Ning Chen. [Effects of overexpression of key enzyme genes on guanosine accumulation in Bacillus amyloliquefaciens]. *Wei Sheng Wu Xue Bao = Acta Microbiologica Sinica*, 52(6):718–725, June 2012.

[61] D. B. Straus, W. A. Walter, and C. A. Gross. The heat shock response of E. coli is regulated by changes in the concentration of sigma 32. *Nature*, 329(6137):348–351, September 1987.

[62] Birgitte H. Kallipolitis and Poul Valentin-Hansen. Transcription of rpoH, encoding the Escherichia coli heat-shock regulator σ32, is negatively controlled by the cAMP-CRP/CytR nucleoprotein complex. *Molecular Microbiology*, 29(4):1091–1099, August 1998.

[63] Q. P. Wang and J. M. Kaguni. dnaA protein regulates transcriptions of the rpoH gene of Escherichia coli. *Journal of Biological Chemistry*, 264(13):7338–7344, May 1989.

[64] A. D. Grossman, Y. N. Zhou, C. Gross, J. Heilig, G. E. Christie, and R. Calendar. Mutations in the rpoH (htpR) gene of Escherichia coli K-12 phenotypically suppress a temperature-sensitive mutant defective in the sigma 70 subunit of RNA polymerase. *Journal of Bacteriology*, 161(3):939–943, January 1985.

[65] Olwen M. Birch, Kirsty S. Hewitson, Martin Fuhrmann, Knut Burgdorf, Jack E. Baldwin, Peter L. Roach, and Nicholas M. Shaw. MioC Is an FMN-binding Protein That Is Essential forEscherichia coli Biotin Synthase Activity in Vitro. *Journal of Biological Chemistry*, 275(41):32277–32280, October 2000.

[66] T. Ogawa and T. Okazaki. Concurrent transcription from the gid and mioC promoters activates replication of an Escherichia coli minichromosome. *Molecular & general genetics: MGG*, 230(1-2):193–200, November 1991.

[67] D. B. Bates, E. Boye, T. Asai, and T. Kogoma. The absence of effect of gid or mioC transcription on the initiation of chromosomal replication in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 94(23):12497–12502, November 1997.

[68] A. Løbner-Olesen, T. Atlung, and K. V. Rasmussen. Stability and replication control of Escherichia coli minichromosomes. *Journal of Bacteriology*, 169(6):2835–2842, January 1987.

[69] E. E.A. Verhoeven. Architecture of nucleotide excision repair complexes: DNA is wrapped by UvrB before and after damage recognition. *The EMBO Journal*, 20(3):601–611, February 2001.

[70] Eralp Arikan, Mahroooda S. Kulkami, David C. Thomas, and Aziz Sancar. Sequences of the E. coli uvrB gene and protein. *Nucleic Acids Research*, 14(6):2637–2650, November 1986.

[71] Fai-Chu Wong, Penny J. Beuning, Maria Nagan, Kiyotaka Shiba, and Karin Musier-Forsyth. Functional Role of the Prokaryotic Proline-tRNA Synthetase Insertion Domain in Amino Acid Editing. *Biochemistry*, 41(22):7108–7115, June 2002.

[72] Z. Zhou and M. Syvanen. Identification and sequence of the drpA gene from Escherichia coli. *Journal of Bacteriology*, 172(1):281–286, January 1990.

[73] E. R. Archibold and L. S. Williams. Regulation of Synthesis of Methionyl-, Prolyl-, and Threonyl-Transfer Ribonucleic Acid Synthetases of Escherichia coli. *Journal of Bacteriology*, 109(3):1020–1026, January 1972.

[74] Akiko Ohba, Tohru Mizushima, Tsutomu Katayama, and Kazuhisa Sekimizu. Amounts of proteins altered by mutations in the dnaA gene of Escherichia coli. *FEBS Letters*, 404(2–3):125–128, March 1997.

[75] Walter Messer and Christoph Weigel. DnaA initiator—also a transcription factor. *Molecular Microbiology*, 24(1):1–6, April 1997.

[76] L. A. Sonna, S. V. Ambudkar, and P. C. Maloney. The mechanism of glucose 6-phosphate transport by Escherichia coli. *Journal of Biological Chemistry*, 263(14):6625–6630, May 1988.

[77] Tomoya Baba, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill A Datsenko, Masaru Tomita, Barry L Wanner, and Hirotada Mori. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology*, 2:2006.0008, 2006.

[78] TJ Merkel, JL Dahl, RH Ebright, and RJ Kadner. Transcription activation at the Escherichia coli uhpT promoter by the catabolite gene activator protein. *J. Bacteriol.*, 177(7):1712–1718, April 1995.

[79] I. N. Olekhnovich and R. J. Kadner. Mutational Scanning and Affinity Cleavage Analysis of UhpA-Binding Sites in the Escherichia coli uhpT Promoter. *Journal of Bacteriology*, 184(10):2682–2691, May 2002.

[80] I. N. Olekhnovich, J. L. Dahl, and R. J. Kadner. Separate contributions of UhpA and CAP to activation of transcription of the uhpT promoter of Escherichia coli. *Journal of Molecular Biology*, 292(5):973–986, October 1999.

[81] Carol A. Webber and Robert J. Kadner. Involvement of the amino-terminal phosphorylation module of UhpA in activation of uhpT transcription in Escherichia coli. *Molecular Microbiology*, 24(5):1039–1048, June 1997.

[82] Jesse S. Wright, Igor N. Olekhnovich, Gail Touchie, and Robert J. Kadner. The Histidine Kinase Domain of UhpB Inhibits UhpA Action at the Escherichia coli uhpT Promoter. *Journal of Bacteriology*, 182(22):6279–6286, November 2000.

[83] John L. Dahl, Bei Yang Wei, and Robert J. Kadner. Protein phosphorylation affects binding of the Escherichia coli transcription activator UhpA to the uhpT promoter. *Journal of Biological Chemistry*, 272(3):1910–1919, 1997.

[84] Daniël T. Verhamme, Jos C. Arents, Pieter W. Postma, Wim Crielaard, and Klaas J. Hellingwerf. Glucose-6-phosphate-dependent phosphoryl flow through the Uhp two-component regulatory system. *Microbiology*, 147(12):3345–3352, 2001.

[85] Donna M. Shattuck-Eidens and Robert J. Kadner. Molecular Cloning of the uhp Region and Evidence for a Positive Activator for Expression of the Hexose Phosphate Transport System of Escherichia coli. *Journal of Bacteriology*, 155(3):1062–1070, January 1983.

[86] Srivatsan Raman, Jameson K. Rogers, Noah D. Taylor, and George M. Church. Evolution-guided optimization of biosynthetic pathways. *Proceedings of the National Academy of Sciences*, 111(50):17803–17808, December 2014.

[87] Anita Loeschcke and Stephan Thies. Pseudomonas putida—a versatile host for the production of natural products. *Applied Microbiology and Biotechnology*, 99(15):6197–6214, June 2015.

[88] Aditya M. Kunjapur and Kristala L. J. Prather. Microbial engineering for aldehyde synthesis. *Applied and Environmental Microbiology*, pages AEM.03319–14, January 2015.

[89] S. Martinenghi, G. Cusella De Angelis, S. Biressi, S. Amadio, F. Bifari, M. G. Roncarolo, C. Bordignon, and L. Falqui. Human insulin production and amelioration of diabetes in mice by electrotransfer-enhanced plasmid DNA gene transfer to the skeletal muscle. *Gene Therapy*, 9(21):1429–1437, November 2002.

[90] Ping Wang, Lydia Robert, James Pelletier, Wei Lien Dang, Francois Taddei, Andrew Wright, and Suckjoon Jun. Robust growth of Escherichia coli. *Current biology: CB*, 20(12):1099–1103, June 2010.

[91] Jonathan W. Young, James C. W. Locke, Alphan Altinok, Nitzan Rosenfeld, Tigran Bacarian, Peter S. Swain, Eric Mjolsness, and Michael B. Elowitz. Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nature Protocols*, 7(1):80–88, January 2012.

[92] Guillaume Baptist, Corinne Pinel, Caroline Ranquet, Jérôme Izard, Delphine Ropers, Hidde de Jong, and Johannes Geiselmann. A genome-wide screen for identifying all regulators of a target gene. *Nucleic Acids Research*, 41(17):e164, September 2013.

[93] Keith A. Martinez, Ryan D. Kitko, J. Patrick Mershon, Haley E. Adcox, Kotiba A. Malek, Melanie B. Berkmen, and Joan L. Slonczewski. Cytoplasmic pH response to acid stress in individual cells of Escherichia coli and Bacillus subtilis observed by fluorescence ratio imaging microscopy. *Applied and Environmental Microbiology*, 78(10):3706–3714, 2012.

[94] Shyam K Sharan, Lynn C Thomason, Sergey G Kuznetsov, and Donald L Court. Recombineering: a homologous recombination-based method of genetic engineering. *Nature protocols*, 4(2):206–23, January 2009.

[95] Minh Hoa Dao-Thi, Laurence Van Melderen, Erwin De Genst, Lieven Buts, An Ranquin, Lode Wyns, and Remy Loris. Crystallization of CcdB in complex with a GyrA fragment. *Acta crystallographica. Section D, Biological crystallography*, 60(Pt 6):1132–1134, 2004.

[96] 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. *Journal of Bacteriology*, 177(14):4121–4130, July 1995.

[97] Kirill A. Datsenko and Barry L. Wanner. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences*, 97(12):6640–6645, June 2000.

[98] Jack A. Bryant, Laura E. Sellars, Stephen J. W. Busby, and David J. Lee. Chromosome position effects on gene expression in Escherichia coli K-12. *Nucleic Acids Research*, 42(18):11383–11392, October 2014.

[99] Elisa Brambilla and Bianca Sclavi. Gene Regulation by H-NS as a Function of Growth Conditions Depends on Chromosomal Position in Escherichia coli. *G3: Genes/Genomes/Genetics*, 5(4):605–614, February 2015.

[100] Ingvild Flåtten, Solveig Fossum-Raunehaug, Riikka Taipale, Silje Martinsen, and Kirsten Skarstad. The DnaA Protein Is Not the Limiting Factor for Initiation of Replication in Escherichia coli. *PLoS Genetics*, 11(6), June 2015.

[101] Irina Mihalcescu, Mathilde Van-Melle Gateau, Bernard Chelli, Corinne Pinel, and Jean-Luc Ravanat. Green autofluorescence, a double edged monitoring tool for bacterial growth and activity in micro-plates. *Physical Biology*, 12(6):066016, December 2015.

[102] G Churchward, E Estiva, and H Bremer. Growth rate-dependent control of chromosome replication initiation in Escherichia coli. *Journal of Bacteriology*, 145(3):1232–1238, March 1981.

[103] H. E. Kubitschek and J. A. Friske. Determination of bacterial cell volume with the Coulter Counter. *Journal of Bacteriology*, 168(3):1466–1467, December 1986.

[104] Howard M. Salis, Ethan A. Mirsky, and Christopher A. Voigt. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, 27(10):946–950, October 2009.

[105] Amin Espah Borujeni, Anirudh S. Channarasappa, and Howard M. Salis. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Research*, 42(4):2646–2659, February 2014.

[106] Yuichi Taniguchi, Paul J. Choi, Gene-Wei Li, Huiyi Chen, Mohan Babu, Jeremy Hearn, Andrew Emili, and X. Sunney Xie. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science (New York, N.Y.)*, 329(5991):533–538, July 2010.

[107] Norunn K. Torheim, Erik Boye, Anders Løbner-Olesen, Trond Stokke, and Kirsten Skarstad. The Escherichia coli SeqA protein destabilizes mutant DnaA204 protein. *Molecular Microbiology*, 37(3):629–638, August 2000.

[108] D. W. Austin, M. S. Allen, J. M. McCollum, R. D. Dar, J. R. Wilgus, G. S. Sayler, N. F. Samatova, C. D. Cox, and M. L. Simpson. Gene network shaping of inherent noise spectra. *Nature*, 439(7076):608–611, February 2006.

[109] Lyle A. Simmons, James J. Foti, Susan E. Cohen, and Graham C. Walker. The SOS Regulatory Network. *EcoSal Plus*, 2008, July 2008.

[110] Qing Chen and Robert J. Kadner. Effect of altered spacing between uhpT promoter elements on transcription activation. *Journal of Bacteriology*, 182(16):4430–4436, 2000.

[111] Nir Friedman, Long Cai, and X. Sunney Xie. Linking Stochastic Dynamics to Population Distribution: An Analytical Framework of Gene Expression. *Physical Review Letters*, 97(16):168302, October 2006.

[112] Jakub Wiktor, Christian Lesterlin, David J. Sherratt, and Cees Dekker. CRISPR-mediated control of the bacterial initiation of replication. *Nucleic Acids Research*, 44(8):3801–3810, May 2016.