Cristiana Isabel Martins Ferreira
Licenciada em Ciências de Engenharia Biomédica

Study of the influence of σ factors on the kinetics of RNA production

Dissertação para obtenção do Grau de Mestre em
Engenharia Biomédica

Orientador: José Manuel Fonseca, Professor Auxiliar, FCT-UNL

Co-orientador: André Sanches Ribeiro, Professor Assistente, TUT, Finlândia

Júri:

Presidente: Professora Doutora Carla Maria Quintão Pereira

Arguente: Professora Doutora Maria João Gomes Trindade Caseiro

Vogal: Professor Doutor José Manuel Matos Ribeiro da Fonseca

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro 2014

Copyright

"Toute réussite déguise une abdication"

"Each success hides an abdication"

Simone de Beauvoir

Acknowledgements

First I would like to thank my supervisor Prof. Dr. José Manuel Fonseca, who proposed me to go to Finland to do the study presented in this Master Thesis, and always believed that I could do it, even when I thought to give up. He always encouraged me to never give up.

To Prof. Dr. André Sanches Ribeiro, who welcomed me in his group and made possible the development of this Master Thesis providing to me all the tools needed and taught me all the background needed.

To Huy Tran, who helped me a lot during my stay in Finland, as well as, after in Portugal. He was always available to explain everything to me and taught me how to use the tools needed for the analysis of the images and to get the results in the stochastic model here developed. Thank you for your help, patience and all the knowledge transmitted.

To Jason Lloyd-Price, who provided me the tools needed to do the processing and segmentation of the time series analyzed. Jason provided me the *Jason masking tool* that was used in all of the time series analyzed and the SGNSim Simulator where all the simulations of the stochastic models were done.

To Vinodh Kandavalli, who provided me the time series data to be analyzed. Without his work in the laboratory it would have been impossible to get the results. And I have to say thank you for all the good moments spent together during my stay in Finland.

To Nádia Gonçalves, who helped me even before I arrived in Finland and who was always present during my five months in Tampere. We spent great moments together that I will never forget, my friend.

To Samuel Matos for the good moments spent together and for all the help on biology subjects.

To everyone in the LBD group, who received me very well, who integrated me so well and who were always available for anything I needed.

To Hanaa Zaki, Carmen Rodríguez, Clémentine Brinon, Ana Domenti, Teresa Bustamante, João Silva and all the exchange students, for all the great moments spent together and specially for being my family in Finland. With you I felt that I was at home and I had the greatest experience in my life. I will never forget you!

To all my friends, from the high school or before, from my journey in Coimbra and from the last 2 years in Lisbon, I want to show my greatest gratitude for all the good times spent together and for being always there to me.

Finally I thank my family, especially my parents and my brother who gave me the opportunity to finish my studies and support my exchange period in Finland, gave me strength to overcome any problem and always supported me in my decisions.

Resumo

A dinâmica da expressão genética da *Escherichia coli* é controlada a nível da iniciação da transcrição, que tem início quando uma holoenzima de RNA polimerase, constituída pelo núcleo da enzima RNA polimerase e o factor σ, reconhece a sequência promotora de um gene. Sob diferentes condições, diferentes factores σ são usados. Para além disto, alguns promotores requerem um factor σ específico, enquanto outros têm menos especificidade. A concentração dos factores σ varia consoante o factor e a fase de crescimento celular.

Dado que a influência dos factores σ na cinética de iniciação da transcrição é desconhecida, pretende-se neste estudo caracterizar a sua dinâmica, em condições de crescimento óptimas, por diferentes factores σ sob controlo de dois promotores, P_{tetA} e P_{BAD}, nas fases exponencial e estacionária de crescimento celular. São utilizadas células mutantes, sem o factor σ^{54} ou sem o factor σ^{38}, e compara-se a dinâmica da iniciação da transcrição com a das células não mutantes, para os promotores nas duas fases de crescimento celular. As moléculas de RNA são detectadas logo que produzidas, através do método de marcação MS2-GFP e são obtidas as distribuições dos intervalos de tempo entre a produção consecutiva de moléculas de RNA.

Dos resultados obtidos conclui-se que: P_{tetA} não é afectado pela composição dos factores σ nas duas fases de crescimento celular em análise, enquanto P_{BAD} o é; a dinâmica da iniciação da transcrição é influenciada pelo promotor usado; existem 3 passos limitantes na iniciação da transcrição sob controlo dos dois promotores para as 3 estirpes durante as fases de crescimento em análise; as distribuições dos intervalos obtidos não são do tipo exponencial; a produção de RNA é sub-Poissonian; os resultados do modelo estão de acordo com as medidas *in vivo* para P_{tetA} enquanto para P_{BAD} existem algumas diferenças.

Palavras-chave: factor σ, iniciação da transcrição, cinética da produção de RNA, simulação estocástica.

Abstract

Gene expression dynamics in *Escherichia coli* is controlled at the transcription initiation stage, which begins when an RNA polymerase holoenzyme, composed of RNA polymerase core enzyme and the σ factor, recognizes the promoter sequence of a gene. Under different conditions, different σ factors are used. Also, some promoters require a specific σ factor, while others have less specificity. The intracellular levels of σ factors vary between σ factors and with the cell growth phase.

It is still unknown whether different σ factors will lead to differing kinetics of transcription initiation, thereby, in this study it will be characterized the dynamics of this process by different σ factors, under optimal growth conditions and under the control of either of two promoters, P_{tetA} and P_{BAD}, during the exponential and stationary growth phases. Mutant cells, lacking σ^{54} or σ^{38} were used and the dynamics of transcription initiation was compared with wild-type cells, for each of the two promoters and during each of the two growth phases. For this, RNA molecules are detected as soon as they are produced in each cell, using an MS2-GFP tagging method, and the distribution of time intervals between consecutive RNA productions are obtained in each condition.

From the results obtained it is concluded that: P_{tetA} is not affected by the σ factors' population composition during the two growth cellular phases studied, while P_{BAD} it is ; the dynamics of transcription initiation is affected by the promoter used; there are 3 rate-limiting steps in transcription initiation under control of the two promoters for the 3 strains during the phases analyzed; the distribution of the intervals are not exponential-like; RNA production is sub-Poissonian; the results of the model developed are in agreement with the observations from *in vivo* measurements under control of P_{tetA}, while for P_{BAD} there are some differences.

Keywords: σ factor, transcription initiation, kinetics of RNA production, stochastic simulation.

Contents

Figure contents

Table of contents

Abbreviations and symbols

μ(s)	Mean duration in seconds
σ	Sigma factor
σ(s)	Standard deviation in seconds
$\alpha_2\beta\beta'\omega$	Catalytic Subunits of RNA polymerase core enzyme
aTc	Anhydrotetracycline
ATP	Adenosine triphosphate
c_j	Reaction probability per time unit
CME	Chemical Master Equation
CVS	Coefficient of variance over mean square
d	Number of steps
DNA	Deoxyribonucleic Acid
E	Amount of RNA polymerase that is free
Eσ	Amount of free holoenzyme
$E\sigma^{70}_b$	Quantity of holoenzyme that bind nonspecifically to DNA
$E\sigma^{38}_b$	
$E\sigma^{54}_b$	
$E\sigma^{70}$	Quantity of holoenzyme that bind specifically to DNA in the promoter region
$E\sigma^{38}$	
$E\sigma^{54}$	
E_b	Amount of nonspecifically binding of E to DNA
E. coli	*Escherichia coli*
In vitro	Latin for "within glass"
In vivo	Latin for "within the living"
k_{cc}	Rate constant of the formation of the closed complex
k_{oc}	Rate constant of the formation of the open complex between RNAp and the DNA
k_p	Translation rate constant
K_{S_X}	Disassociation constant of the specific binding of the holoenzyme to the promoter
K_X	Disassociation constant between RNAp core enzyme and the sigma factor
LBD	Laboratory of Biosystem Dynamics

mRNA	Messenger Ribonucleic Acid
MS2-GFP	MS2 coat protein fused with Green Fluorescence Protein (GFP)
N	Number of existing species in the solution
$n_{\sigma^{70}}$	Total number of sigma factors
$n_{\sigma^{38}}$	
$n_{\sigma^{54}}$	
$n_{E\sigma}$	Total number of holoenzyme
P	Protein
Pr	Promoter region
Pr_c	Closed complex of RNA polymerase and the DNA
Pr_{o_X}	Open complex formation with respective sigma factor associated
P_{tetA}	Reporter gene – *tet* operon
P_{Lac}	Reporter gene – Lactose operon
r	Ribosomes
RNA	Ribonucleic Acid
RNAp	Ribonucleic Acid Polymerase
RNApσ	RNA polymerase holoenzyme
SSA	Stochastic Simulation Algorithm
T	Temperature
UV	Ultraviolet
V	Constant volume

1. Introduction

Every living organism ensures its survival by following the cell dogma (Crick, 1970), which is scripted in the chromosome inherited from its parent(s) and passed down to its children. Following this dogma, the pieces of information stored in the chromosome, known as genes, specific sequences of nucleotides encoded in the deoxyribonucleic acid (DNA), are expressed through two complex processes. The first process, known as transcription, is a crucial step in gene expression and its regulation. The second process is translation. In transcription, an enzyme called RNA polymerase (RNAp) reads the DNA and makes a complementary messenger RNA strand (mRNA). Upon released, mRNA is either modified or immediately translated by ribosome to create proteins, which are the functional units in cells (Alberts et al., 2008).

Gene expression has a stochastic nature, which causes cell to cell variability in the number of RNA and protein molecules in cells of a genetically identical population (Süel, Garcia-Ojalvo, Liberman, & Elowitz, 2006). Single-cell experiments have shown that there are fluctuations in rate of production of RNA and proteins over time (Elowitz, Levine, Siggia, & Swain, 2002).

The bacterium *Escherichia coli (E. coli)* has been used as a model organism in studies related to transcription. *In vitro* studies established that transcription initiation is a multi-step process (Buc & McClure, 1985) which takes a long time and plays an important role in determining the mean and noise in mRNA. The duration of each step vary between promoters (Lutz, Lozinski, Ellinger, & Bujard, 2001), also with temperature (Buc & McClure, 1985) and with the concentration of Mg^{2+} and others metabolites (Suh, Leirmo, Record, & Jr., 1992), among other reasons. It is known that transcription is a stochastic process (H H McAdams & Arkin, 1999) and recent *in vivo* studies showed that it is a sub-Poissonian process (Kandhavelu et al., 2011) under weak and medium induction levels. The same studies revealed that transcription initiation *in vivo* has at least two elementary steps.

Recent studies also recognized that in transcription, there is a sensoring factor that enables specific binding of RNA polymerase to gene promoters. This factor, σ (sigma) factor, is a single subunit of the transcription machinery of *E. coli* that acts as a sensor guiding RNA polymerase to specific binding sites on promoters (Gruber & Gross, 2003). It is known that, under different conditions, different σ factors are used in transcription. Also, while some promoters appear to require a specific σ factor to initiate transcription, others have less σ factor specificity (Gruber & Gross, 2003). Additionally, some σ factors transcribe genes expressed during the exponential growth phase while others carried out the transcription of

genes expressed during stationary growth phase (Loewen & Hengge-aronis, 1994). So far, it remains unclear how different σ factors affect the transcription dynamics, particularly the kinetics of the multi-step transcription initiation process. Therefore, it would be of interest to study how the dynamics of transcription differ between mutant cells (lacking one specific σ factor) and wild-type cells (containing all σ factors), when under the same, optimal growth conditions. This test will be performed here in two promoters, P_{tetA} and P_{BAD}, in order to infer if existing differences in RNA production kinetics between mutant cells are solely σ factor-dependent or are also promoter-dependent. Because the number of σ factors is cell phase-dependent, this test will be performed here for two growth phases of *E. coli*, more specifically it will be compared the results from wild-type cells and mutant cells during the exponential phase with the ones from the same strains during the stationary phase, for both promoters used.

In order to study the role of σ factors on the dynamics of transcription, a data analysis of the measurements was performed and a stochastic model of σ factors was developed. The model simulates the biological processes at the single event level using a modified version of the Stochastic Simulation Algorithm (SSA) (Gillespie, 1977) that allows delays in reaction events and aims at the prediction of the statistics of transcription, which would not be feasible using deterministic kinetics. As mentioned before, the transcription by the RNAp takes some time, thereby the model developed from a delayed stochastic model of transcription that contains time delays in reaction events (A. Ribeiro, Zhu, & Kauffman, 2006). The duration of transcription initiation, in this model, is modelled following a Gaussian distribution, to take into account the rate-limiting steps inherent to this process (McClure, 1980). The model developed also includes explicitly the steps of transcription initiation, as the formation of the closed complex and its isomerization, which leads to the open complex formation (Buc & McClure, 1985) and the elongation process. It contains also the reactions of the translation process, like the formation of proteins and the time needed for that.

Once the different σ factors in study are included in our model, it is possible to simulate all the different strains of *E. coli* as well as the two different growth phases in study, changing the values of parameters like the intracellular level of the holoenzymes and dissociation constants. The results of these simulations are compared with the results from the *in vivo* measurements. For the simulations under control of P_{tetA} the results are in agreement with the results of the *in vivo* measurements. On the other hand, for P_{BAD} although the distribution of RNA production follows the same trend, the values of the mean production intervals are different from the *in vivo* results.

This study is derived from a study conducted in the Laboratory of Biosystem Dynamics (LBD), Department of Signal Processing, Tampere University of Technology, Finland. Here, the author acquired knowledge about the models of gene expression and σ factors, and became familiar with the simulation and data analysis tools. We performed the data analysis from the measurements and the author assisted in the development of the first stochastic model of transcription that includes σ factors.

The results should offer insights on the prospective of modifying the array of σ factors in *E. coli* mutants, whether to restrict the environments it can live in, or to expand them to more extreme conditions, beneficial to the synthetic biology field or the pharmaceutical industry.

This thesis contains 5 Chapters besides the first, introductory one. Chapter 2 gives an introduction of the main topics of this thesis: gene expression, the role of σ in transcription, the models of gene expression dynamics in *E. coli* and the *in vivo* measurements of time intervals between consecutive RNA productions. In chapter 3 we present all the methodology used in this study, such as, the use of fluorescence probes to obtain bright fluorescent spots, how we do the measurements with the microscope, how we analyse the images obtained in the microscope and how we extract the results. Chapter 4 contains the results and their discussion, of the *in vivo* measurements in wild-type cells, in mutant cells lacking σ^{54} and in mutant cell lacking σ^{38} during exponential phase and during stationary phase under control of P_{tetA} or under control of P_{BAD}. This chapter also contains the results obtained when we fitted the model with the empirical results. The discussion of the results of the simulations made for the three strains under control of the two promoters during the two growth phases are also presented in this chapter. The conclusion of this work is presented in Chapter 5 as well as the perspectives on future developments. Finally, Chapter 6 contains all the references used in this work.

2. State of the art

In this chapter we present a theoretical description of the main concepts needed to understand the study performed in this thesis. First, we give insights on gene expression. Second, we describe in greater detail the first step of this process, transcription, in *E. coli* and the role of σ factors in this process. Then, we explain how the gene expression dynamics in *E. coli* is modeled and how the *in vivo* measurements of time intervals between consecutive RNA productions are made. These last two subchapters are important since they describe how to obtain the results with the analysis performed in the present study.

2.1. Gene expression

All living cells contain their genetic instructions stored in genes, which are specific sequences of nucleotides encoded in DNA. These instructions are copied and transmitted from mother to daughter cells. The flow of these genetic instructions within a cell is explained by the central dogma of molecular biology (Crick, 1970), which states that "DNA (Deoxyribonucleic Acid) makes RNA (Ribonucleic Acid) which makes protein". The molecule of DNA contains the genetic code which is inherited from the mother cell. The process by which the genetic code is used by cells to direct protein synthesis is denominated by gene expression. This process consists of two main steps: transcription and translation.

Prokaryotes transcription occurs in three phases known as initiation, elongation and termination. In transcription initiation, transcription factors bind to RNA polymerase (RNAp) allowing the RNAp to be tightly bound in the promoter region in DNA. Once the RNAp is attached to the DNA strain, a small portion of the DNA double helix is opened and unwound, in order to expose the bases on each DNA strand. Only one strand of DNA is used as a template at any one time for the synthesis of a messenger RNA (mRNA) molecule. The nucleotide sequence of the RNA chain is determined like in DNA replication, which means determined by the complementary base-pairing between incoming nucleotides and the DNA template. The incoming ribonucleotides are covalently linked to the growing RNA chain when there is complementarity of its bases. When the RNAp binds the terminator region, transcription is over and it is released the DNA template and the completed messenger RNA (mRNA) molecule. Upon released, mRNA is either modified or immediately translated by ribosome to create proteins, which are the functional units in cells (Alberts et al., 2008).

2.2. The role of σ factors in transcription

RNA polymerase (RNAp), the central enzyme of gene expression in bacteria, consists of five protein subunits, two α subunits, together with single copies of the two largest (β and β') and the smallest (ω) subunits, and it is responsible for the polymerization or synthesis of RNA. This core enzyme (RNAp) is able to copy DNA into RNA but transcription is not initiated at the correct site in a gene, which means that it does not recognize the promoter region. It is required that the core enzyme binds to the transcription initiation factor, σ factor, which is a single regulatory subunit that recognizes the signal on the DNA strand and indicates that the RNA polymerase should initiate the synthesis of the RNA. When the σ factor binds the core enzyme, it is formed the RNA polymerase holoenzyme (RNApσ) (reaction 2.1), increasing the affinity to various promoters and decreasing the affinity of the RNAp for nonspecific DNA. It is known that, when reaction 2.1 occurs, the σ factor provides most of the determinants for promoter recognition and DNA melting (Gruber & Gross, 2003).

$$RNAp + \sigma \leftrightarrow RNAp\sigma \qquad (2.1)$$

Prokaryotes transcription (Figure 2.1) occurs in three phases: initiation, elongation and termination. Transcription initiation involves a reversible binding of RNAp holoenzyme to a special DNA sequence at the beginning of the gene, known as the promoter region (Pr). This step is referred as closed complex formation (Pr_c) because the DNA is not unwound (reaction 2.2) (Figure 2.1 – step 1). The closed complex is a relatively weak, unstable formation.

$$RNAp\sigma + Pr \rightarrow Pr_c \qquad (2.2)$$

The next step is the unwound of approximately 10 bases of DNA around the initiation site in order to form an open complex (Figure 2.1 – step 2), much stronger than the closed complex, in which one strand of DNA is a template for transcription. When few nucleotides are added, the σ factor is released stochastically from the RNA polymerase (Figure 2.1 – step 3), which then leaves the promoter and moves along the template strand of DNA to continue the elongation of the growing RNA chain (Figure 2.1 – step 4, 5 and 6). It was found that sometimes the σ factor remains on RNA polymerase until termination, which can be used as an elongation regulator (Mooney, Darst, & Landick, 2005).

During elongation, the RNAp unwinds the template strand of DNA ahead of it and rewinds the DNA behind it, maintaining an unwound region in the region of transcription. The synthesis

of the RNA continues until the RNAp encounters a termination signal, at which point transcription stops and the messenger RNA (mRNA) is released from the RNAp and this enzyme is dissociated from its DNA strand template (reaction 2.3) (Figure 2.1 – step 7).

$$Pr_c \rightarrow Pr + RNAp + \sigma + mRNA \qquad (2.3)$$

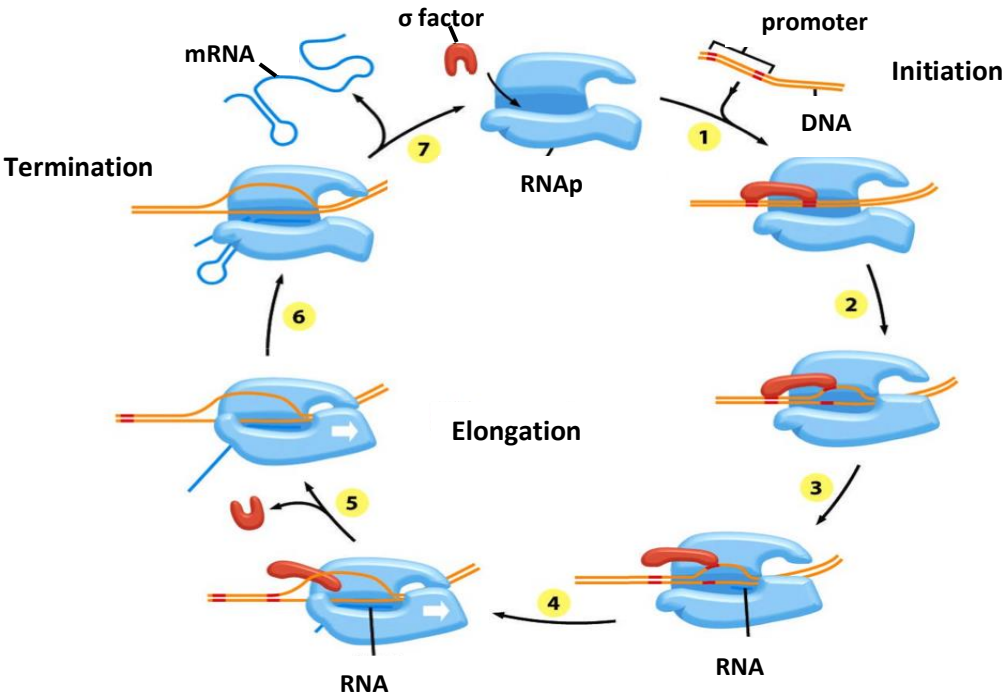

Figure 2.1: Scheme of the three phases of transcription in Escherichia coli. The σ factor binds to the RNA polymerase, forming the holoenzyme. When this holoenzyme finds and binds the promoter region, it initiates transcription (1) and a closed complex is formed. The next step is the formation of an open complex (2) where one of the two strands will act as a template for complementary base pairing with the incoming ribonucleotides. This reaction proceeds and after about 10 nucleotides of RNA synthesis, the RNAp breaks its interaction with the promoter region and the σ factor is dissociated (3). This dissociation increases the affinity of RNA for DNA which make RNAp highly processive, moving along the DNA strand and synthesizing RNA. This step is called elongation (4,5,6). When the RNAp finds the termination site, the newly formed messenger RNA molecule and the RNAp dissociate (7). This phase, known as termination, allows the mRNA release.[1]

In *Escherichia coli,* there are seven different σ subunits that can participate in the transcription of a specific set of genes (Ishihama, 2000). These σ factors, that possess different promoter-recognition properties, can be generally divided into two groups: the σ^{70} family of σ factors and the σ^{54} group. The first group includes σ factors that share structural similarities and the ones in the second group have differences in sequence, promoter architecture and

function from the σ^{70} family (Wösten, 1998). Despite the overall similarity in their structures, the σ^{70} family can recognize different classes of promoters (Gruber & Gross, 2003). Some promoters can only be activated by a specific σ factor, while others can be activated by any σ factor. Similarly, some σ factors appear to become activated at specific temperature ranges. There is also anti-σ factors that inhibit the function of the σ factors (Ishihama, 2000). Under stress, its repertoire of σ factor is altered along with the cells' transcriptional program (Gruber & Gross, 2003). It is known that in response to growth transitions and environmental conditions there is changes in the intracellular levels of each individual σ factor (Jishage, Iwata, Ueda, & Ishihama, 1996).

E. coli contains six σ factors of the σ^{70} family, σ^{70} (encoded by RpoD), σ^{38} (RpoS), σ^{32} (RpoH), σ^{28} (RpoF), σ^{24} (RpoE) and σ^{FecI}, each participating in the transcription of a specific set of genes (Ishihama, 2000). σ^{70} is the house-keeping σ factor that can transcribe most genes expressed during the exponential phase. The intracellular concentration of σ^{70} subunit remains at a constant level in the transition from the exponential growth phase to the stationary phase, although the levels of core enzyme subunits decrease concomitantly with the stopping of cell growth. σ^{38} is the master regulator of the general stress response, transcribing more than 70 genes that confer resistance against such diverse insults as oxidative stress, UV-radiation, heat shock, hyperosmolarity, acidic pH and ethanol. Due to the generality of the response, σ^{38} plays both a preventative and a combative role (Gruber & Gross, 2003). The level of σ^{38} increases when the cell enters the stationary growth phase and plays an important role in the stress response during these translation to that phase (Jishage & Ishihama, 1995). σ^{32} and σ^{24}can transcribe the heat shock genes. σ^{28} is involved in transcription of flagellar formation and chemotaxis genes. σ^{FecI} is used in the ferric citrate transport system and has extracytoplasmic functions.

σ^{54} transcribe genes which are activated by a deficiency of nitrogen (Merrick, 1993) and some other stress response genes (Shingler, 1996). It is known that the amount of σ^{54} present in the cell is approximately one tenth of the amount of σ^{70}, during exponential and stationary phase growth (Jishage et al., 1996). Some differences between this group of σ factors and the σ^{70} family are: σ^{54} is able to bind promoter DNA even in the absence of core RNA polymerase and σ^{54} requires an additional ATP-dependent activation event provided by transcriptional activators before they initiate transcription.

2.3. Models of gene expression dynamics in *E. coli*

The dynamics of the two main steps of gene expression, transcription and translation, has two main properties. First, these are both stochastic processes (H H McAdams & Arkin, 1999). Second, both of these processes are multi-stepped (Buc & McClure, 1985). That is, they possess more than one event that is 'rate-limiting' and thereby affects the durations of the intervals between consecutive productions of the product molecules. The modelling strategy here presented was first proposed in (A. Ribeiro et al., 2006), and aims to capture both of these features. For that, the simulation of these models is performed using the 'Stochastic Simulation Algorithm' (SSA) (Gillespie, 1977) which is a Monte Carlo method that simulates numerically the time evolution of well stirred reaction systems. The time goes forward in discrete steps. A reaction is explicitly executed in each step and the effect on the number of each molecule is settled. Since these models do not explicitly include σ factors and due to the aim of this thesis, it is required the development of a new model containing the σ factors in study of *E. coli*. With that, it is possible to study its influence on the kinetics of RNA production. The model of σ factors developed in this thesis as well as its results, are presented in chapter 4, in section 4.3..

Transcription is usually modelled as a 2-step process (reaction 2.4). The first step is named *closed complex formation* and consists on the finding of the transcription start site by an RNA polymerase (here modelled with constant rate k_{cc})(Figure 2.1). The second step is named *open complex formation* and consists on the formation of the open complex between the RNA polymerase and the DNA (here modelled with rate constant k_{oc})(Figure 2.1). Since these are the two major rate-limiting steps of transcription under optimal conditions (Buc & McClure, 1985), in general, the outcome of this step includes not only a free promoter region and RNAp but also a complete messenger RNA molecule (Figure 2.1):

$$Pr + RNAp \xrightarrow{k_{cc}} PrRNAp \xrightarrow{k_{oc}} Pr + RNAp + mRNA \qquad (2.4)$$

As the gene exists at single-copy level, each transcription event only produces one molecule of messenger RNA.

$$mRNA + r \xrightarrow{k_P} mRNA + r + P \qquad (2.5)$$

Reaction 2.5 is used to model, as a single-step process, the translation of mRNA by ribosomes (r) in order to produce proteins (P), where k_p is the translation rate constant. The values of the reaction rates vary between promoters and determine the dynamics of RNA production.

Studies have shown that the lifespan of mRNA is usually limited to minutes, while proteins have a much longer half-life (from several minutes to hours). The lifespan of mRNA can be fitted with an exponential distribution (Bernstein, Khodursky, Lin, Lin-Chao, & Cohen, 2002). The concentration of the proteins decreases due to cell elongation and division. The decaying process of these molecules can be modelled by the first order reactions listed below (reactions 2.6 and 2.7) (Greive & von Hippel, 2005).

$$mRNA \xrightarrow{d_M} 0 \qquad (2.6)$$

$$P \xrightarrow{d_P} 0 \qquad (2.7)$$

2.4. *In vivo* measurements of time intervals between consecutive RNA productions

Due to the development of new techniques for tagging RNA molecules with MS2 coat protein fused with Green Fluorescence Protein (GFP), MS2-GFP proteins (Peabody, 1993)(Peabody, 1997) (Fusco et al., 2003)(Golding & Cox, 2004) and using time-lapse fluorescence microscopy, it is possible to detect RNA molecules soon after completion or even while elongating (Golding & Cox, 2004), once the GFP produces a green light when a source of UV hits the protein and a green spot appears when a molecule of RNA is formed.

Before these new techniques appeared, the knowledge of RNA transcription and its dynamics came from population studies or *in vitro* studies with purified components ((Harada et al., 1999) (Shaevitz, Abbondanzieri, Landick, & Block, 2003)(Skinner, Baumann, Quinn, Molloy, & Hoggett, 2004)). However, once these studies were not a cell-to-cell study, it was difficult to understand some processes which occurs at a single cell level. Thus, in our study the technique developed by Golding et al. (Golding, Paulsson, Zawilski, & Cox, 2005) will be used. With this technique, the cell-to-cell diversity in RNA numbers of a population at a given moment in time was firstly quantified (Golding et al., 2005).

The *in vivo* kinetics of RNA production at single cell level can also be measured with this technique. Once each RNA molecule is tagged during the elongation process or shortly after

(Golding et al., 2005) if the time when the first RNA molecules appear is registered, it is possible to measure the time interval between consecutive transcription events and calculate the mean duration as well as the variability of these intervals (Kandhavelu et al., 2011, 2012; Muthukrishnan et al., 2012). It is suspected that the cell-to-cell diversity in RNA and protein numbers in populations of sister cells (Kandhavelu et al., 2011; Paulsson, 2004; A. Ribeiro et al., 2006) comes from the expected noise of the underlying chemical processes in gene expression (Peccoud & Ycart, 1995). If the noise in transcription is estimated by measuring the intervals between transcription events (Kandhavelu et al., 2011) instead of by measuring cell-to-cell diversity in RNA numbers, the results are more reliable once that in the latter method the noise is influenced by several phenomena other than transcription, like errors in RNA partitioning in cell division or noise in RNA degradation (Huh & Paulsson, 2011a, 2011b; Lloyd-Price, Gupta, & Ribeiro, 2012).

3. Methods and Materials

In this chapter the methodology used in our experiments is explained. First the laboratory procedure will be introduced, explaining how the fluorescent cells are obtained and how the microscope is set to allow obtaining the images for analysis using custom software written in Matlab. The results of this analysis will be compared with the ones from the model developed in the aim of this study. The modeling and the simulating strategies used for the model here developed are explained in the last two subchapters of this chapter.

3.1. Use of fluorescent probes

The mRNA detection system contains two elements: a reporter gene on a medium copy plasmid and a target gene on a single-copy-F-plasmid. In figure 3.1 the mRNA detection system used in our study is described. The reporter gene (P_{Lac}) codes for a fluorescence protein, GFP, fused to a dimmer of the RNA bacteriophage MS2 coat protein (MS2d). The target gene codes for the target RNA, which contains several MS2-binding sites (Golding et al., 2005). As mentioned before, we will used different promoters to analyse its influence in RNA production and due to this, we represent the promoter in Figure 3.1 with P_X, where X can be tetA or BAD.

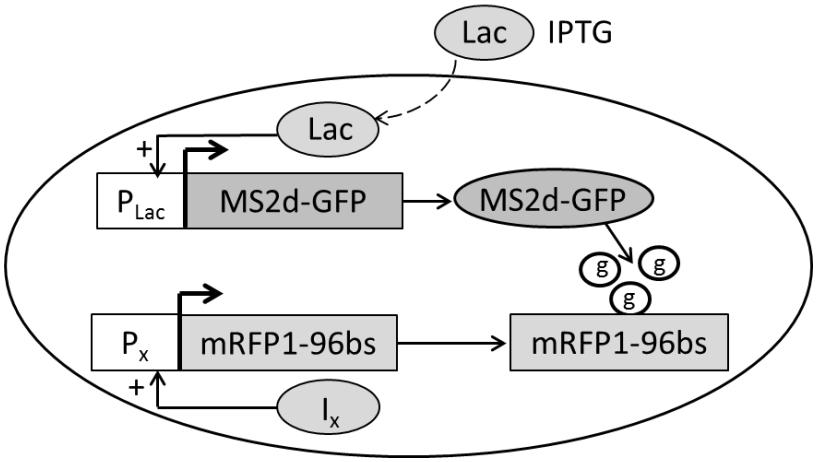

Figure 3.1: Measurement system. Components of the detection system. The reporter gene (P_{Lac}) controls the expression of the tagging protein (MS2d-GFP) and is inducible by Lac. The target construct is on a single-copy F-plasmid and its expression is controlled by the promoter P_X whose activity is regulated by specific inducer (I_X). The spots (represented with 'g' in the figure) appear when MS2d-GFP bind to a newly transcribed RNA (Mäkelä et al., 2013).

A bright fluorescent spot (Figure 3.1, represented by 'g') appears in the cell when multiple MS2d-GFP fusion proteins bind to a newly transcribed RNA.

For our measurements, cells with both MS2d-GFP and transcript target plasmids will be grown overnight in Miller LB medium at 37°C with aeration, diluted into fresh medium to maintain exponential growth until reaching an optical density of OD600 ≈ 0.5 for both promoters during exponential phase and an optical density of OD600 ≈ 0.4 for both promoters during stationary phase and supplemented by antibiotics according to the specific plasmids. The reporter plasmids will be induced with IPTG (1.0 mM%) during 45 minutes at 37 °C. The target P_{tet}, single-copy-F-plasmid, will be induced with anhydrotetracycline (aTc) (15 ng) during 5 minutes at the same temperature (37 °C). The target P_{BAD}, also a single-copy-F-plasmid, will be induced with arabinose (Ara) (0.2%) during 5 minutes at 37 °C.

3.2. Microscopy measurements

Microscopy measurements are done as in (Kandhavelu et al., 2011). After the induction of target RNA, the cells are placed on a microscopic slide between a cover slip and 3% LB-agarose gel pad set, and visualized by fluorescence microscopy, using a Nikon Eclipse inverted C1 confocal laser-scanning system with a 100x Apo TIRF objective.

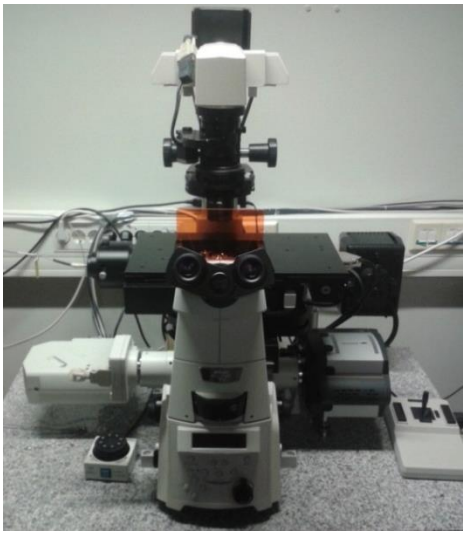

Figure 3.2: Nikon Eclipse (Ti-E, Nikon, Japan) inverted microscope with a 100x Apo TIRF objective (1.49 NA, oil). MS2-GFP fluorescence was measured with this microscope by a C2 confocal laser-scanning system with a 488 nm laser (Melles-Griot) and a 515/30 nm detection filter, using a pixel dwell of 2.4 µs and a resolution of 1024x1024 pixels. Phase contrast images are captured with a 2560x1920 pixel resolution CCD camera (DS-Fi2, Nikon, Tokyo, Japan). The software used for image acquisition is Nikon NIS-Elements.

In order to measure GFP fluorescence, we use a C2 confocal laser-scanning system with 488 nm laser and a 515/30 nm detection filter. For each slide, images of cells are captured using C1 with Nikon software EZ-C2 (Figure 3.2), approximately 5 min after induction, one image each 30 seconds, for approximately 2 hours. Microscopy will be performed in a temperature chamber to maintain the temperature constant.

3.3. Image analysis

Once the images of the cells are acquired by confocal microscopy (Figure 3.3), we will perform their analysis as in (Kandhavelu et al., 2012). In particular, the image analysis will be done semi-automatically using custom software written in Matlab.

The methods used follow (Kandhavelu et al., 2011). However, although all the similarities with the methods used in (Kandhavelu et al., 2011), the masking process is different. Each image is divided in three classes: background, cell border and cell region. Clumped cells are identified based on their size and edge information using an iterative cell segmentation process. A threshold is defined based on cell size and cells whose size goes beyond this threshold are discarded. RNA spots are segmented using a Kernel Density Estimation method (Ruusuvuori et al., 2010).

Figure 3.3 represents an example of the three stages of the processing. It is chosen one region of one original image (Figure 3.3 – A) taken with the microscope of our laboratory (Figure 3.2) and it is shown this region during the masking process (Figure 3.3 – B) and during the spot detection (Figure 3.3 – C). During the masking process (Figure 3.3 – B) it is necessary to do manual corrections of the masking made automatically. Only when all the frames of the time series are well masked, the spot detection is made (Figure 3.3 – C). After this processing, the individual cells are shown in blue and the spots are shown in green. Then, the number of RNA molecules in each spot is quantified by normalizing the MS2d-GFP-RNA spot intensity distribution, which means, dividing a spot's intensity by the intensity of the first peak in the histogram of spot intensities (Golding et al., 2005).

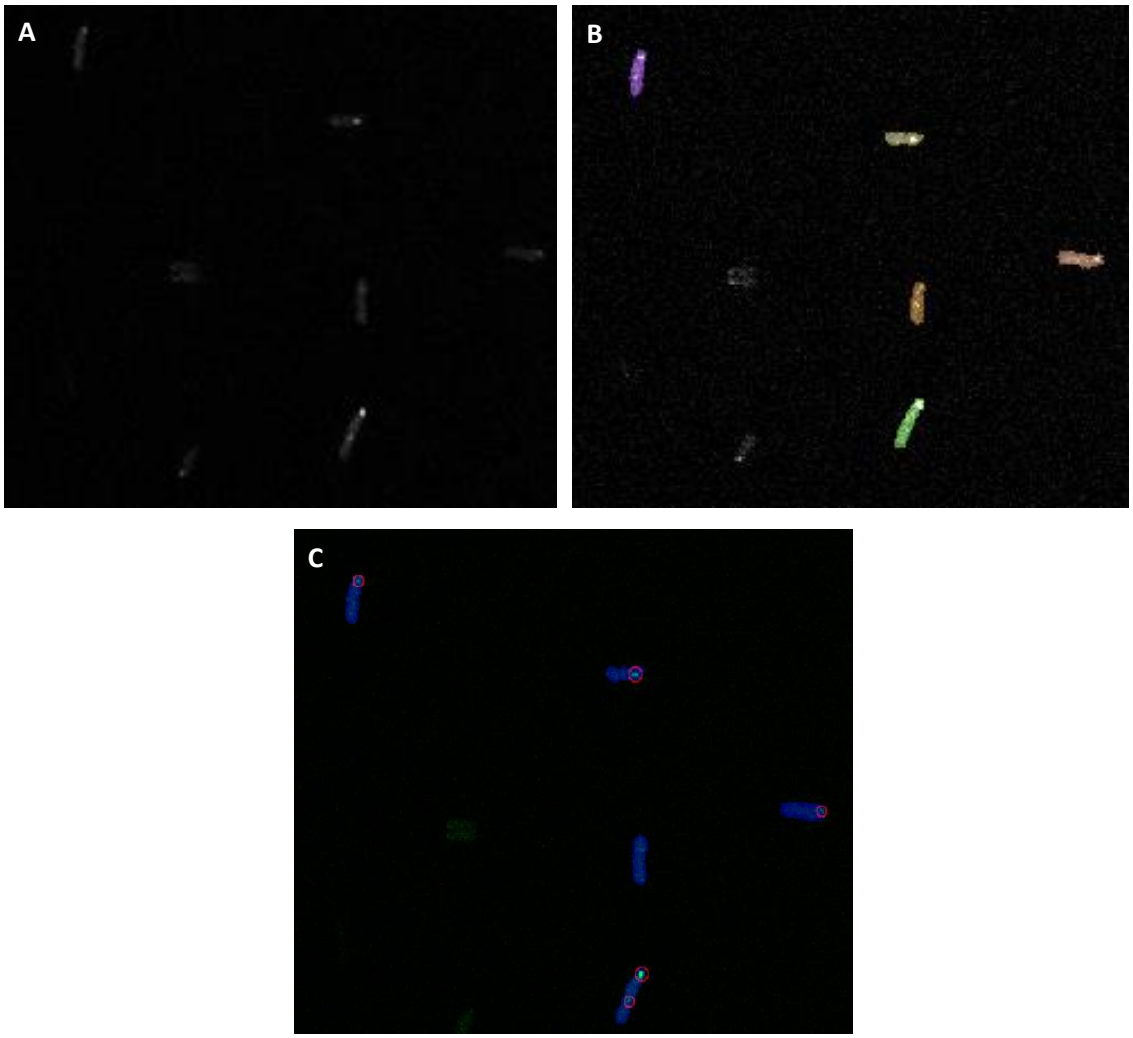

Figure 3.3: Part of an image taken by confocal microscope of MS2-GFP tagged molecules in E. coli cells at different stages of its processing. (A) Original image of some cells. (B) The same cells during the masking process. (C) After the masking, the result of the spot detection. Here the spots are circled with a red line and appear green in the blue cells.

The masking process (Figure 3.3) is made for each individual image independently, for a time-series measurement. The overall distribution of spot intensity is generated and obtained from all cells at each time point. This allows obtaining the number of RNA molecules in each cell at each time point. It is possible to determine when a new RNA appears and the time between the appearance of consecutive RNA molecules in individual cells when counting the number of RNAs in each cell at each moment. From that, the distributions of intervals between consecutive transcription events in a cell population subject to the same level of induction and temperature can be generated (Kandhavelu et al., 2011).

3.4. Modelling strategies

The chemical reactions used to represent elementary biological processes in cells could be mathematically described using models, enabling the simulation and inspection of the chemical system dynamics. These models can be either deterministic or stochastic.

The deterministic model is particularly useful when assessing processes with a population approach. However, it does not work properly when modelling gene expression due to the low copy number of RNA molecules (Taniguchi et al., 2010) and the stochastic nature of chemical reactions (Harley H. McAdams & Arkin, 1999) involved in this process. In this way, we employ a stochastic model of gene expression. Stochasticity in gene expression affects the functioning of cells and organisms and contributes to the phenotypic diversity in a genetically homogeneous population (Harley H. McAdams & Arkin, 1997; Ozbudak, Thattai, Kurtser, Grossman, & van Oudenaarden, 2002; Samoilov, Arkin, & Ross, 2002).

The kinetics of molecules in a solution is studied under the following assumptions:

- the system is well-stirred and of constant volume V, which requires that the spatial distribution of all species' molecules is uniform within the volume V and the position of molecules is independent of each other;
- the system is in thermal equilibrium at constant temperature T, which means that every molecule in the solution moves independently;
- reactions occur only when two or more molecules collide, while most collisions do not lead to reactions.

The dynamics of a solution with N species from X_1 to X_N and M reactions can be inspected by the analysis of the j^{th} reaction, described as follow:

$$\sum_{i=1}^{N} s_{ij} X_i \xrightarrow{c_j} \sum_{i=1}^{N} r_{ij} X_i \qquad (3.1)$$

In this reaction, the reaction constant c_j is the "reaction probability per time unit" and indicates how likely the reaction j^{th} is to happen given the reactants' molecule number at a given time. s_{ij} and r_{ij} indicates how many molecules of the substance X_i are consumed/produced via the j^{th} reaction. For stochastic models there is the propensity function (equation 3.2) which is equivalent to the rate equation of the deterministic models.

$$a_j(x) = c_j \cdot h(x) \qquad (3.2)$$

$a_j(x)$ dt specifies the probability for the j^{th} reaction to occur in the infinitesimal time window [t, t + dt). $h(x)$ indicates the number of possible reactant combinations of a reaction at a given time, with x the reactants' molecule number vector. Once the propensity function at a

specific time depends only on the current state, the system dynamics can be considered as a Markov process, where each reaction marks a change in state. Using the stochastic model, we can calculate the time for the change in the state to occur and the possible next state.

3.5. Simulating strategies

In order to find the solution for the stochastic models, the first order Chemical Master Equation (CME) (Gillespie, 1977), a mathematical method used to describe the time-evolution of probability density at fixed rates was originally employed. The most serious issues of CME are when it is applied to complex systems involving a large number of substances and when the probability densities are calculated on the continuous time scale. In those cases, the solutions are usually intractable. To address these problems, the Stochastic Simulation Algorithm (SSA) (Gillespie, 1977) was developed. The SSA is a Monte Carlo method that simulates numerically the time evolution of well stirred reaction systems. Time goes forward in each discrete step where a reaction is explicitly executed and the effect on the number of each molecule is settled. The time of the next reaction is determined using probability distributions. SSA takes into account the fact that the time evolution of a spatially homogeneous chemical system is a stochastic process. This algorithm numerically simulates the Markov process using random sampling.

Equation (3.3) is the basis for the realization of SSA.

$$P_{\tau,j}(\tau, j; x, t) = a'_j(x)a(x)\exp(-a(x)\tau) \qquad (3.3)$$

In equation (3.3) $a'_j(x)$ is the normalized flux of the reaction, j^{th} and indicates how likely the reaction is to occur and $a(x)\exp(-a(x)\tau)$ represents the exponential distribution of the probability of one reaction occurs at time t+τ. The time τ for the next reaction to occur can be calculated for any state *x* in the system's state space, if the distribution $a(x)\exp(-a(x)\tau)$ is inverted as in equation (3.4). Two uniform random numbers r_1 and r_2 are used to do the inverse transformation. *j* can be obtained from equation (3.5).

$$\tau = -a(x)^{-1}\ln(1 - r_1) \qquad (3.4)$$

$j = j'$ such that

$$\sum_{i=1}^{j'-1} a_i'(x) \leq r_2 < \sum_{i=1}^{j'} a_i'(x) \qquad (3.5)$$

SSA consists of the following steps:

1. Initialize the step n=0 with time $t_n \leftarrow t_0$ and state $x_n \leftarrow x_0$;
2. Calculate $a_j(x_n)$ and $a(x_n)$ from the current state x_n ;
3. Generate r_1 and r_2 from a uniform distribution [0,1);
4. Calculate τ from equation (3.4) and j from equation (3.5);
5. Perform reaction R_j with the update of $t_{n+1} = t_n + \tau$ and $x_{n+1} = x_n + Sj$, where Sj is the stoichiometric vector indicating the changes in molecule numbers after one reaction j^{th} occurs;
6. Set n=n+1 and return to step 2.

SSA is implemented in SGNSim (A. S. Ribeiro & Lloyd-Price, 2007). SGNSim models a wide range of systems of chemically interacting elements. The extended version of SGNSim, SGNS2 (Lloyd-Price et al., 2012) was the first simulator to include multi-delayed events, dynamic compartments and molecule partitioning schemes in division. Thereby, in our study, the SGNS2 simulator will be used.

4. Results and Discussion

This chapter contains the results obtained in our study of potential differences in dynamics of transcription between mutant cells and wild-type cells under the same, optimal growth conditions. This study was performed under control of P_{tetA} and P_{BAD}, to determine if existing differences in RNA production kinetics between mutant cells are solely σ factor-dependent or are also promoter-dependent.

Firstly, we present the results of the *in vivo* measurements of tagged RNA molecules made in wild-type cells and in mutant cells lacking σ^{38} or lacking σ^{54}. As mentioned before, these experiments were done under control of two different promoters during exponential growth phase and stationary growth phase. For all the experiments, assuming that transcription initiation consists of a sequence of exponentially distributed steps, it is inferred the number of steps as well as the duration of the underlying rate-limiting steps. The modeling strategy used to model the dynamics of transcription and the results of its simulations to explore the dynamics of gene expression under stress conditions are presented below. The discussion of the results obtained is done for all the measurements.

4.1. *In vivo* measurements of tagged RNA molecules in wild-type cells, in mutant cells lacking σ^{38} and in mutant cells lacking σ^{54}

When a bacterium is inoculated in a medium, it passes through four growth phases. The first growth phase is known as lag phase and corresponds to the time required for the adaptation to the new environment. In this phase its growth rate is 0. The second growth phase is the exponential phase, where the mass of the cell increases in an exponential manner. When the nutrients became exhaust or when the toxic metabolic products accumulate or inhibit growth, the cell enters in another phase, known as stationary phase. At this point, the growth ceases completely and the death of the bacteria starts. The last phase, denominated by death phase, is where there is a progressive death of the cell. Our measurements were done during the exponential and stationary phase, because there are σ factors responsible for transcribing genes expressed during exponential phase and other σ factors responsible for transcribing genes expressed during the stationary phase (Jishage et al., 1996). The results and the discussion of these experiments are presented below.

4.1.1. Measurements under control of P_{tetA}

These *in vivo* studies of the kinetics of transcription initiation were performed under control of P_{tetA} for three different *E. coli* strains. For the *in vivo* measurements of wild-type cells, which means containing all the σ factors presents in *E. coli*, it is used the *E. coli* strain BW25113 (Baba et al., 2006). The *E. coli* strain JW5437 (Baba et al., 2006) is used in *in vivo* measurements of mutant cells lacking σ^{38} and the *E. coli* strain JW3169 (Baba et al., 2006) is used in mutant cells lacking σ^{54}.

Once the sequence of the target gene contains 96 binding sites for the MS2 coat protein, the reporter proteins (MS2-GFP) can bind to the target RNA and a fluorescent spot is formed. These fluorescent spots can be observed in fluorescence microscopy images. The cells are placed under the confocal microscope during 2 hours with a measurement done at each 30 seconds. Analysis of these images is performed by a semi-automatic method (Kandhavelu et al., 2011) which does the detection and the masking of the cells from the images obtained (Figure 3.3).

As mentioned before, the experiments are made during the exponential and the stationary growth phases. Bellow, the results for the measurements made during these two phases, as well as, its discussion is addressed.

4.1.1.1. Measurements during exponential phase

From the images we extracted the number of cells as well as the number of intervals between productions of consecutive RNA molecules detected in individual cells (Number of samples). Table 4.1 shows these values for the three strains of *E. coli* analysed during the exponential growth phase, as well as the mean duration to complete a transcription initiation event once initiated (μ(s)), the standard deviation (σ(s)) of this interval duration. Both values are represented in seconds, and the variance over mean square value (CVS) obtained per each experiment. The CVS value is an important value once it indicates how spread the probability density of the protein number it is and how noisy the regulation of one gene to another it is. However, this value will be discussed later on the sub-chapter of the inference of step in transcription initiation.

Studies have shown that some σ factors are responsible for the transcription of genes that are expressed during the exponential growth phase while others are responsible to transcribe

genes that are expressed during the stationary phase (Jishage et al., 1996). It is also known that the intracellular concentration of σ^{70} as well as of σ^{54} is the same in the exponential growth phase and in the stationary phase, when analysing one strain of E. coli (Table 4.2) (Jishage et al., 1996). On the other hand, the intracellular concentration of σ^{38} is approximately zero during the exponential growth phase, but it increases significantly in the stationary phase (Table 4.2) (Jishage et al., 1996).

Table 4.1: Statistics on the time intervals between consecutive transcription events in individual cells under control of P_{tetA} during exponential phase. Number of cells analyzed, number of intervals between production of consecutive RNA molecules detected in individual cells (Number of samples), mean duration of production intervals in seconds (μ(s)), the standard deviation (σ(s)) and the square of the coefficient of variation (CVS) of the interval duration obtained in our experiment with wild-type cells and with mutant cells lacking σ^{38} and σ^{54} during exponential phase under control of P_{tetA}.

Strains	Wild-type cells	Mutant cells (lacking σ^{38})	Mutant cells (lacking σ^{54})
Number of cells	306	545	800
Number of samples	268	351	177
μ (s)	1024	960	976
σ (s)	772	881	627
CVS	0.57	0.84	0.41

Table 4.2: Intracellular levels of σ^{70}, σ^{54} and σ^{38} subunits in E.coli W3110 and MC4100. This information is obtained from (Jishage et al., 1996).

σ subunit	Level (fmol/μg) of σ subunit in strain:			
	W3110		MC4100	
	Exponential phase	Stationary phase	Exponential phase	Stationary phase
σ^{70}	150-170	150-170	50-80	50-80
σ^{54}	20-30	20-30	3-5	3-5
σ^{38}	0	40-60	0	20-30

From the data in Table 4.1, we observe that the time needed to complete a transcription initiation event once initiated is approximately the same for all the three strains analysed.

From Table 4.2 it is possible to observe that the intracellular level of σ^{54} is not negligible (around one tenth of the intracellular level of the total amount of σ^{70} ((Jishage et al., 1996))) during exponential phase. Thus, it was expected a difference between kinetics of transcription initiation of wild-type cells and of mutant cells lacking this σ factors. Once this difference is not verified, it is possible to conclude that P_{tetA} does not contain any consensus for σ^{54} binding.

For the mean production interval of mutant cells lacking σ^{38} comparing with wild-type results was not expected any difference, since during exponential phase the intracellular level of this σ factor is null (Table 4.2). From these results, it is possible to conclude that P_{tetA} is not affected by σ factor composition during the exponential phase.

4.1.1.2. Measurements during stationary phase

Since during stationary phase the intracellular level of σ^{38} is not negligible, it was important to evaluate the dynamics of transcription initiation when this σ factor is lacking when compared with the dynamics of transcription of wild-type cells.

Table 4.3: Statistics on the time intervals between consecutive transcription events in individual cells under control of P_{tetA} during stationary phase. Number of cells analyzed, number of intervals between production of consecutive RNA molecules detected in individual cells (Number of samples), mean duration of production intervals in seconds (μ(s)), the standard deviation (σ(s)) and the square of the coefficient of variation (CVS) of the interval duration obtained in our experiment with wild-type cells and with mutant cells lacking σ^{38} and σ^{54} during stationary growth phase under control of P_{tetA}.

Strains	Wild-type cells	Mutant cells (lacking σ^{38})	Mutant cell (lacking σ^{54})
Number of cells	177	191	272
Number of samples	267	94	506
μ (s)	1105	1086	1207
σ (s)	852	972	873
CVS	0.60	0.80	0.52

Thereby the measurements during the stationary growth phase are made using the same strains and procedure of the measurements during exponential growth phase. Further, the same parameters are extracted for the three strains analysed, as the number of cells analysed, the number of samples, the mean duration of production intervals, the standard deviation and the CVS of the interval duration. These values are represented in Table 4.3.

Comparing between strains and from Table 4.3, it is possible to observe that the time needed to complete a transcription initiation event in wild-type cells is approximately the same than for the measurements made during exponential phase (Table 4.1). The transcription time for mutant cells lacking σ^{38} becomes longer, indicating an increase in the proportion of σ^{38} in wild-type strain (Table4.2). On the other hand, the mean production interval of mutant lacking σ^{54} is higher when comparing with the same measurements during exponential phase.

With the data from Tables 4.1 and 4.3 it is possible to infer that the transcription kinetics under control of P_{tetA}, during exponential and stationary phases, is not affected by σ factor composition, due the dynamics of transcription almost does not change between strains during exponential and stationary growth phase.

4.1.2. Measurements under control of P_{BAD}

The measurements under control of P_{BAD} were made following the same procedure used for P_{tetA}, differing only on the inducer used to induce the target. In these measurements, the *E. coli* strains under study are the same than the strains used under control of P_{tetA}. This sub-chapter contains the results and the discussion of the results of these measurements during exponential phase and stationary phase, similar to the previous.

4.1.2.1. Measurements during exponential phase

Likewise as the analysis made for the measurements under control of P_{tetA}, Table 4.4 presents the values obtained for the experiments made during exponential phase under control of P_{BAD}. As mentioned before and from Table 4.2, it is known that the intracellular level of σ^{38} during exponential phase is not significant. Therefore, it was not expected a marked difference in RNA production rates between wild-type cells and mutant cells lacking this σ

factor, which can be confirmed by μ(s) represented in Table 4.4. However, as the intracellular level of σ^{54} cannot be neglected during exponential phase (Table 4.2), it was expected a difference in RNA production rates between wild-type cells and mutant cells lacking σ^{54}. From Table 4.4, it is possible to observe that mutant cells lacking σ^{54} need less time to complete a transcription initiation event once initiated than wild-type cells. Therefore, it is possible to conclude that P_{BAD} is preferentially transcribed by σ^{70} than by σ^{54}.

Table 4.4: Statistics on the time intervals between consecutive transcription events in individual cells under control of P_{BAD} during exponential phase. Number of cells analyzed, number of intervals between production of consecutive RNA molecules detected in individual cells (Number of samples), mean duration of production intervals in seconds (μ(s)), the standard deviation (σ(s)) and the square of the coefficient of variation (CVS) of the interval duration obtained in our experiment with wild-type cells and with mutant cells lacking σ^{38} and σ^{54} during exponential growth phase under control of P_{BAD}.

Strains	Wild-type cells	Mutant cells (lacking σ^{38})	Mutant cell (lacking σ^{54})
Number of cells	611	832	680
Number of samples	482	281	395
μ (s)	733	645	441
σ (s)	623	580	399
CVS	0.72	0.80	0.82

In the exponential phase, there is a difference in the dynamics of RNA production between mutant cells lacking σ^{54} and the wild-type strain but no significant difference between wild-type and mutant cells lacking σ^{38}. These results are in agreement with the measurements of the intracellular levels (Table 4.2) of σ^{54} and σ^{38} in cells under optimal conditions during the exponential phase. We conclude that the number of σ^{54} is a rate-limiting factor of transcription of the P_{BAD} under optimal conditions during the exponential phase, while σ^{38} is not.

4.1.2.2. Measurements during stationary phase

These measurements follow the procedure described before for the experiments during the exponential phase. From Table 4.5 it is possible to observe that the difference in time needed to complete a transcription initiation event once initiated between wild-type cells and mutant cells lacking σ^{38} becomes greater when comparing with the results of the

measurements during the exponential phase. This might indicate an increase in the proportion of σ^{38} in the wild-type strain. On the other hand, the difference between wild-type cells and mutant lacking σ^{54} is smaller (Table 4.5), indicating a decrease in the proportion of σ^{54} in the wild-type strain. Since σ^{54} quantity is not reported to be affected by the growth phase (Table 4.2), we suggest that σ^{70} is not much affected by the growth phase as well.

Table 4.5: Statistics on the time intervals between consecutive transcription events in individual cells under control of P_{BAD} during stationary phase. Number of cells analyzed, number of intervals between production of consecutive RNA molecules detected in individual cells (Number of samples), mean duration of production intervals in seconds (μ(s)), the standard deviation (σ(s)) and the square of the coefficient of variation (CVS) of the interval duration obtained in our experiment with wild-type cells and with mutant cells lacking σ^{38} and σ^{54} during stationary growth phase under control of P_{BAD}.

Strains	Wild-type cells	Mutant cells (lacking σ^{38})	Mutant cell (lacking σ^{54})
Number of cells	174	445	588
Number of samples	218	47	1215
μ (s)	1657	970	1156
σ (s)	1342	629	848
CVS	0.66	0.42	0.54

Analysing the data from Tables 4.4 and 4.5 it is possible to conclude that the dynamics of transcription initiation under control of P_{BAD}, during both cellular growth phases, is affected by the σ factor composition, due the significant differences of the mean production intervals between strains during exponential and stationary growth phases. Further, as there are significant differences in RNA production kinetics between wild-type cells and mutant cells under control of P_{BAD}, and under control of P_{tetA}, the dynamics of transcription initiation is promoter-dependent.

4.2. Inference of steps in transcription initiation

The number and duration of the sequential steps in transcription initiation can be inferred by maximum-likelihood from the distribution of intervals between productions of consecutive RNA molecules (Kandhavelu et al., 2011). Following the method used in (Kandhavelu et al., 2011), it is possible to fit the measured distribution using a small number of steps, here represented by *d*, which is in agreement with the number of steps believed to be rate-limiting from *in vitro* studies (Buc & McClure, 1985; Lutz et al., 2001). Using this method only the number of sequential steps inferred can be assessed, being impossible to determine their temporal order.

The distribution of intervals between productions of consecutive RNA molecules per each experiment is represented in Figure 4.1, for the measurements under control of P_{tetA} during the exponential growth phase, in Figure 4.2, for the same promoter during the stationary growth phase, in Figure 4.3 for the experiments under control of P_{BAD} during exponential phase and in Figure 4.4 for the ones under the same promoter during stationary phase. In these figures are also shown the curves that best fit their distribution, for a number of steps (*d*) varying from 1 to 3. From these figures it is possible to observe that for all the strains under control of the two promoters during the two cellular growth phases, the shape of the distributions of intervals are not exponential-like, which it is consistent with (Mäkelä et al., 2013; Muthukrishnan et al., 2012). These studies show that the process of RNA production under the control of P_{tetA} (Muthukrishnan et al., 2012) or P_{BAD} (Mäkelä et al., 2013) is not Poissonian. Namely, since the value of CVS (Table 4.1 and Table 4.3 to Table 4.5), for all the experiments made, is below 1, it is possible to affirm that this process is sub-Poissonian, in agreement with (Mäkelä et al., 2013; Muthukrishnan et al., 2012). The differences in CVS value between strains and cellular growth phases are due to changes in shape of the distributions (Figure 4.1 to Figure 4.4).

In Tables 4.6 and 4.8, the log-likelihood values are shown and the duration of the inferred steps for *d* ranging from 1 step to 4 steps models for the three strains under control of P_{tetA} during exponential phase and during stationary phase, respectively. These results are compared with the likelihood-ratio values test between pairs of models for the measurements in exponential phase and stationary phase represented in Tables 4.7 and 4.9, respectively. The inference of the number of steps can be made selecting a higher-degree model in detriment of a lower-degree model. Tables 4.10 and 4.12 represent the log-likelihood values for the measurements in exponential phase and stationary phase, respectively, under control of P_{BAD}.

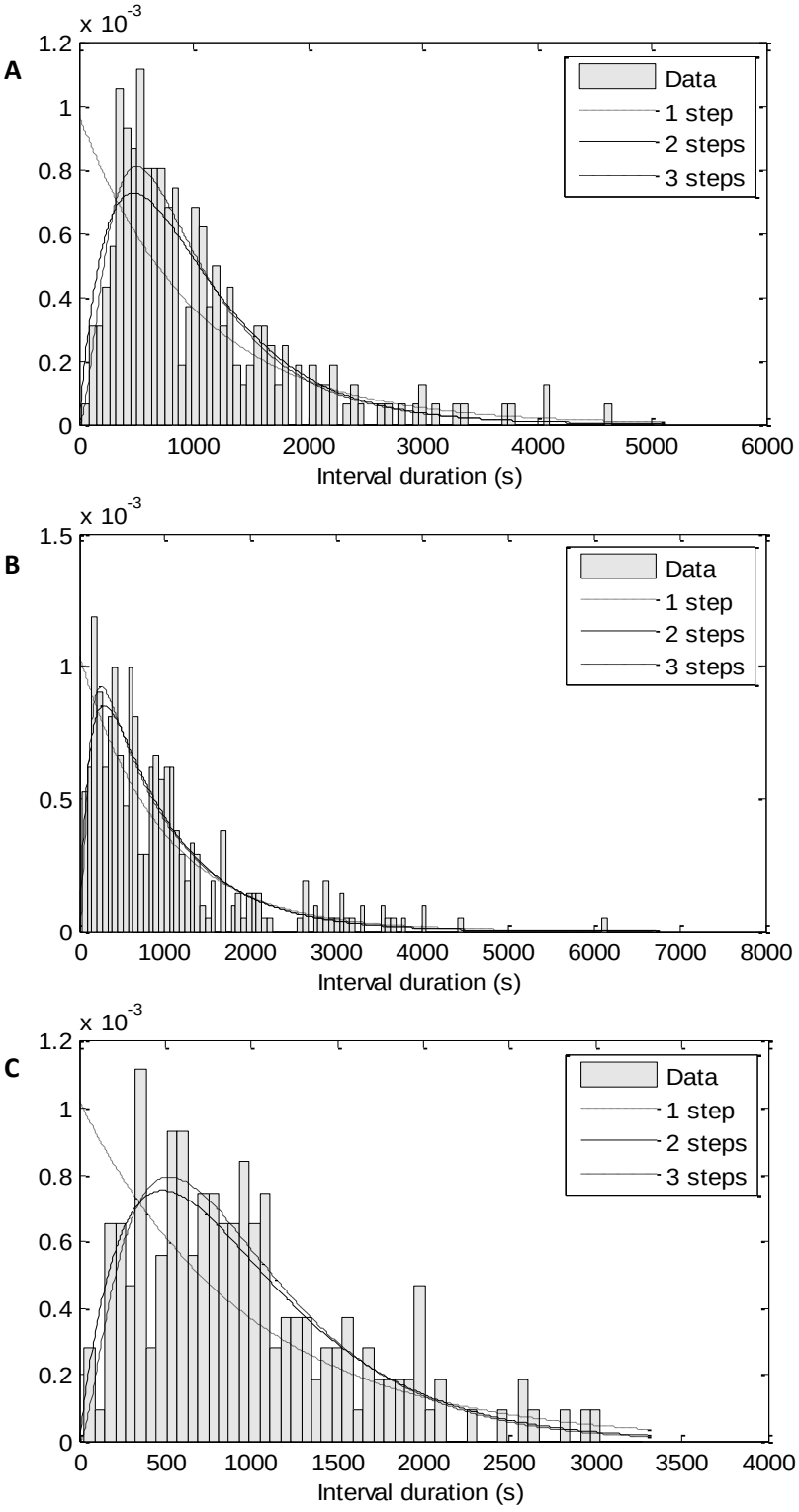

Figure 4.1: Histogram of the measured intervals between consecutive transcription events under control of P_{tetA} during exponential phase. The distribution of intervals between consecutive productions of transcripts events represented is obtained for wild-type cells (A), mutant cells lacking σ^{38} (B) and mutant cells lacking σ^{54} (C) under full induction and under control of P_{tetA} during exponential phase. Each bar represents 60 seconds and the measurement time is 2 hours (measured every 30 seconds). The histogram of measured intervals is superimposed with probability density functions of models with 1 step (dotted line), 2 steps (solid line) and 3 steps (dashed line) that best fit the data.

The inference of the number of steps is made comparing these results with the ones from Tables 4.11 and 4.13, selecting a higher-degree model in detriment of a lower-degree model, likewise as the analysis made with the measurements under control of P_{tetA}.

Table 4.6: Log-likelihood and duration of the steps of the models with d equal to 1 to 4 steps for three strains under control of P_{tetA} during exponential growth phase.

Wild-type cells					
d	**Log-likelihood**	**Duration of steps (s)**			
1	-2126	1025			
2	-2088	354	670		
3	-2082	156	156	712	
4	-2081	98	98	98	731

	Mutant cells (lacking σ^{38})					Mutant cells (lacking σ^{54})				
d	**Log-likelihood**	**Duration of steps (s)**				**Log-likelihood**	**Duration of steps (s)**			
1	-2761	960				-1411	977			
2	-2741	133	827			-1382	487	489		
3	-2740	58	58	844		-1381	60	458	458	
4	-2740	70	21	21	847	-1381	18	33	463	463

Table 4.7: Likelihood-ratio tests. P values between pairs of models for the three strains under control of P_{tetA} during exponential phase. d_x represents the null model and can have values from 1 to 3, while d_{x+1} can vary between 2 to 4 and represents the alternative model.

(d_x,d_{x+1})	Wild-type cells	Mutant cells (lacking σ^{38})	Mutant cells (lacking σ^{54})
(1,2)	0	0	0
(2,3)	0.001	0.081	0.095
(3,4)	0.123	0.629	0.789

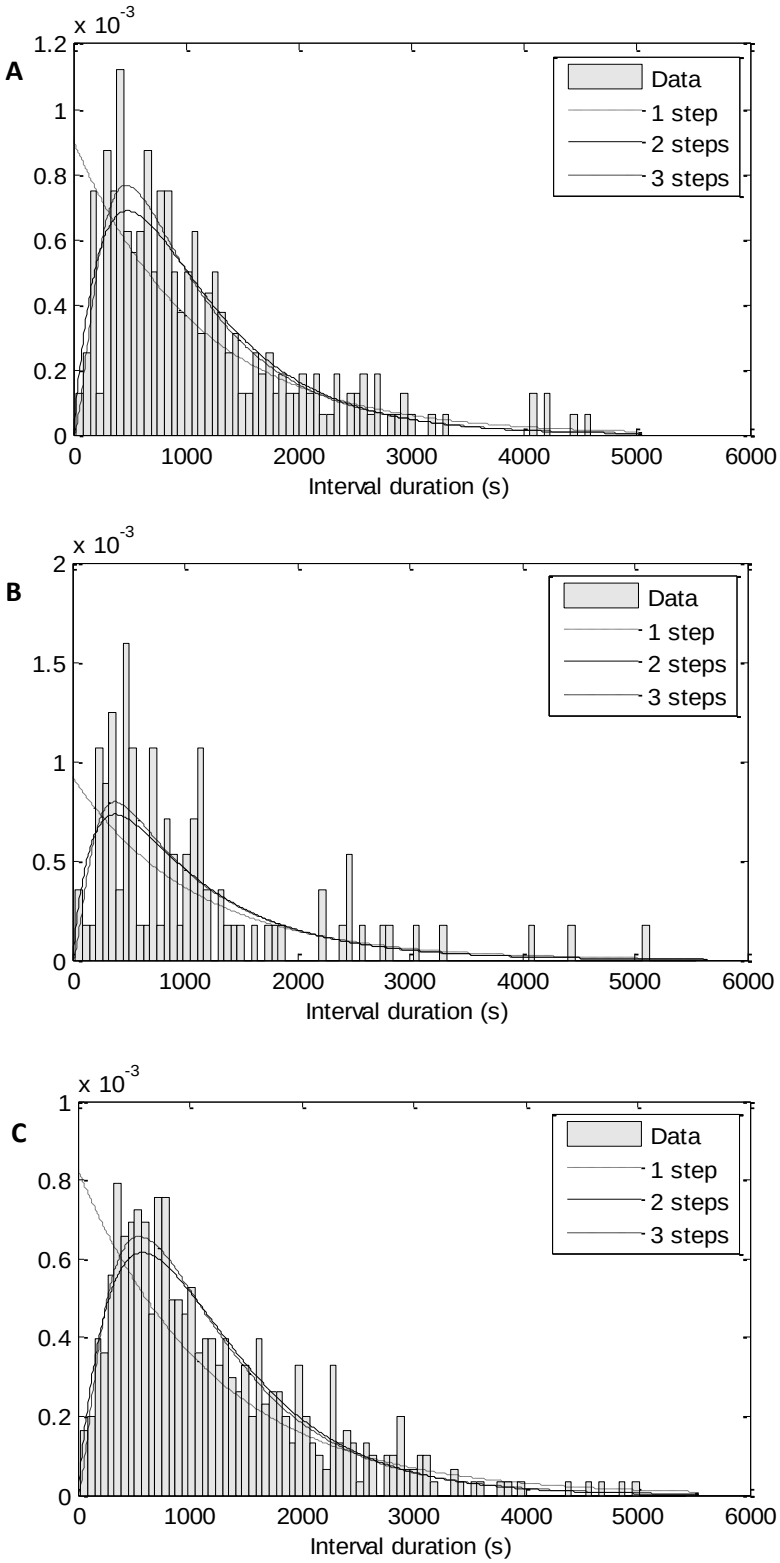

Figure 4.2: : Histogram of the measured intervals between consecutive transcription events under control of P_{tetA} during stationary phase. The distribution of intervals between consecutive productions of transcripts events represented is obtained for wild-type cells (A), mutant cells lacking σ^{38} (B) and mutant cells lacking σ^{54} (C) under full induction and under control of P_{tetA}. Each bar represents 60 seconds and the measurement time is 2 hours (measured every 30 seconds). The histogram of measured intervals is superimposed with probability density functions of models with 1 step (dotted line), 2 steps (solid line) and 3 steps (dashed line) that best fit the data.

Table 4.8: Log-likelihood and duration of the steps of the models with d equal to 1 to 4 steps for three strains under control of P_{tetA} during stationary growth phase.

Wild-type cells						
d	**Log-likelihood**	**Duration of steps (s)**				
1	-2138	1105				
2	-2109	318	788			
3	-2107	133	133	839		
4	-2107	132	2	132	840	

	Mutant cells (lacking σ^{38})					Mutant cells (lacking σ^{54})				
d	**Log-likelihood**	**Duration of steps (s)**				**Log-likelihood**	**Duration of steps (s)**			
1	-751	1086				-4097	1207			
2	-743	196	889			-4037	459	748		
3	-742	97	97	892		-4036	52	839	316	
4	-742	94	94	5	893	-4035	23	23	327	834

Table 4.9: Likelihood-ratio tests. P values between pairs of models for the three strains under control of P_{tetA} during stationary phase. d_x represents the null model and can have values from 1 to 3, while d_{x+1} can vary between 2 to 4 and represents the alternative model.

(d_x, d_{x+1})	Wild-type cells	Mutant cells (lacking σ^{38})	Mutant cells (lacking σ^{54})
(1,2)	0	0	0
(2,3)	0.037	0.167	0.062
(3,4)	0.932	0.890	0.500

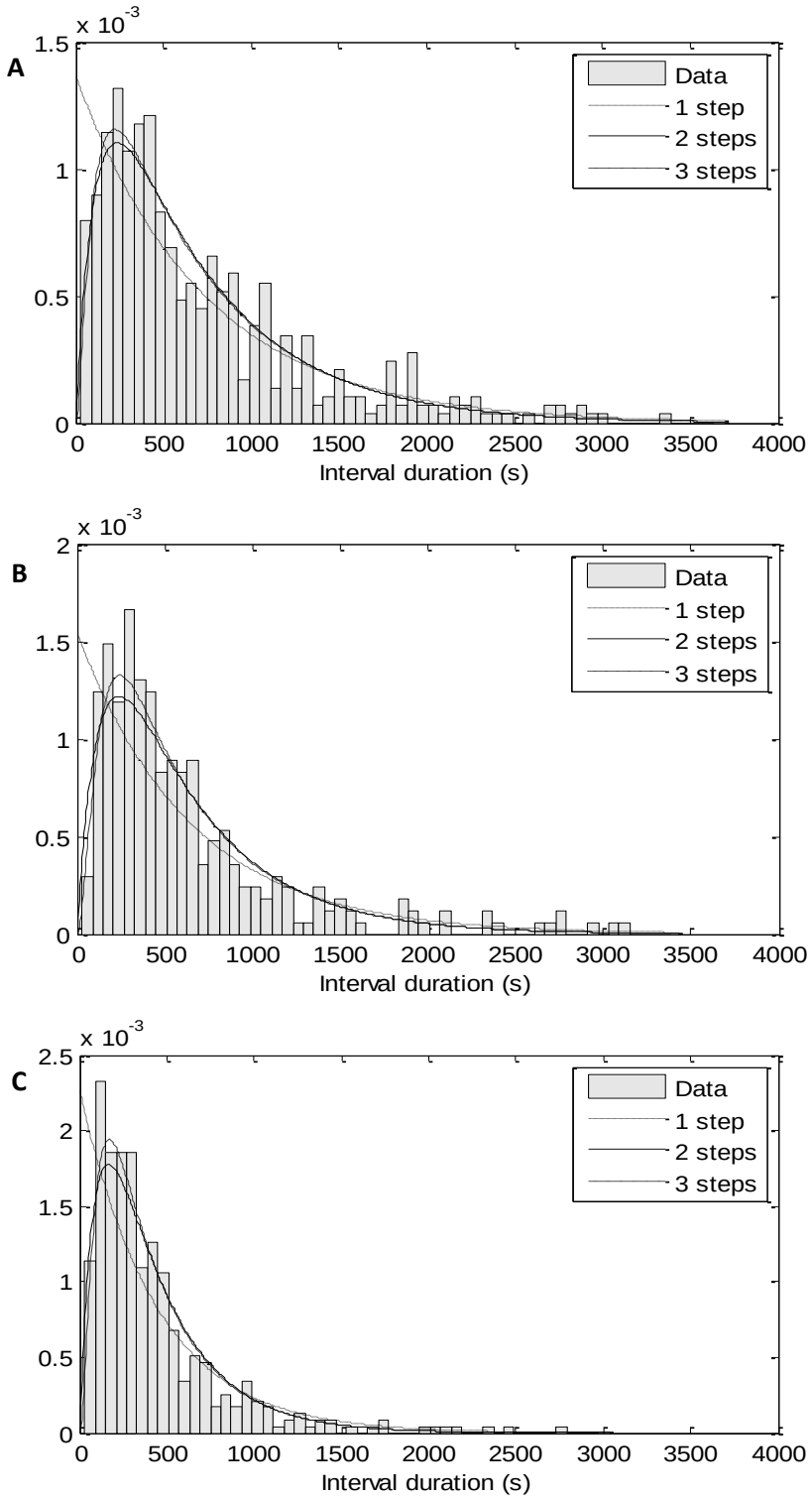

Figure 4.3: : Histogram of the measured intervals between consecutive transcription events under control of P_{BAD} during exponential phase. The distribution of intervals between consecutive production of trancript events represented is obtained for wild-type cells (A), mutant cells lacking σ^{38} (B) and mutant cells lacking σ^{54} (C) under full induction and under control of P_{BAD} . Each bar represents 60 seconds and the measuremet time is 2h (measured every 30 seconds). The histogram of measured intervals is superimposed with probability density functions of models with 1 step (dotted line), 2 steps (solid line) and 3 steps (dashed line) that best fit the data.

Table 4.10: Log-likelihood and duration of the steps of the models with d equal to 1 to 4 steps for three strains under control of P_{BAD} during exponential growth phase.

Wild-type cells					
d	**Log-likelihood**	**Duration of steps (s)**			
1	-3662	733			
2	-3632	145	618		
3	-3631	88	19	625	
4	-3631	95	7	7	624

	Mutant cells (lacking σ^{38})					Mutant cells (lacking σ^{54})				
d	**Log-likelihood**	**Duration of steps (s)**				**Log-likelihood**	**Duration of steps (s)**			
1	-2100	647				-2800	441			
2	-2073	129	518			-2761	93	347		
3	-2068	64	64	519		-2754	46	46	349	
4	-2066	41	41	41	523	-2753	20	29	29	353

Table 4.11: Likelihood-ratio tests. P values between pairs of models for the three strains under control of P_{BAD} during exponential phase. d_x represents the null model and can have values from 1 to 3, while d_{x+1} can vary between 2 to 4 and represents the alternative model.

(d_x, d_{x+1})	Wild-type cells	Mutant cells (lacking σ^{38})	Mutant cells (lacking σ^{54})
(1,2)	0	0	0
(2,3)	0.167	0.002	0
(3,4)	0.789	0.067	0.066

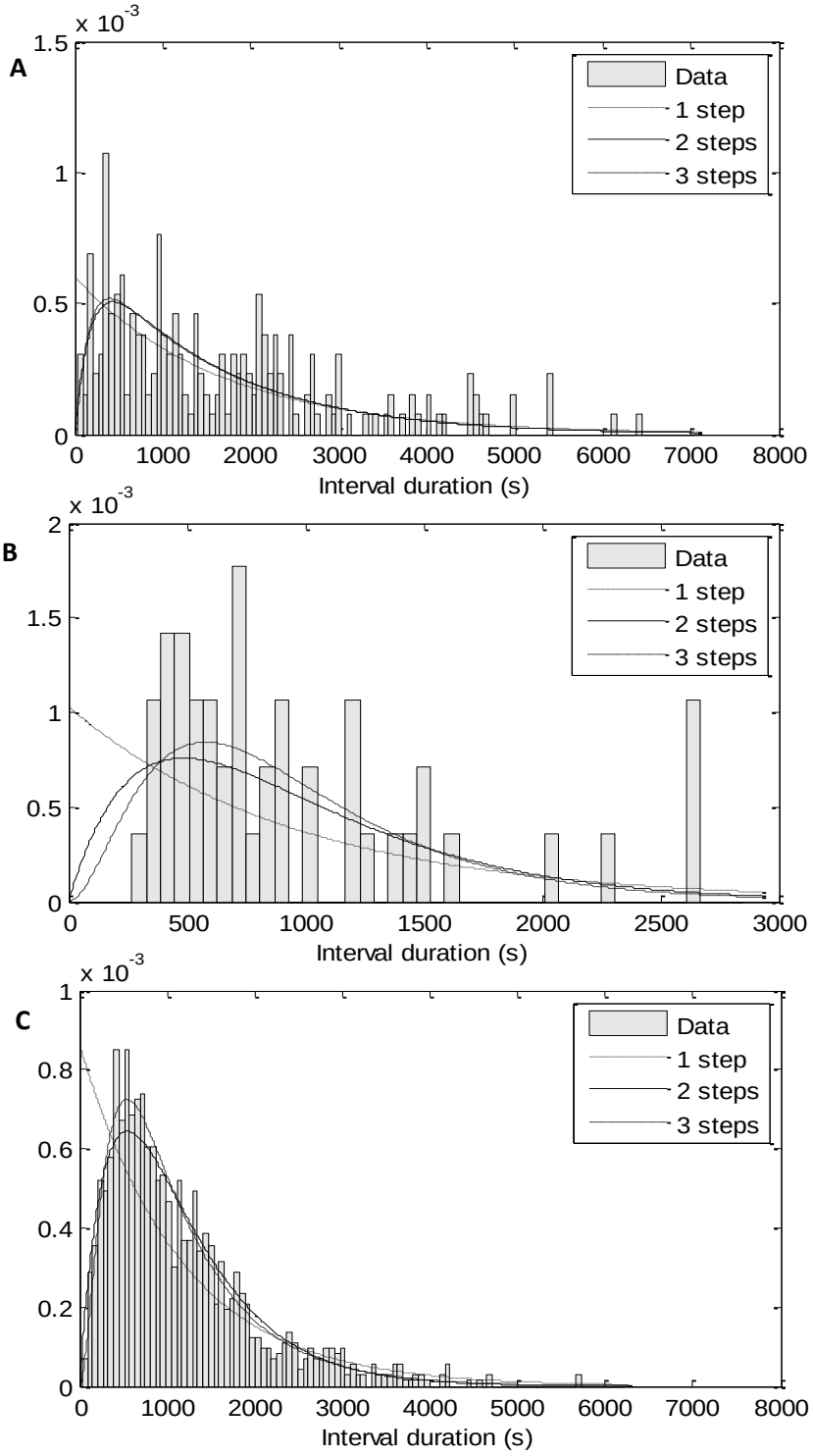

Figure 4.4: : Histogram of the measured intervals between consecutive transcription events under control of P_{BAD} during stationary phase. The distribution of intervals between consecutive production of trancript events represented is obtained for wild-type cells (A), mutant cells lacking σ^{38} (B) and mutant cells lacking σ^{54} (C) under full induction and under control of P_{BAD} . Each bar represents 60 seconds and the measuremet time is 2h (measured every 30 seconds). The histogram of measured intervals is superimposed with probability density functions of models with 1 step (dotted line), 2 steps (solid line) and 3 steps (dashed line) that best fit the data.

Table 4.12: Log-likelihood and duration of the steps of the models with d equal to 1 to 4 steps for three strains under control of P_{BAD} during stationary growth phase.

Wild-type cells					
d	**Log-likelihood**	**Duration of steps (s)**			
1	-1834	1657			
2	-1825	180	1477		
3	-1825	17	1491	149	
4	-1825	9	9	145	1493

	Mutant cells (lacking σ^{38})					Mutant cells (lacking σ^{54})				
d	**Log-likelihood**	**Duration of steps (s)**				**Log-likelihood**	**Duration of steps (s)**			
1	-370	970				-9784	1156			
2	-360	485	485			-9629	420	737		
3	-358	212	212	545		-9618	161	161	834	
4	-356	133	133	133	571	-9618	213	13	99	831

Table 4.13: Likelihood-ratio tests. P values between pairs of models for the three strains under control of P_{BAD} during stationary phase. d_x represents the null model and can have values from 1 to 3, while d_{x+1} can vary between 2 to 4 and represents the alternative model.

(d_x, d_{x+1})	Wild-type cells	Mutant cells (lacking σ^{38})	Mutant cells (lacking σ^{54})
(1,2)	0	0	0
(2,3)	0.554	0.023	0
(3,4)	0.760	0.113	0.506

From Tables 4.7, 4.9, 4.11 and 4.13, we can infer that the single step model is insufficient to explain the measurements when comparing to the multi-step models, once the p-value is

equal to 0 for all the strains under control of the two promoters during the two cellular growth phases. On the other hand, the 2 steps as well as the 3 steps models fits the measurements in agreement with previous studies (Buc & McClure, 1985; Lutz et al., 2001) where it is concluded that both closed and open complex formation are rate-limiting. We find, by comparing the log-likelihood values for *d* equal to 3 and for *d* equal 4 (Table 4.6, 4.8, 4.10 and 4.12), that increasing the number of steps beyond two does not result in significantly better fit of the model to the data. Therefore, we conclude that there are three rate-limiting steps in transcription initiation by P_{tetA} and by P_{BAD} for all the strains during exponential and stationary phase.

4.3. Model of σ factors

As mentioned before, the goal of this thesis is to study how the dynamics of transcription differ between mutant cells (lacking σ^{38} or σ^{54}) and wild-type cells (containing all σ factors), under the same optimal growth conditions. It was also mentioned that σ factor is the transcription initiation factor and its number varies from bacteria to bacteria. In case of *E. coli* it is known that it contains seven different σ factors each transcribing specific sets of genes (Ishihama, 2000). Further, some σ factors transcribe genes expressed during exponential phase while others are responsible for the transcription of genes expressed during stationary phase (Jishage et al., 1996). The proposed study is focused in three of them: the housekeeping σ^{70}, the master regulator of the general stress response σ^{38}, and the one responsible for expression of genes which are activated by a deficience of nitrogen or other stress response σ^{54}.

The modeling strategy used to model the dynamics of transcription was the development of a deterministic model of RNAp dynamics coupled with stochastic gene expression (Buc & McClure, 1985). The model developed is based in the model proposed in (Grigorova, Phleger, Mutalik, & Gross, 2006). However, in our model the target promoter (P_{tetA} or P_{BAD}) is a single copy promoter. In order to simulate all the *in vivo* measurements previously made during exponential and stationary phases, and assert some conclusions of the influence of changing some parameters, we adjust some values like the intracellular concentration of the σ factors, which vary between strains of *E. coli* and cellular growth phase and the value of the disassociation constants, which vary between target promoter.

In our study we use an equilibrium model of RNAp binding to σ factors to explore the influence of some parameters of this model in the dynamics of transcription initiation in E. coli (Figure 4.5). The reactions used in our model, as well as the parameters and its values, are presented below.

As only one σ factor can bind to the free RNAp at a specific moment, we represent generic reactions using the letter 'X' to represent the σ factor. If X is equal to 70, 38 or 54, it is respectively σ^{70}, σ^{38} or σ^{54} that is in study. Note that the value of X does not change from reaction to reaction during transcription, which means that if X is defined as 70 in the first reaction (4.1) in the last one (4.4) X is still 70. We adopt this methodology to simply the understanding of our model (Figure 4.5). However, in the model developed in custom software, it is taking into account all the reactions per each σ factor.

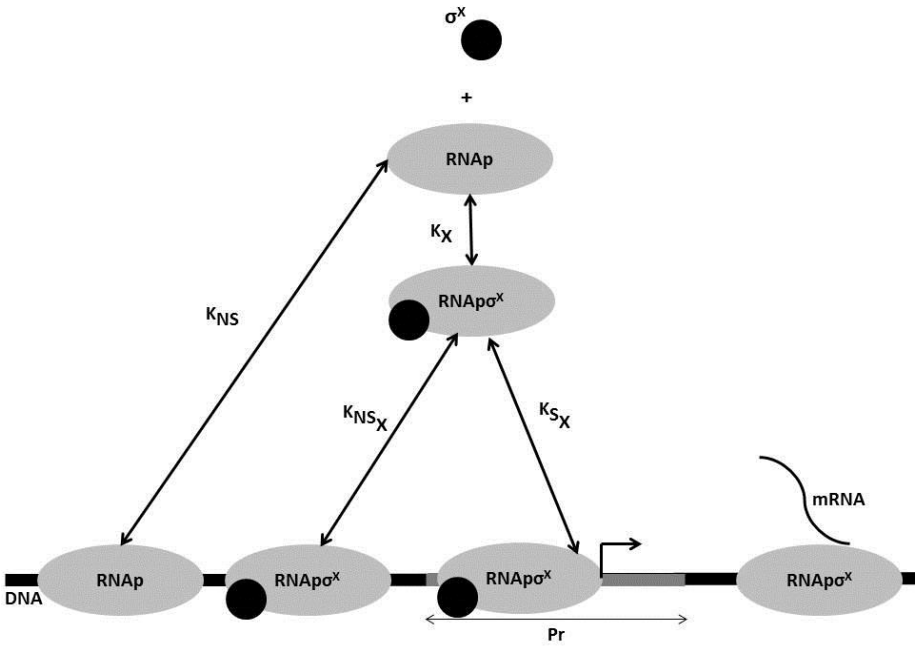

Figure 4.5: Scheme of the model used in our study. Here only one σ factor is represented, to simplify the representation of the reactions. In this figure, X can be 70, 54 or 38, in order to represent the three σ factors used, respectively σ^{70}, σ^{54} or σ^{38}, as well as the rates associated. The σ^{X} factor can bind specifically to the free RNAp, forming the holoenzyme ($RNAp\sigma^{X}$) with a disassociation constant of K_X. On the other hand, the free RNAp can bind non-specifically to the DNA chain with a rate, represented by K_{NS}. The holoenzyme formed can bind specifically the promoter (Pr) with a dissociation rate of K_{SX} and start the transcription of a molecule of messenger RNA which is release as soon as the termination site is reached. However the holoenzyme can also bind non-specifically the DNA chain with a disassociation rate here represented by K_{NSX}.

As mentioned before, in order to initiate transcription, it is necessary that the RNAp holoenzyme ($RNAp\sigma^{X}$) (Figure 4.5) is formed. This reaction is represented in 4.1 where K_X is the disassociation constant between RNAp core enzyme and σ^{X}.

$$RNAp + \sigma^{X} \overset{K_X}{\leftrightarrow} RNAp\sigma^{X} \qquad (4.1)$$

This holoenzyme can find the specific gene for each σ factor (reactions 4.2) and bind the promoter Pr with a disassociation constant represented by K_{S_X} (Figure 4.5).

$$RNAp\sigma^X + Pr \overset{K_{S_X}}{\leftrightarrow} Pr_X \qquad (4.2)$$

The promoter specifically bound by the specific holoenzyme (Pr_X) takes an interval of time until forming a closed complex (Pr_{c_X}) (reaction 4.3). This interval is also known as the dissociation constant, here represented by k_{c_X}. This step is not represented in Figure 4.5 but can be observed in Figure 2.1.

$$Pr_X \overset{k_{c_X}}{\rightarrow} Pr_{c_X} \qquad (4.3)$$

The closed complex isomerizes to form an open complex Pr_{o_X}, (reactions 4.4) with the transcription bubble (Figure 2.1). In this reaction, k_{o_X} represents the rate constant of formation of the open complex. This open complex allows that after transcription of some nucleotides, RNAp leaves the promoter and starts to elongate the RNA chain (Figure 4.5) until it finds the termination signal, at which point transcription stops and the molecule of messenger RNA is released. This molecule of mRNA will be translated by ribosomes in order to produce proteins.

$$Pr_{c_X} \overset{k_{o_X}}{\rightarrow} Pr_{o_X} \rightarrow P + RNA \qquad (4.4)$$

Note that although the σ factors are released after transcription of a few nucleotides, as mentioned in the chapter of state of the art, it is known that this does not affect the amount of σ factors in general. Thereby this step is not modeled and is not represented in Figure 4.5.

Since most RNAp binds to DNA (Grigorova et al., 2006; Shepherd, Dennis, & Bremer, 2001), the proportion of the holoenzymes in cytosol that is subject to σ factor competition is determined by the amount of unbound σ factors of each type. We assume that σ factors can only bind to free RNAp, which means that it is not bound to DNA, therefore, the proportion of unbound σ factors of each type determines the composition of not only the free holoenzymes but also holoenzymes that are bound to DNA.

As in (Grigorova et al., 2006), since the intracellular concentration of σ factors is smaller than the amount of RNAp (both holoenzyme and core enzyme), the amount of free RNAp is not a rate-limiting factor and as such does not require an explicit representation in our model.

Thereby the parameters of the model here presented are: the amount of RNAp core enzyme that is free (here modelled with E); the amount of holoenzyme that it is free (Eσ); the amount of nonspecifically binding of E to DNA (here modelled with E_b); the quantity of holoenzyme that bind nonspecifically to DNA ($E\sigma^{70}_b$, $E\sigma^{38}_b$ or $E\sigma^{54}_b$) and that bind specifically to DNA in the promoter region ($E\sigma^{70}$, $E\sigma^{38}$ or $E\sigma^{54}$); and the total number of σ^{70} ($n_{\sigma^{70}}$) σ^{38} ($n_{\sigma^{38}}$)and σ^{54} ($n_{\sigma^{54}}$).

Recent evidences (Grigorova et al., 2006) predict that σ factors only compete to bind to E when their total number is higher than the total amount of RNAp, rather than the amount of free RNAp, and σ^{70} is in excess of total of E. Other study (Shepherd et al., 2001) suggests that only a small percentage of the total amount of RNAp in the cell is represented by both free and non-specifically bound holoenzyme. As mentioned before our target is a single promoter. Thereby the non-specific binding is the binding of holoenzyme that blocks the target promoter, preventing the binding of other holoenzyme. Moreover, since our target promoter is single copy, non-specific binding should not affect the sigma factor composition but only the promoter dynamics.

Once the value of free RNAp and RNAp that is non-specifically bound to DNA corresponds to a small percentage of the total amount of RNAp (Grigorova et al., 2006), we ignore these values (E and E_b). It is known that the total number of σ^{70} , $n_{\sigma^{70}}$, consists of the sum of σ^{70} that it is free with the amount of σ^{70} that it is bound to RNAp, which means $E\sigma^{70}_b$ (equation 4.5). The same is verified to σ^{38} and σ^{54} (equations 4.6 and 4.7). The total amount of free holoenzyme (Eσ) could be expressed by the sum of $E\sigma^{70}_b$ with $E\sigma^{38}_b$ and $E\sigma^{54}_b$ (equation 4.8).

$$n_{\sigma^{70}} = \sigma^{70} + E\sigma^{70}_b \qquad (4.5)$$

$$n_{\sigma^{38}} = \sigma^{38} + E\sigma^{38}_b \qquad (4.6)$$

$$n_{\sigma^{54}} = \sigma^{54} + E\sigma^{54}_b \qquad (4.7)$$

$$n_{E\sigma} \approx E\sigma^{70}_b + E\sigma^{38}_b + E\sigma^{54}_b \qquad (4.8)$$

From the prediction that to initiate transcription the total number of σ factors has to be higher than the total amount of RNAp (Grigorova et al., 2006), we infer the following expressions (equations 4.9 and 4.10).

$$\frac{E\sigma^{54}}{E\sigma^{70}} = \frac{\sigma^{54}}{\sigma^{70}} = \frac{E\sigma_b^{54}}{E\sigma_b^{70}} \qquad (4.9)$$

$$\frac{E\sigma^{38}}{E\sigma^{70}} = \frac{\sigma^{38}}{\sigma^{70}} = \frac{E\sigma_b^{38}}{E\sigma_b^{70}} \qquad (4.10)$$

Figure 4.5 represents two rates which have not yet been mentioned. These rates are the dissociation constant for nonspecific binding of E to DNA chain (K_{NS}) and the dissociation constant for nonspecific binding of the holoenzyme to DNA chain (K_{NS_X}). Once that these rates do not affect the amount of free holoenzymes their values are not included in our model.

4.3.1. Results of simulations of the model of σ factors

As mentioned before, the difference between strains and cellular growth phase are modelled changing some parameters. In this sub-chapter, first, the values set per each simulation will be addressed as well as the results of the simulations for the three strains under control of P_{tetA} during stationary and exponential phases. Second, the results of the simulations for the three strains under control of P_{BAD} during the two cellular growth phases are presented. Further, the results of some simulations under stress response will be shown.

4.3.1.1. Under control of P_{tetA} during exponential phase

In this sub-chapter the values of the parameters used will be described, as well as the results obtained for the three strains under control of P_{tetA} during the exponential phase. The value of the total number of molecules of σ^{70}, σ^{38}, σ^{54} and Eσ are obtained by (Maeda, Fujita, & Ishihama, 2000). Note that the value itself is not the most important point, but the relationship between these values. Therefore, the total number of σ^{70}, $n_{\sigma^{70}}$, is defined as 700 molecules per cell (Maeda et al., 2000), the total number of σ^{54}, $n_{\sigma^{54}}$, is set as 110 molecules per cell (Maeda et al., 2000), the total number of σ^{38}, $n_{\sigma^{38}}$, is 0 molecules per cell (Table 4.2) (Maeda et al., 2000) and the total number of holoenzyme, $n_{E\sigma}$, is 600 molecules per cell. Using the equations 4.5 to 4.10 and applying some math it is possible to infer that the σ^{70} is 181 molecules per cell, σ^{54} is 29 molecules per cell, σ^{38} does not exist in exponential phase, $E\sigma_b^{70}$ is 519 molecules per cell, $E\sigma_b^{54}$ is 81 molecules per cell and $E\sigma_b^{38}$ is 0.

To simulate the three strains used in our study, three simulations were done, where to simulate the mutant cells, the total number of σ factor lacking and the number of corresponding holoenzymes is set as 0. Between strains, the total amount of holoenzyme, $n_{E\sigma}$, does not change, but between cellular growth phases it changes.

The disassociation constant of the specific binding of the holoenzyme, $E\sigma^{70}$, to P_{tetA} is set as 900 seconds, once that we are only interested in the amount of the holoenzyme unbound from the DNA and its copy number is higher when compared to the promoter copy number. Moreover, the specific binding occurs only around the promoter region and the specific binding is only done by $E\sigma^{70}$. Thereby the disassociation constants of specific binding of the holoenzyme $E\sigma^{54}$ and $E\sigma^{38}$ are set with a higher number (10^{13} seconds), in order to not influence the process of transcription. The value of the disassociation constants of the closed and open complex formation were first set as the values represented in Table 4.6, respectively $k_{c_{\sigma^{70}}}$ equal to 156 seconds and $k_{o_{\sigma^{70}}}$ as 712 seconds. Once P_{tetA} is not affected by σ composition, the higher value should correspond to the open complex formation and the smaller to the closed complex formation. As with these values the results were not acceptable, we adjusted the values for: $k_{c_{\sigma^{70}}}$ was set as 100 seconds as well as the elongation rate and the value of $k_{o_{\sigma^{70}}}$ was set as 600 seconds. The disassociation constants of the open complex formation for the other σ factors are set as 10^{11} seconds, due to the specific binding is only done by $E\sigma^{70}$ and the disassociation constant of the closed complex formation for σ^{54} and σ^{38} and the elongation rate are set as 10^{3}.

From Figure 4.6 and Table 4.14 it is possible to infer that the results from the simulations of the stochastic model are in agreement with the *in vivo* measurements. The distributions of the intervals between consecutive transcription events in the three strains under control of P_{tetA} during exponential phase and the CVS value shows that this process is a sub-Poissonian process. Thereby the results are also in agreement with previous studies made for this promoter (Muthukrishnan et al., 2012).

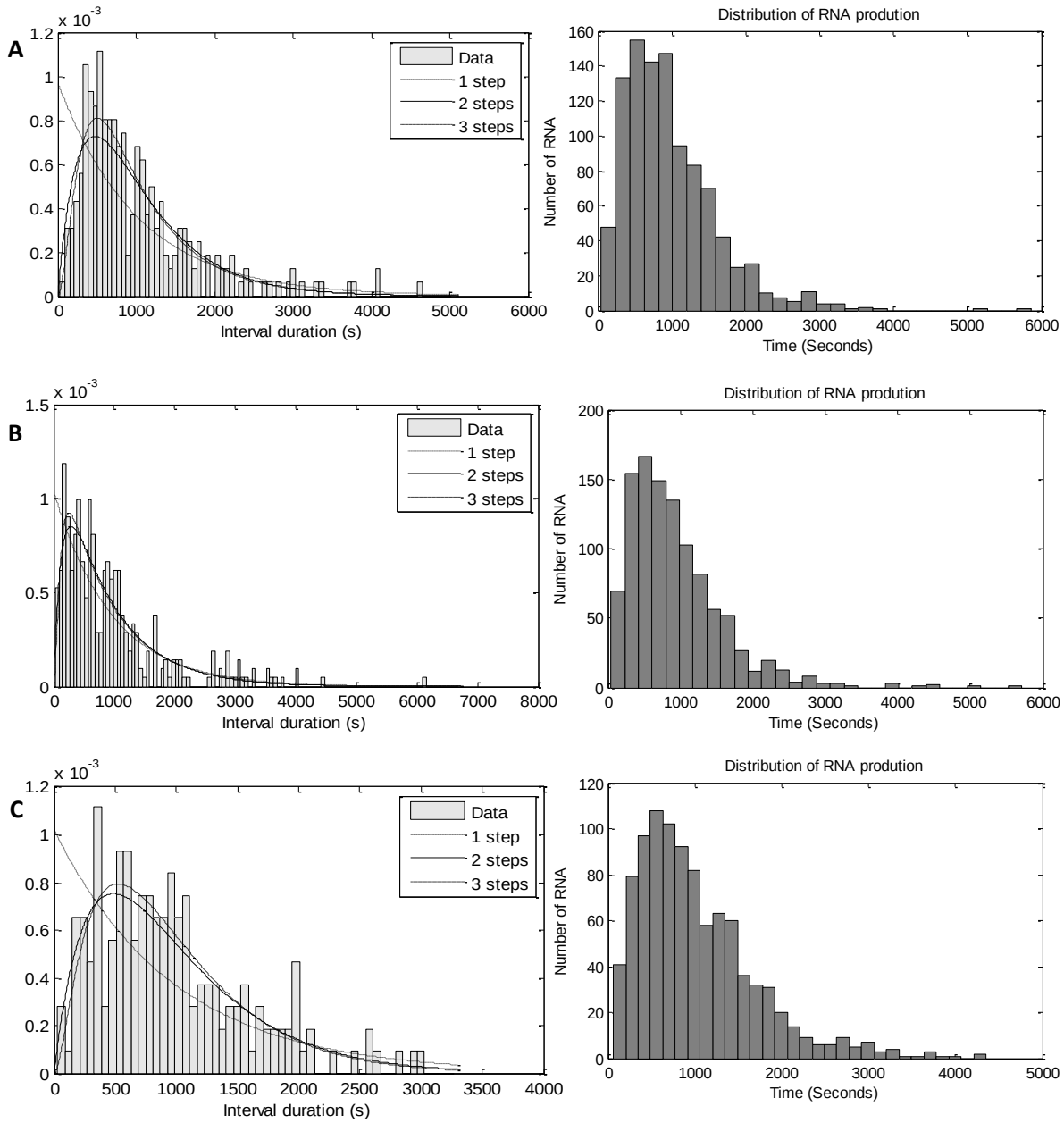

Figure 4.6: Distribution of RNA production under control of P_{tetA} during exponential phase. Histogram of the duration of the intervals between consecutive productions of RNA for the three strains (A – wild-type cells, B – mutant cells lacking σ^{38} and C – mutant cells lacking σ^{54}) under control of P_{tetA} during exponential phase resulting from in vivo measurements (left) and from simulations of the model (right).

In Figure 4.6, the distribution of production intervals obtained for each measurement (Figure 4.1) is repeated in the left side of Figure 4.6 in order to compare these results with the distribution of production intervals obtained from our model (Figure 4.6 - right side). From that, we conclude that both distributions follow the same trend. The mean production interval (μ(s)) (Table 4.14) is in agreement with the *in vivo* measurements (Table 4.1), once the value

between the three strains is similar. This similarity is due to P_{tetA} is not affected by σ factor composition.

Table 4.14: Results of the stochastic model developed in this study for the three strains under control of P_{tetA} during exponential phase. The mean production interval (μ(s)), its standard deviation (σ(s)) and the CVS value are presented for wild-type cells and mutant cells lacking σ^{38} and σ^{54}.

Strains	μ(s)	σ(s)	CVS
Wild-type cells	987	646	0.43
Mutant cells (lacking σ^{38})	937	659	0.49
Mutant cells (lacking σ^{54})	1025	690	0.45

4.3.1.2. Under control of P_{tetA} during stationary phase

For the simulations during the stationary phase some parameters were changed. The value of the total number of molecules of σ^{70}, σ^{38}, σ^{54} and Eσ were obtained by (Jishage et al., 1996; Maeda et al., 2000). The total number of σ^{70}, $n_{\sigma^{70}}$, were defined as 700 molecules per cell (Maeda et al., 2000), the total number of σ^{54}, $n_{\sigma^{54}}$, was set as 110 molecules per cell (Maeda et al., 2000), the total number of σ^{38}, $n_{\sigma^{38}}$, 230 molecules per cell (Table 4.2) (Jishage et al., 1996) and the total number of holoenzyme, $n_{E\sigma}$, 670 molecules per cell. Using the same procedure of the simulations during the exponential phase, it is possible to infer that σ^{70} is 249 molecules per cell, σ^{54} is 39 molecules per cell, σ^{38} is 82 molecules per cell, $E\sigma_b^{70}$ is 451 molecules per cell, $E\sigma_b^{54}$ is 71 molecules per cell and $E\sigma_b^{38}$ is 148.

Similarly to the simulations made during exponential phase, three simulations were also done, one per each strain, where to the mutant cells, the total number of σ factor which is lacking and the number of corresponding holoenzymes was as well set to 0. Between strains, the total amount of holoenzyme, $n_{E\sigma}$, does not change.

The disassociation constants are the same for the simulations during exponential phase, once these values change between promoters but not between strains and cellular growth phases.

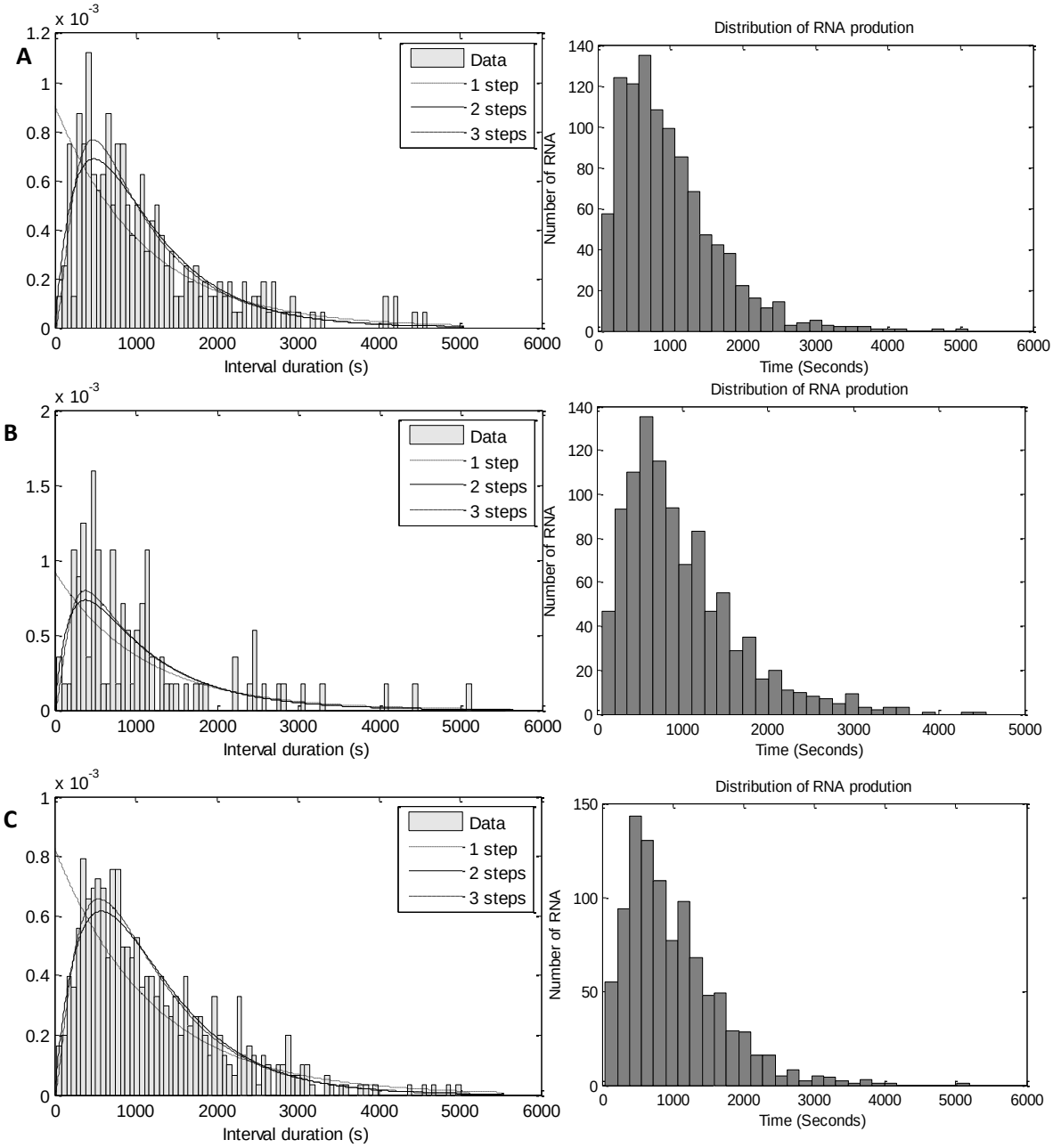

Figure 4.7: Distribution of RNA production under control of P_{tetA} during stationary phase. Histogram of the duration of the intervals between consecutive productions of RNA for the three strains (A – wild-type cells, B – mutant cells lacking σ^{38} and C – mutant cells lacking σ^{54}) under control of P_{tetA} during stationary phase resulting from in vivo measurements (left) and from simulations of the model (right).

Similar to the analysis made for the simulations during exponential phase, from Figure 4.7 and from Table 4.15 it is possible to infer that the results from the simulations of the stochastic model developed in this study are also in agreement with the *in vivo* measurements. Since the distributions of the intervals between consecutive transcription events in the three strains under control of P_{tetA} during stationary phase and the CVS value shows that this process is a

sub-Poissonian process, which is also in agreement with previous studies made for this promoter (Muthukrishnan et al., 2012). Likewise Figure 4.6, Figure 4.7 also contains the distribution of production intervals obtained with the in vivo measurements for these conditions (Figure 4.2 and left side of Figure 4.7) and the distributions of production intervals resulting from simulations of our model (right side of Figure 4.7). Comparing both distributions it is possible to observe that in vivo measurements and our simulations follow the same trend. The mean production interval ($\mu(s)$) (Table 4.15) is also in agreement with the *in vivo* measurements (Table 4.3) once the value between the three strains is similar. This similarity can be explained likewise as the *in vivo* measurements, due to P_{tetA} that is not affected by σ factor composition.

Table 4.15: Results of the stochastic model developed in this study, for the three strains under control of P_{tetA} during stationary phase. The mean production interval ($\mu(s)$), its standard deviation ($\sigma(s)$) and the CVS value are presented for wild-type cells and mutant cells lacking σ^{38} and σ^{54}.

Strains	$\mu(s)$	$\sigma(s)$	CVS
Wild-type cells	986	676	0.47
Mutant cells (lacking σ^{38})	988	678	0.47
Mutant cells (lacking σ^{54})	1004	672	0.45

4.3.1.3. Under the control of P_{BAD} during the exponential phase

The total number of molecules of σ^{70}, σ^{38}, σ^{54} and Eσ change between cellular growth phases, but do not change between promoters. Thereby, the values of these parameters for the simulations under control of P_{BAD} are the same than the ones from the simulations under control of P_{tetA}.

Three simulations were also done, one per each strain of *E. coli* and the simulations of the mutant cells were also made setting the total number of σ factor which is lacking and the number of corresponding holoenzymes to 0. Between strains, the total amount of holoenzyme, $n_{E\sigma}$, does not change.

The disassociation constant of the specific binding of the holoenzyme, $E\sigma^{70}$, to P_{BAD} is set as 900 seconds. Further, also for this promoter the specific binding occurs only around the promoter region and the specific binding is only done by $E\sigma^{70}$. Thereby, the disassociation constants of specific binding of the holoenzyme $E\sigma^{54}$ and $E\sigma^{38}$ are set with a higher number (10^{13} seconds), in order to not influence the transcription process, as for P_{tetA}. The value of the

disassociation constants of the closed complex and open complex formation were first set as the values represented in Table 4.10, respectively $k_{c_{\sigma^{70}}}$ equal to 625 seconds and $k_{o_{\sigma^{70}}}$ as 87 seconds, due to P_{BAD} is strongly affected by σ composition and the higher value should correspond to the closed complex formation and the smaller to the open complex formation (note that the smaller value is around 20 seconds and it will be set as the elongation rate). The results using $k_{c_{\sigma^{70}}}$ equal to 598 seconds were not acceptable once that P_{BAD} is not always saturated by $E\sigma^{70}$ and the prediction of this value, represented in Table 4.4, is higher than it should be. In the end, the value of $k_{c_{\sigma^{70}}}$ should be slower. Thus it was set as 250 seconds and the value of $k_{o_{\sigma^{70}}}$ was set as 51 seconds as well as the elongation rate. The disassociation constants of the closed complex formation for the other σ factors are set as 10^{11} seconds, for the same reason than the disassociation constant of the specific binding, and the disassociation constant of the open complex formation for σ^{54} and σ^{38} and the elongation rate are set as 10^{3}.

From Figure 4.8 and from Table 4.16 it is possible to infer that the results from the simulations of the stochastic model developed in this study are in agreement with the in vivo measurements. The distributions of the intervals between consecutive transcription events in the three strains under control of P_{BAD} during exponential phase and the CVS value shows that this process is a sub-Poissonian process, which is also in agreement with previous studies made for this promoter (Mäkelä et al., 2013).

Comparing the left side of Figure 4.8, which represent the distributions of production intervals obtained for the in vivo measurements, with the right side, where the distributions of production intervals of simulations of our model are represented, we concluded that both distributions follows the same trend, as for P_{tetA}. On the other hand, the mean production interval (μ(s)) (Table 4.16) is not in agreement with the *in vivo* measurements (Table 4.10), once the values are similar to all the strains and it was expected that mutant cells lacking σ^{54} needed less time to produce RNA once this σ factor does not initiate P_{BAD} transcription.

Table 4.16: Results of the stochastic model here developed for the three strains under control of P_{BAD} during exponential phase. The mean production interval (μ(s)), its standard deviation (σ(s)) and the CVS value are presented for wild-type cells and mutant cells lacking σ^{38} and σ^{54}.

Strains	μ(s)	σ(s)	CVS
Wild-type cells	773	691	0.79
Mutant cells (lacking σ^{38})	782	715	0.84
Mutant cells (lacking σ^{54})	802	714	0.79

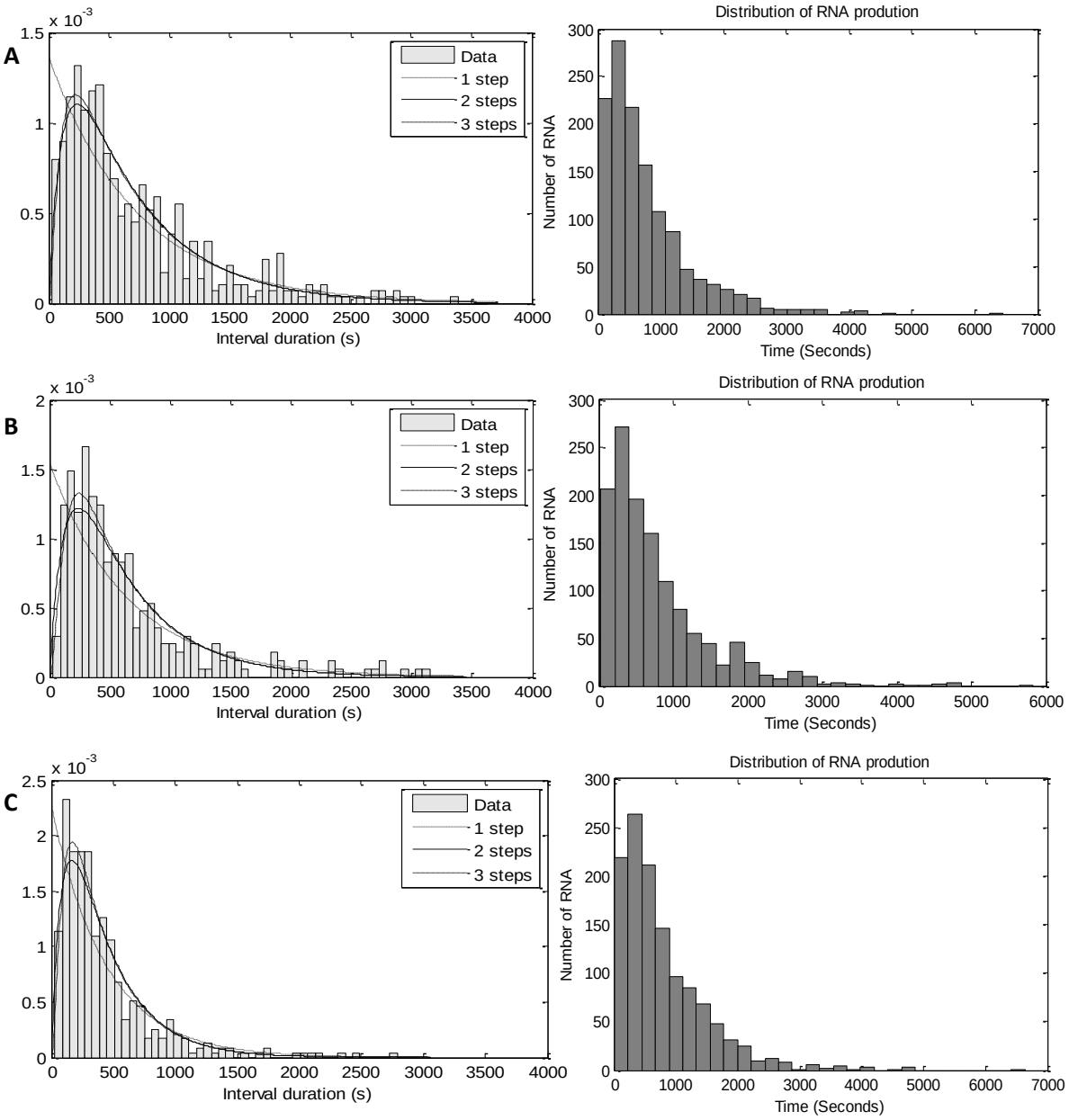

Figure 4.8: Distribution of RNA production under control of P_{BAD} during exponential phase. Histogram of the duration of the intervals between production of RNA for the three strains (A – wild-type cells, B – mutant cells lacking σ^{38} and C – mutant cells lacking σ^{54}) under control of P_{BAD} during exponential phase resulting from in vivo measurements (left) and from simulations of the model (right) developed in the aim of this study.

4.3.1.4. Under control of P_{BAD} during stationary phase

Once the value of the total number of molecules of σ^{70}, σ^{38}, σ^{54} and Eσ does not change between promoters, these parameters are set as the simulations under P_{tetA} during exponential phase. Similarly to the others simulations, here are also done three simulations following the same procedure.

The disassociation constants are the same for the simulations during exponential phase, once these values changes between promoters but not between cellular growth phases, therefore, these values will not be repeated in this sub-chapter.

From Figure 4.9 and from Table 4.17 it is possible to infer that the results from the simulations of the stochastic model developed in this study are in agreement with the in vivo measurements during stationary phase. The distributions of the intervals between consecutive transcription events in the three strains under control of P_{BAD} and the CVS value shows that this process is a sub-Poissonian process which is also in agreement with previous studies made for this promoter (Mäkelä et al., 2013).

It is also compared the distributions of production intervals obtained with *in vivo* measurements (left side of Figure 4.9) with the distributions obtained with our simulations (right side of Figure 4.9) and observe that both distribution follow the same trend.

The *in vivo* results for the mean production interval (μ(s)) (Table 4.5) suggest that the three strains need more time to complete a transcription initiation event once initiated comparing with the results for the stochastic simulations (Table 4.17). However, for both results, the strain which needs more time is the wild-type cells and the strain which complete a transcription initiation event quickly is the mutant cells lacking σ^{38}. As for the *in vivo* measurements (Table 4.5), when going from exponential growth phase to stationary phase, the difference between wild-type cells and mutant cells lacking σ^{38} becomes greater, which can be explained for an increasing in the proportion of σ^{38} in the wild-type strain. Although this difference is significantly smaller comparing with the difference obtained in the *in vivo* measurements, it is possible to conclude that P_{BAD} is affected by σ factor composition.

Table 4.17: Results of the stochastic model here developed for the three strains under control of P_{BAD} during stationary phase. The mean production interval (μ(s)), its standard deviation (σ(s)) and the CVS value are presented for wild-type cells and mutant cells lacking σ^{38} and σ^{54}.

Strains	μ(s)	σ(s)	CVS
Wild-type cells	861	748	0.76
Mutant cells (lacking σ^{38})	842	727	0.74
Mutant cells (lacking σ^{54})	853	724	0.72

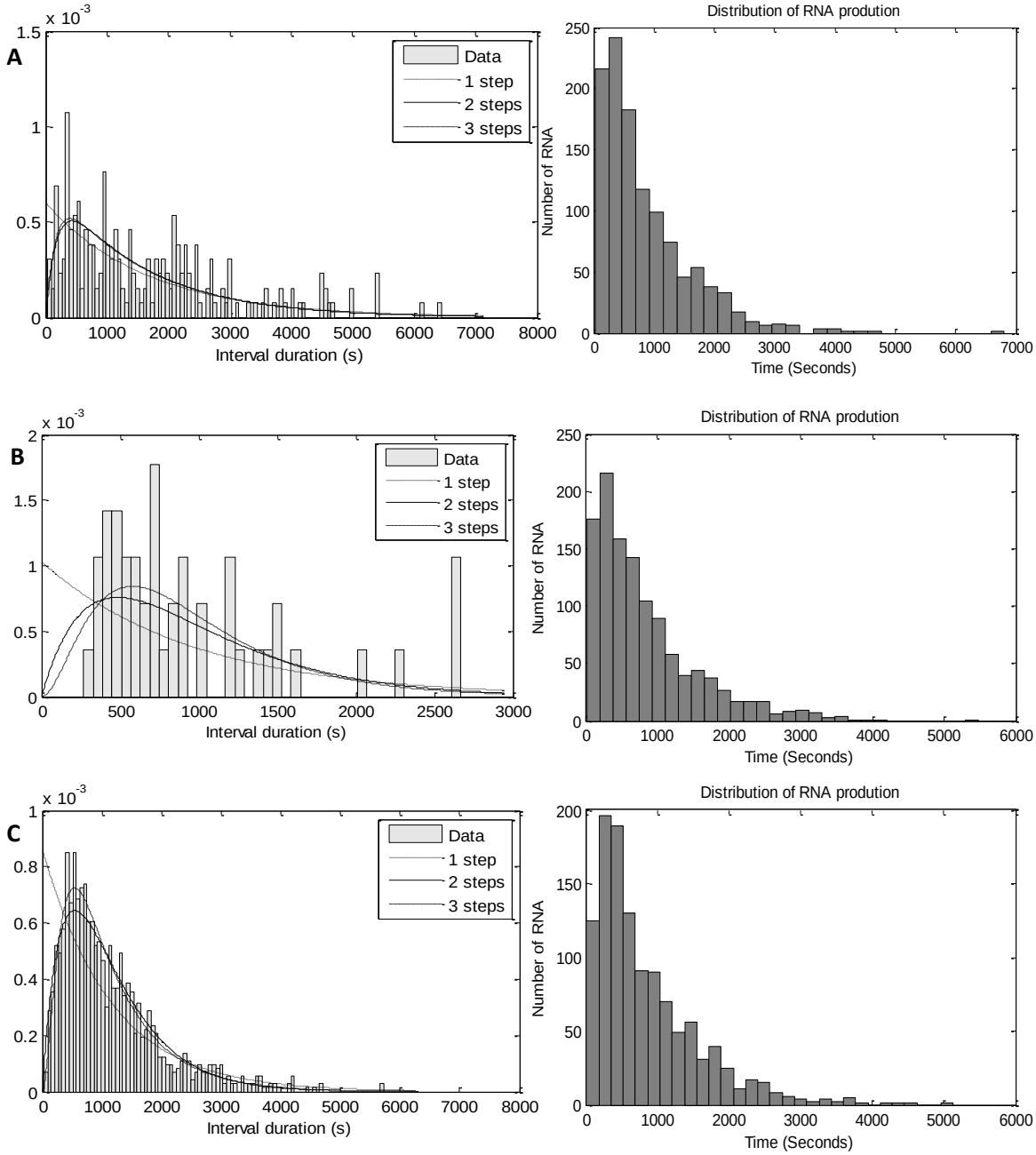

Figure 4.9: Distribution of RNA production . Histogram of the duration of the intervals between consecutive productions of RNA for the three strains (A – wild-type cells, B – mutant cells lacking σ^{38} and C – mutant cells lacking σ^{54}) under control of P_{BAD} during stationary phase resulting from in vivo measurements (left) and from simulations of the model (right).

5. Conclusion

We studied the potential differences in dynamics of transcription between mutant cells lacking σ^{54} or lacking σ^{38} and wild-type cells under the same, optimal growth conditions. This study of transcription dynamics dependence on σ factors was performed under control of two promoters, P_{tetA} and P_{BAD}, during two cellular growth phases, in order to determine if existing differences in RNA production kinetics between mutant cells are solely σ factor-dependent (which differ with the cell phase and mutations) or are also promoter-dependent.

For this study, we did the analysis of the time series obtained in our laboratory for each strain during exponential and stationary phase, by a semi-automatic method (Kandhavelu et al., 2011). For all the experiments, assuming that transcription initiation consists of a sequence of exponentially distributed steps, it was inferred the number of steps, as well as the duration of the underlying rate-limiting steps. We also developed a stochastic model which takes into account the presence of σ factors. Further, as transcription by RNAp takes some time, our model contains these time delays, being a delayed stochastic model. Thereby, the model developed includes explicitly the steps of transcription initiation, as well as the formation of the closed complex and its isomerization, which leads to the open complex formation (Buc & McClure, 1985) and the elongation process. It further contains the reactions of the translation process, namely the formation of proteins and the time needed for that.

From this model, it is extracted the dynamics of transcription for all the simulations using the delayed stochastic simulation algorithm. Then we compare the *in vivo* dynamics with the ones from the simulations of the model, developed in this study.

From the analysis of the mean duration to complete a transcription initiation event once initiated of the experiments made under control of P_{tetA} during exponential and stationary phase, we conclude that this promoter is not affected by σ factor composition, due the dynamics of transcription almost does not change between strains during the two growth phase.

On the other hand, for P_{BAD}, during both cellular growth phases, there are significant differences of the mean production intervals between strains. As such, we conclude that the dynamics of transcription from this promoter is affected by the σ factor population composition in the cells. Moreover, as for one promoter there are significant differences in RNA production kinetics between strains and for the other there are not, we conclude that the dynamics of transcription initiation is promoter-dependent.

Relatively to the shape of the distributions of the intervals obtained either with the *in vivo* measurements or with the simulations of the model, for all the strains under control of both promoters during the two cellular growth phase, it is conclude that the distributions are not exponential-like, which is consistent with the literature (Mäkelä et al., 2013; Muthukrishnan et al., 2012). Further, the process of RNA production under the control of P_{tetA} or P_{BAD} is sub-Poissonian also in agreement with previous studies.

Finally, from the data, it is possible to infer the number and duration of the sequential steps in transcription initiation by maximum-likelihood from the distribution of intervals between production of consecutive RNA molecules and it is concluded that there are three rate-limiting steps in transcription initiation by P_{tetA} and by P_{BAD} for all the strains during exponential and stationary phase. However, the steps durations differ significantly.

The results from our model are in agreement with the in vivo measurements, when analysing the dynamics per each strain under control of P_{tetA} during both growth phases. Under control of P_{BAD} there are some differences between our model and *in vivo* measurements. Future research is required to explain such differences.

The results of the analysis of the *in vivo* measurements of tagged RNA molecules in wild-type cells and mutant cells (lacking σ^{54} and σ^{38}) during exponential phase under control of the P_{tetA} and P_{BAD} were presented in conference entitled Views into Nuclear Function, Patras, Greece, on 11-13 September 2014 [2].

7. References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2008). *Molecular Biology of the Cell*. (M. Anderson & S. Granum, Eds.) (Fifth Edit., pp. 329–410). New York: Garland Science.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., … Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, *2*, 2006.0008. doi:10.1038/msb4100050

Bernstein, J. A., Khodursky, A. B., Lin, P.-H., Lin-Chao, S., & Cohen, S. N. (2002). Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(15), 9697–702. doi:10.1073/pnas.112318199

Buc, H., & McClure, W. R. (1985). Kinetics of open complex formation between Escherichia coli RNA polymerase and the lac UV5 promoter. Evidence for a sequential mechanism involving three steps. *Biochemistry*, *24*(11), 2712–2723. doi:10.1021/bi00332a018

Crick, F. (1970, August). Central Dogma of Molecular Biology. *Nature*, *227*, 561–564.

Elowitz, M. B., Levine, A. J., Siggia, E. D., & Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science (New York, N.Y.)*, *297*(5584), 1183–6. doi:10.1126/science.1070919

Fusco, D., Accornero, N., Lavoie, B., Shenoy, S. M., Blanchard, J. M., Singer, R. H., & Bertrand, E. (2003). Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Current Biology : CB*, *13*(2), 161–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12546792

Gillespie, D. T. (1977). Exact Stochastic Simulation of Coupled Chemical Reactions. *The Journal of Physical Chemistry*, *81*(25), 2340–2361.

Golding, I., & Cox, E. C. (2004). RNA dynamics in live Escherichia coli cells. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(31), 11310–5. doi:10.1073/pnas.0404443101

Golding, I., Paulsson, J., Zawilski, S. M., & Cox, E. C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell*, *123*(6), 1025–1036. doi:10.1016/j.cell.2005.09.031

Greive, S. J., & von Hippel, P. H. (2005). Thinking quantitatively about transcriptional regulation. *Nature Reviews. Molecular Cell Biology*, *6*(3), 221–32. doi:10.1038/nrm1588

Grigorova, I. L., Phleger, N. J., Mutalik, V. K., & Gross, C. a. (2006). Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(14), 5332–7. doi:10.1073/pnas.0600828103

Gruber, T. M., & Gross, C. a. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. *Annual Review of Microbiology*, *57*, 441–66. doi:10.1146/annurev.micro.57.030502.090913

Harada, Y., Funatsu, T., Murakami, K., Nonoyama, Y., Ishihama, a, & Yanagida, T. (1999). Single-molecule imaging of RNA polymerase-DNA interactions in real time. *Biophysical Journal*, *76*(2), 709–15. doi:10.1016/S0006-3495(99)77237-1

Huh, D., & Paulsson, J. (2011a). Non-genetic heterogeneity from stochastic partitioning at cell division. *Nature Genetics*, *43*(2), 95–100. doi:10.1038/ng.729

Huh, D., & Paulsson, J. (2011b). Random partitioning of molecules at cell division. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(36), 15004–9. doi:10.1073/pnas.1013171108

Ishihama, A. (2000). Functional Modulation of Escherichia Coli RNA Polymerase. *Annual Review of Microbiology*, *54*, 499–518.

Jishage, M., & Ishihama, A. (1995). Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli : intracellular levels of sigma 70 and sigma 38 . *Journal of Bacteriology*, *177*(23).

Jishage, M., Iwata, a, Ueda, S., & Ishihama, a. (1996). Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli: intracellular levels of four species of sigma subunit under various growth conditions. *Journal of Bacteriology*, *178*(18), 5447–51. Retrieved from

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=178365&tool=pmcentrez&rendertype=abstract

Kandhavelu, M., Lloyd-Price, J., Gupta, A., Muthukrishnan, A.-B., Yli-Harja, O., & Ribeiro, A. S. (2012). Regulation of mean and noise of the in vivo kinetics of transcription under the control of the lac/ara-1 promoter. *FEBS Letters*, *586*(21), 3870–5. doi:10.1016/j.febslet.2012.09.014

Kandhavelu, M., Mannerström, H., Gupta, A., Häkkinen, A., Lloyd-Price, J., Yli-Harja, O., & Ribeiro, A. S. (2011). In vivo kinetics of transcription initiation of the lar promoter in Escherichia coli. Evidence for a sequential mechanism with two rate-limiting steps. *BMC Systems Biology*, *5*(1), 149. doi:10.1186/1752-0509-5-149

Lloyd-Price, J., Gupta, A., & Ribeiro, A. S. (2012). SGNS2: A Compartmentalized Stochastic Chemical Kinetics Simulator for Dynamic Cell Populations. *Bioinformatics (Oxford, England)*, *28*(22), 3004–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23014631

Loewen, P. C., & Hengge-aronis, R. (1994). THE ROLE OF THE SIGMA FACTOR SIGMA s (KatF) IN BACTERIAL GLOBAL REGULATION. *Microbiol.*, 53–80.

Lutz, R., Lozinski, T., Ellinger, T., & Bujard, H. (2001). Dissecting the functional program of Escherichia coli promoters: the combined mode of action of Lac repressor and AraC activator. *Nucleic Acids Research*, *29*(18), 3873–81. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=55909&tool=pmcentrez&rendertype=abstract

Maeda, H., Fujita, N., & Ishihama, a. (2000). Competition among seven Escherichia coli sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Research*, *28*(18), 3497–503. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=110723&tool=pmcentrez&rendertype=abstract

Mäkelä, J., Kandhavelu, M., Oliveira, S. M. D., Chandraseelan, J. G., Lloyd-Price, J., Peltonen, J., … Ribeiro, A. S. (2013). In vivo single-molecule kinetics of activation and subsequent

activity of the arabinose promoter. *Nucleic Acids Research*, *41*(13), 6544–52. doi:10.1093/nar/gkt350

McAdams, H. H., & Arkin, a. (1999). It's a noisy business! Genetic regulation at the nanomolar scale. *Trends in Genetics : TIG*, *15*(2), 65–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10098409

McAdams, H. H., & Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(3), 814–9. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=19596&tool=pmcentrez&rendertype=abstract

McClure, W. R. (1980). Rate-limiting steps in RNA chain initiation. *Proceedings of the National Academy of Sciences of the United States of America*, *77*(10), 5634–8. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=350123&tool=pmcentrez&rendertype=abstract

Merrick, M. J. (1993), In a class of its own — the RNA polymerase sigma factor σ;^{54} (σ^N). Molecular Microbiology, 10: 903–909. doi: 10.1111/j.1365-2958.1993.tb00961.x

Mooney, R. A., Darst, S. a, & Landick, R. (2005). Sigma and RNA polymerase: an on-again, off-again relationship? *Molecular Cell*, *20*(3), 335–45. doi:10.1016/j.molcel.2005.10.015

Muthukrishnan, A.-B., Kandhavelu, M., Lloyd-Price, J., Kudasov, F., Chowdhury, S., Yli-Harja, O., & Ribeiro, A. S. (2012). Dynamics of transcription driven by the tetA promoter, one event at a time, in live Escherichia coli cells. *Nucleic Acids Research*, *40*(17), 8472–83. doi:10.1093/nar/gks583

Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., & van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nature Genetics*, *31*(1), 69–73. doi:10.1038/ng869

Paulsson, J. (2004). Summing up the noise in gene networks. *Nature*, *427*(6973), 415–8. doi:10.1038/nature02257

Peabody, D. S. (1993). The RNA binding site of bacteriophage MS2 coat protein. *The EM*, *12*(2), 595–600.

Peabody, D. S. (1997). Role of the coat protein-RNA interaction in the life cycle of bacteriophage MS2. *Molecular & General Genetics : MGG*, *254*(4), 358–64. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9180688

Peccoud, J., & Ycart, B. (1995). Markovian Modelling of Gene Product Synthesis.

Ribeiro, A. S., & Lloyd-Price, J. (2007). SGN Sim, a Stochastic Genetic Networks Simulator. *Bioinformatics (Oxford, England)*, *23*(6), 777–779. doi:10.1093/bioinformatics/btm004

Ribeiro, A., Zhu, R., & Kauffman, S. a. (2006). A general modeling strategy for gene regulatory networks with stochastic dynamics. *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology*, *13*(9), 1630–1639.

Ruusuvuori, P., Aijö, T., Chowdhury, S., Garmendia-Torres, C., Selinummi, J., Birbaumer, M., … Yli-Harja, O. (2010). Evaluation of methods for detection of fluorescence labeled subcellular objects in microscope images. *BMC Bioinformatics*, *11*, 248. doi:10.1186/1471-2105-11-248

Samoilov, M., Arkin, A., & Ross, J. (2002). Signal Processing by Simple Chemical Systems. *The Journal of Physical Chemistry A*, *106*(43), 10205–10221. doi:10.1021/jp025846z

Shaevitz, J. W., Abbondanzieri, E. a, Landick, R., & Block, S. M. (2003). Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature*, *426*(6967), 684–7. doi:10.1038/nature02191

Shepherd, N., Dennis, P., & Bremer, H. (2001). Cytoplasmic RNA Polymerase in Escherichia coli. *Journal of Bacteriology*, *183*(8), 2527–34. doi:10.1128/JB.183.8.2527-2534.2001

Shingler, V. (1996, September). MicroReview Signal sensing by s54 -dependent regulators : derepression as a control mechanism. *Molecular Microbiology*, *19*, 409–416.

Skinner, G. M., Baumann, C. G., Quinn, D. M., Molloy, J. E., & Hoggett, J. G. (2004). Promoter binding, initiation, and elongation by bacteriophage T7 RNA polymerase. A single-

molecule view of the transcription cycle. *The Journal of Biological Chemistry*, *279*(5), 3239–44. doi:10.1074/jbc.M310471200

Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M., & Elowitz, M. B. (2006). An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*, *440*(7083), 545–50. doi:10.1038/nature04588

Suh, W. C., Leirmo, S., Record, M. T., & Jr. (1992, February). Roles of Mg2+ in the Mechanism of Formation and Dissociation of Open Complexes between Escherichia coli RNA polymrase and the Pr Promoter: Kinetic Evidence for a Second Open Complex Requiring Mg2+. *Biochemistry*, 7815–7825.

Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., … Xie, X. S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science (New York, N.Y.)*, *329*, 533–538. doi:10.1126/science.1188308

Wösten, M. M. S. M. (1998). Eubacterial sigma-factors. *FEMS Microbiology Reviews*, *22*, 127–150.

Website:

[1] http://amolecularmatter.tumblr.com/post/21213985951/transcription-in-prokaryotes, consulted in 10/07/2014.

Publication:

[2] V Kandavalli, H Tran, J Chandraseelan, C Ferreira and AS Ribeiro (2014). The role of different σ factors in the dynamics of transcription during the exponential growth phase. Proceedings of the Symposium on Views into Nuclear Function. Patras, Greece, Sept 11–13, 2014.