TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Anantha-Barathi Muthukrishnan

Studies of the Plasticity of Transcription in *Escherichia coli* using Single-Molecule, *in vivo* Detection Techniques

Julkaisu 1238 • Publication 1238

Tampere 2014

Tampereen teknillinen yliopisto. Julkaisu 1238
Tampere University of Technology. Publication 1238

Anantha-Barathi Muthukrishnan

Studies of the Plasticity of Transcription in *Escherichia coli* using Single-Molecule, *in vivo* Detection Techniques

Thesis for the degree of Doctor of Philosophy to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB111, at Tampere University of Technology, on the 16th of September 2014, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2014

ISBN 978-952-15-3344-0 (printed)
ISBN 978-952-15-3360-0 (PDF)
ISSN 1459-2045

Abstract

The phenotypic characteristics of living organisms are shaped by the interactions of the genotype with the environment. In their lifetime, organisms are subject to various environmental changes, some of which are stressful. They cope up with these by means of phenotypic plasticity. In prokaryotes, this plasticity is achieved mostly by rapid adaptations of the gene expression profile. To better understand this, it is critical to study the mechanisms by which these adaptations are implemented.

In *Escherichia coli*, transcription initiation is the first and most regulated step in gene expression. *In vitro* studies suggest that this is a complex, sequential process. Its rate-limiting steps regulate both the rate and the fluctuations of RNA production. These then determine the protein numbers and, thus, the cellular phenotype.

In this work, we make use of state-of-the-art techniques in microscopy imaging, image processing and molecular probing to perform a quantitative analysis of the *in vivo* dynamics of transcription initiation in different environments in the prokaryotic model organism, *E. coli*. From the measurements, we characterize the plasticity of this process. For this, we used MS2-GFP fluorescent tagging of mRNA that allows detection of single mRNA molecules with confocal microscopy, shortly after their production. We also developed a tool to automatically track cell lineages in a time-lapse movie, and extract the spatiotemporal distribution of fluorescently tagged molecules in individual cells.

From the analysis of the results, we show that, *in vivo*, the process of transcription initiation in *E. coli* is multi-stepped, as *in vitro* measurements had previously suggested. Also, the kinetics of each step can be independently controlled by different regulatory molecules. Further, the number and timing of the rate-limiting steps are affected by physiological changes that occur in cells when subject to changing environmental conditions. We conclude that the phenotypic plasticity of *E. coli* arises, partially, from the plasticity of the kinetics of the rate-limiting steps in transcription initiation.

Preface

Foremost, I would like to express my sincere gratitude to my supervisor, Assistant Professor & Adjunct Professor Andre. S. Ribeiro, for his constant support, guidance and motivation throughout the research work for this thesis. I truly thank him for providing this opportunity to work in a multi-disciplinary field. I would like to thank my instructor, Professor Olli Yli-Harja for his motivation and encouragement during this thesis work. Furthermore, I wish to thank Professor Isabel Gordo and Assistant Professor Shenbagarathai Rajaiah for reviewing the manuscript of this thesis and for providing insightful comments.

I wish to thank Assistant Professor Jose M. Fonseca, New University of Lisbon, Portugal, for his support in the collaborative research work. Also, I wish to thank Assistant Professor Meenakshisundaram Kandhavelu, for his support during the initial stage of this work. I wish to express my sincere thanks to all co-authors and members of the Laboratory of Biosystem Dynamics, for their inspiring collaboration. Especially, to Jason Lloyd-Price, Jarno Mäkelä and Antti Häkkinen, for the fruitful discussions and great help in fulfilment of this work.

I wish to express my sincere gratitude to the TUT President's Doctoral Programme and all other funding bodies for providing financial support for this thesis work. This work was carried out in the Department of Signal Processing**,** Tampere University of Technology, Finland. I wish to thank all the members of the Faculty for providing immense support. My special thanks to the department secretary Virve Larmila, lab manager Juha Peltonen and the coordinators Elina Orava and Ulla Siltaloppi, for their kind support during my studies here.

I sincerely thank my parents, R. Tamilselvi and G. P. Muthukrishnan, for all their love and moral support. Especially, to my mother, for her great support and motivation in all my pursuits. Also, I wish to thank my sister, friends and my in-laws for their love.

I am deeply grateful to my husband Bhupesh Nagarajan, for his unconditional love and immense support throughout this period. Finally, I thank my little daughter, Indira Bhupesh for her patience, understanding and great co-operation during this period.

Tampere, April 2014

Anantha-Barathi Muthukrishnan

List of abbreviations

Frequently used abbreviations are presented below.

aTc	anhydrotetracycline
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
FISH	Fluorescence *in situ* hybridization
FRET	Fluorescence resonance energy transfer
GFP	Green Fluorescent Protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
mRNA	messenger RNA
MS2	Bacteriophage MS2
ppGpp	Guanosine tetra phosphates
qPCR	Quantitative Polymerase chain reaction
RNA	Ribonucleic acid
RNAp	RNA polymerase
TSS	Transcription start site

Contents

List of publications

This thesis is based on the following studies. Studies I, II and IV are published and are referred to in the text as "Publication I", "Publication II" and "Publication IV" respectively. The study submitted for review is referred to as "study III" in the text. The publications are reproduced with kind permissions from the publishers.

I. M. Kandhavelu, J. Lloyd-Price, A. Gupta,, A-B. Muthukrishnan, O. Yli-Harja, and A.S. Ribeiro, *Regulation of mean and noise of the in vivo kinetics of transcription under the control of the lac/ara-1 promoter*, FEBS Lett. 586, 3870–3875, 2012.

II. A-B. Muthukrishnan, M. Kandhavelu, J. Lloyd-Price, F. Kudasov, S. Chowdhury, O. Yli-Harja, and A.S. Ribeiro, *Dynamics of transcription driven by the tetA promoter, one event at a time, in live Escherichia coli cells*, Nucleic Acids Res. 40(17), 8472-8483. DOI:10.1093/nar/gks583, 2012.

III. A.-B. Muthukrishnan, A. Martikainen, R. Neeli Venkata, and A.S. Ribeiro, *In vivo single-molecule probing of the transcriptional dynamics of non-stress-responsive genes in stressed Escherichia coli cells*, FEBS Journal, under review.

IV. A. Häkkinen, A.-B. Muthukrishnan, A. Mora, J.M. Fonseca, and A.S. Ribeiro, *CellAging: A tool to study segregation and partitioning in division in cell lineages of Escherichia coli*, Bioinformatics 29 (13), 1708-1709. DOI: 10.1093/bioinformatics/btt194, 2013.

In this Thesis, a quantitative analysis was performed of the *in vivo* dynamics of transcription initiation in *Escherichia coli* when subject to different environments, using state-of-the-art techniques in microscopy imaging, image processing and, particularly, molecular probing. From the measurements, the plasticity of transcription was characterized. Here, we describe the contribution of each scientific paper to this effort, as well as the contribution of the author of this thesis to each of these scientific papers.

In Publication I (published in FEBS Letters, Impact Factor 3.5), it was shown, using live cell imaging and single-molecule detection techniques capable of detecting individual RNA molecules as soon as these are produced in the cells, that transcription factors are capable of regulating the number and duration of rate-limiting steps in transcription initiation. Relevantly, the work shows the quantitative effects on the

duration of these steps, which occur early in initiation, of altering the concentration of the inducers of transcription of a gene. More importantly, the work further shows that it is possible, by changing the amount of these inducers, to regulate the degree of noise in RNA production in live cells.

The author of this thesis contributed to Publication I as follows. First, the author conceived and tested, along with M Kandhavelu, what were the environmental and technical conditions necessary in order to test how different inducers affect the kinetics of transcription. Also, the author, in this regard, determined for which concentrations of inducers it would be possible to compare results from the microscopy measurements with the results from an independent method, namely quantitative polymerase chain reactions (qPCR). Next, the author designed and conducted all the qPCR measurements (and calculations associated with it), which were crucial in validating the results from the microscopy measurements. Several problems occurred when conducting the qPCR measurements, all of which were solved by the author alone. Finally, the author had an active participation in the writing of the paper (and, later on, in the revisions following the reviewers' reports). In addition, the author was solely responsible for writing section "2.2. qPCR analysis of target RNA", 6 paragraphs in the results section concerning the comparison of results from the two methods, 3 paragraphs in the discussion about this issue, and also 2 paragraphs in the introduction concerning previous studies in the topic.

In Publication II (published in Nucleic Acids Research, Impact factor 8.3), it was shown that the dynamics of transcription and, in particular, the number and duration of the rate-limiting steps in transcription initiation are strongly affected by temperature changes. From the results, this work proves, for the first time, that the kinetics of transcription of a natural gene is a sub-Poissonian process. Further, it shows for the first time that the noise in RNA production changes with small temperature changes.

The author of this thesis contributed to Publication II as follows. The author (first author of Publication II) jointly conceived the study with A. S. Ribeiro. Also, the author conceived and conducted the microscopy experiments and the conditions of these measurements. The author also assisted in image analysis. Finally, the author proposed, designed and conducted all qPCR and protein measurements, which validated all the results from live cell imaging. Further, the author actively participated in the writing of the manuscript (at least half of the manuscript was written by the author and later revised by all authors).

In Study III, (submitted to FEBS Journals, Impact factor 4.25), it is presented the first study of the in vivo transcription dynamics of genes uninvolved in stress-response pathways in cells under sub-lethal acidic shift and oxidative stress, by following with single-molecule sensitivity the activity of a probe gene encoding RNA target for MS2d-GFP. The results inform on how the dynamics of transcription is adjusted by cells to survive to stress conditions, and further, in general, they inform on the plasticity of transcription dynamics in *Escherichia coli*.

The author of this thesis, first author of Study III, contributed to it as follows. First, the author proposed the study. Second, the author jointly conceived how the study should be conducted along with A.S. Ribeiro. Next, the author designed and conducted all experiments. The author then performed the microscopy measurements, assisted by R. N. Venkata. The author designed and then conducted most of the qPCR measurements and data analysis, assisted by R. N. Venkata in some conditions. Finally, the manuscript was written jointly by the author and A.S. Ribeiro. Sections 1 to 7 of the Materials and Methods, were solely written by the author.

In Publication IV (published in Bioinformatics, Impact factor 5.3), a software tool was proposed to assist in studies making use of live cell imaging and single-molecule detection techniques capable of detecting individual RNA molecules as soon as these are produced in the cells, among other. Studies I-III included in this thesis make use of a novel method to investigate processes in live cells, named 'single-cell biology'. This method relies on the observation of individual, live cells, for significant periods of time, as well as (usually) fluorescent spots within the cells, usually individual proteins or fluorescently tagged RNA molecules. As such, in the course of this work, a tool was developed for this purpose with all the functionalities that were necessary to develop to extract signals from the data, using methods from Signal Processing. To design, develop, and evaluate this tool, multiple experiments were necessary. Further, the results of this tool needed to be evaluated by a specialist in single-cell biology.

The author of this thesis contributed to Publication IV as follows. First, the author designed and performed all the biological experiments. Second, the author decided which experiments would be useful in testing the software tool, based on their relevance for biological studies. Third, the author proposed the addition of features to the tool that would assist users with a biological background. Finally, the author actively participated in the writing of the manuscript. Aside from participating in the final revision of the entire manuscript, the author was sole responsible for writing some sections in the Methods and three sections in the supplement to this manuscript.

Chapter 1

Introduction

1.1. Background and Motivation

Living organisms are robust to changes in the environment (Stelling *et al*., 2004). They have evolved mechanisms which allow them to adapt to perturbations in their environments (Henge-Aronis, 1999; Roszak & Colwell, 1987; Sleator and Hill, 2002). Particularly, rapid adaptation is critical for microorganisms such as bacteria, as they are facing constant changes in their environment (Lopez-Maury *et al*., 2008; Ramos *et al*., 2001). Identifying how organisms achieve this robustness and adaptability is essential for our understanding of biological systems.

Several studies have contributed to our knowledge on adaptability (Kannan *et al*., 2008; Stoebel *et al*., 2009). Genome-wide studies on gene expression suggest that the dynamic response of an organism to environmental signals is achieved by tuning transcription levels (Causton *et al*., 2001; Chen *et al*., 2003; Ma & Bohnert, 2007; Murray *et al*., 2004; Gunasekara *et al*., 2008).

Transcription of DNA to RNA is the first step in gene expression. Translation is the next step where proteins are synthesized from RNA transcripts (Crick, 1970). Transcription consists of three sequential phases, namely, initiation, elongation and termination (Bai *et al*., 2006). In prokaryotes, transcription initiation, rather than degradation, is the major point of regulation for tuning gene expression levels (Bernstein *et al*., 2002; McClure, 1985). Transcription initiation is a multi-step process (Buc & McClure, 1985; McClure, 1980). The kinetics of the intermediate steps in transcription initiation such as the closed and open complex formations is rate-limiting, and therefore controls the transcription rate (Lutz *et al*., 2001; McClure, 1985).

The kinetics of the rate-limiting steps in transcription initiation is regulated by intrinsic factors such as the promoter sequence and extrinsic factors such as transcription factors (Browning and Busby, 2004; Reznikoff *et al*., 1985). Much of our

knowledge of the kinetics of transcription initiation has been derived from *in vitro* biochemical and biophysical studies, which are conducted on molecules isolated from living cells (Bertrand-Burggraf, 1984; Lutz *et al*., 2001; McClure, 1980; Buc & McClure, 1985). However, a living cell is a much more complex system, and includes dynamic interactions with the environment (Xie *et al*., 2008).

Further, in a living cell, transcription involves interactions of molecules of which there are few copies in the cell (Xie *et al*., 2008). Therefore, small fluctuations from the random biochemical events involved in the process can generate significant noise in the levels of messenger RNA (mRNA) (Kaern *et al*., 2005). These levels determine protein levels and thus affect the phenotype. Consequently, stochasticity in the kinetics of transcription can create phenotypic diversity in a population of genetically identical cells (Elowitz *et al*., 2002).

Studies at the single-cell level suggest that phenotypic diversity is advantageous for adaptation as it can enhance the fitness of a population (Maamar *et al*., 2007; Zhuravel *et al*., 2010). However, the stochasticity in transcription is masked in the population level *in vitro* measurements (Xie *et al*., 2008). Since the rate-limiting steps in transcription initiation also affect the amount of noise in the process (Ribeiro *et al*., 2010), *in vivo* studies with single-molecule sensitivity are essential to understand how the kinetics of transcription initiation regulates both the transcription rate the noise in gene expression.

Recent advancements in fluorescent live cell imaging, synthetic gene construction, models of gene expression and computational tools for data analysis have opened new possibilities for studying the *in vivo* kinetics of transcription (Golding *et al*., 2005; Kandhavelu *et al*., 2011). Fluorescent live cell imaging allows measurements of individual molecules in live cells, with high spatial and temporal resolution (Golding *et al*., 2005; Kandhavelu *et al*., 2011; Xie *et al*., 2008).

From these measurements, image analysis and signal processing tools enable the consistent and automated extraction of the relevant dynamics (Meijering, 2012). Novel stochastic models of gene expression can then be used to predict the behaviour in terms of the underlying mechanisms, to interpret experimental results, and to suggest novel hypotheses to test (McAdams & Arkin, 1997; Ribeiro *et al*., 2007; Ribeiro, 2010; Ribeiro *et al*., 2010; Roussel & Zhu, 2006). Single-molecule *in vivo* studies of transcription kinetics with this approach assist in better understanding the factors influencing the mean transcription rate, the regulation of transcriptional noise

and noise-mediated phenotypic diversity (Elowitz *et al*., 2002; Golding *et al*., 2005; Kandhavelu *et al*., 2011).

1.2. Thesis objectives

In this thesis, we study the plasticity of transcription in *Escherichia coli* using single-molecule, *in vivo* detection techniques.

In *E. coli*, the rate of transcription initiation of most promoters is regulated by more than one transcription factor (Browning and Busby, 2004). In this context, the first objective of this thesis is to study how the kinetics of RNA production and the kinetics of the underlying rate-limiting steps in transcription initiation (McClure, 1980; McClure, 1985) are affected by two different regulatory molecules. Specifically, we aimed to determine the effects of IPTG and arabinose on the kinetics of *lac/ara-1* promoter in *E. coli* DH5α-PRO (Lutz *et al*., 2001). We used different concentrations of the regulatory molecules to study their effect on the kinetics of the rate-limiting steps and noise levels in transcript production.

Previous *in vitro* studies suggest that inducers and environmental factors such as temperature and pH affect the kinetics of intermediate steps in transcription initiation (Bertrand-Burggraf, 1984; Buc & McClure, 1985; Lutz *et al*., 2001). In this context, the next objective is to determine how induction and temperature affect the kinetics of the underlying rate-limiting steps in transcription initiation, and thus control RNA production. For this, we studied the kinetics of the *tetA* promoter with and without induction in optimal and sub-optimal temperatures, in *E. coli* DH5α-PRO.

The stress response in *E. coli* includes down regulation of many non-stress related genes responsible for metabolism and proliferation (Jozefczuk *et al*., 2010). In light of this, the third objective is to study how the kinetics of the underlying rate-limiting steps in transcription initiation of non-stress related genes, and thus the kinetics of their RNA production are altered in stress conditions. For this, we studied the effects of oxidative stress and acidic shift on the kinetics of a probe gene, namely, *mRFP1-96BS*, under the control of the *lac/ara-1* promoter, in *E. coli* DH5α-PRO.

Finally, we aimed to develop a computational tool for automatic extraction of data from time-lapse fluorescent microscopy images using corresponding bright-field images. This tool was constructed for automatic data analysis of *in vivo* single-molecule measurements spanning several generations, such that gene expression

dynamics could be studied and possible temporal correlations in the cell lineages be detected.

1.3. Thesis outline

This thesis is organized as follows: Chapter 2 introduces the biological background in detail, with emphasis on the mechanisms of transcription initiation, the intermediate steps involved, and their regulation. Also introduced are the fundamental concepts of noise in prokaryotic gene expression. Next, Chapter 3 presents an overview of empirical methods used for studying transcriptional dynamics. Those used in the works included in this thesis are then presented in detail. Chapter 4 presents the computational tools used in this thesis, namely, the tools for segmentation of cells and spots, quantification of single RNA molecules, and the models used to infer the rate-limiting steps in transcription initiation. The concepts studied using the data from these tools are also presented. Finally, the conclusions and final discussion are presented in Chapter 5.

Chapter 2

Biological Background

This chapter is an overview of the biological concepts related to this thesis. These include the mechanism of gene expression in *E. coli*, the steps in the process of transcription initiation and their regulation, and the transcriptional noise and phenotypic plasticity in *E. coli* that result from the stochasticity in gene expression.

2.1 Gene expression in *Escherichia coli*

The prokaryotic bacterium, *E. coli* is one of the most studied organisms (Blattner *et al*., 1997; Faith *et al*., 2007; Lee *et al*., 2003). Thus, it presently serves as a model organism to understand the central processes of life, such as DNA replication, RNA production and regulation of gene expression (Elowitz *et al*., 2002; Golding and Cox, 2004; Mott and Berger, 2007).

DNA is the genetic material of living organisms with few exceptions (Hershey and Chase, 1952). The process by which genetic information encoded in DNA is transformed into functional units is called gene expression. In *E. coli*, this process occurs as stated by the central dogma of molecular biology (Crick, 1970). It starts with a transfer of information from double stranded DNA to single stranded RNA chains by transcription (McClure, 1985). The mRNA chains are subsequently decoded by ribosomes to synthesize proteins by translation (See Fig. 1) (Garrett, 1999).

E. coli, a prokaryote, lacks a nucleus. Thus, there is no spatial separation between transcription and translation, in contrast to eukaryotes (Brown and Doolittle, 1997). Without this separation, transcription and translation have been shown to be coupled in the bacterial cytoplasm (See Fig. 2) (Miller *et al*., 1970; Yarchuck *et al*., 1992).

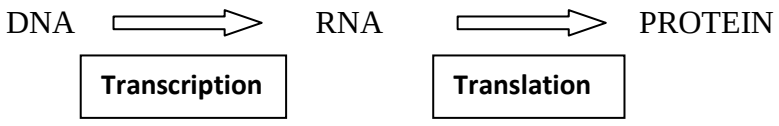

Figure 1: Flow of genetic information in *E. coli*. It follows the general transfer of information as stated in the central dogma of molecular biology (Crick, 1970). The arrows indicate the direction of the flow of information.

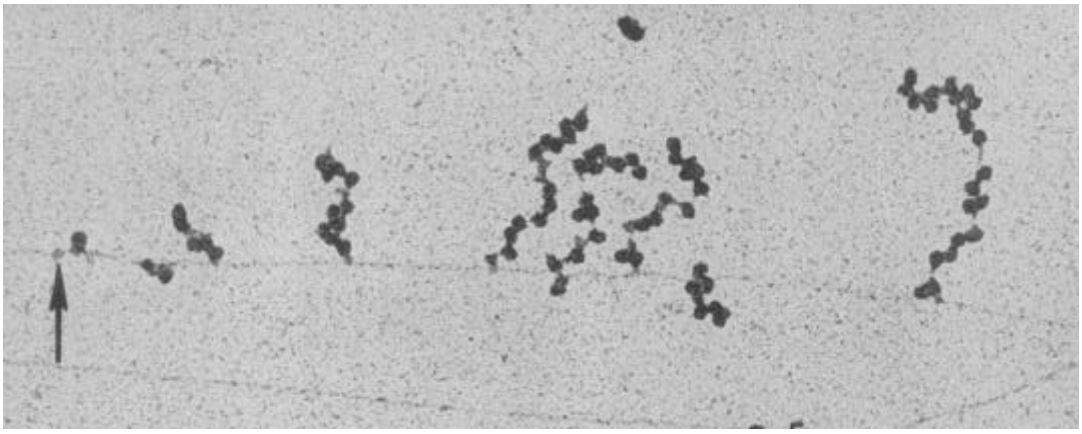

Figure 2: Electron microscope image of coupled transcription and translation in *E. coli*. The arrow indicates an RNA polymerase (RNAp) near the transcription start site (TSS). Multiple ribosomes are visible, simultaneously translating the RNA as it is transcribed (From Miller *et al*., 1970. Reprinted with permission from AAAS).

2.2 Mechanisms of Transcription

In *E. coli* there is a single copy of chromosomal DNA, which is a double stranded highly compact supercoiled structure. It contains ~ 4288 densely packed genes coding for structural and regulatory proteins (Blattner *et al*., 1997). Additionally, these organisms contain extra-chromosomal DNA, named plasmids, that code for genes that, in general, confer microbial resistance (Eliasson *et al*., 1992).

In bacteria, genes coding for proteins performing related functions are generally organized into a single transcriptional unit called an operon, under the control of a single promoter. The *lac* operon of *E. coli* provides a typical example that has a promoter, an operator, three adjacent structural genes (*lacZ, lacY and lacA*) coding for lactose metabolism, followed by a terminator region (See Fig. 3). Another gene (*lacI*), located upstream from the promoter region, is responsible for regulation of transcription of the lac operon (Jacob *et al*., 1960). In an operon, the genes are

transcribed as a polycistronic mRNA that codes for multiple polypeptides (Jacob *et al*., 1960).

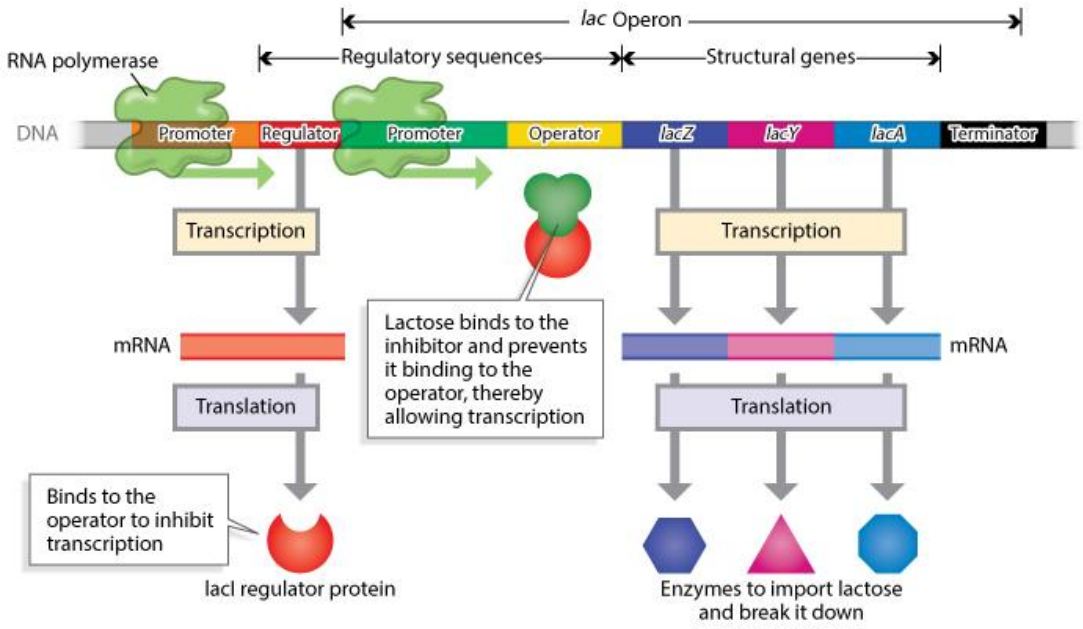

Figure 3: The *lac* operon in *E. coli*. Three lactose metabolism genes (*lacZ*, *lacY*, and *lacA*) are organized together in a cluster called the *lac* operon. The coordinated transcription and translation of the *lac* operon structural genes are controlled by a shared promoter, operator, and terminator. A *lac* regulator gene with its promoter is found upstream of the *lac* operon (Reused with permission from Nature education (Adapted from Pierce, Benjamin.*Genetics: A Conceptual Approach*, 2nd ed. All rights reserved).

Transcription occurs in three sequential steps (See Fig. 4) (Bai *et al*., 2006), namely, initiation (McClure, 1985), elongation (Uptain *et al*., 1997) and termination (Nudler *et al*., 2002). The process is initiated by the binding of RNAp enzyme to the promoter region, followed by elongation of the RNA chain until the termination site is reached (Bai *et al*., 2006). In *E. coli*, a single DNA dependent RNAp enzyme plays a central role in transcription, starting from promoter recognition to termination of the process. From molecular crystallography and FRET (Fluorescence resonance energy transfer) studies, the structure of the RNAp enzyme has been elucidated. The enzyme has multiple subunits (α, α', β, β' and σ) (Young *et al*., 2002). It exists in two forms, as holoenzyme with all subunits and as a catalytic core enzyme without the regulatory σ subunit (Murakami *et al*., 2002).

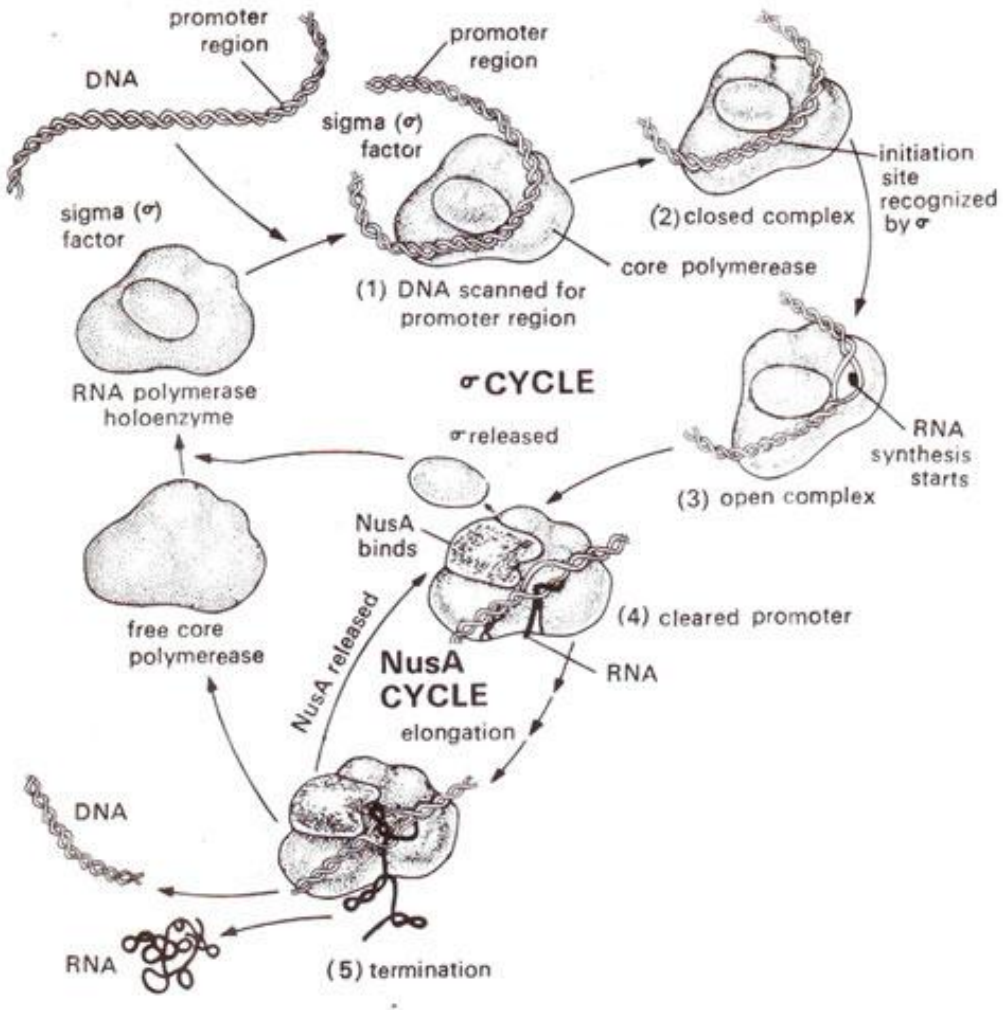

Figure 4: Mechanisms of transcription in *E. coli*. Free core RNAp forms a holoenzyme with a σ factor and searches for the promoter region. After binding to the promoter region, forming the closed complex, it initiates open complex formation by DNA unwinding. Following this, the synthesis of RNA from the template strand is initiated and the promoter is cleared for other interactions. After initiation, the σ factor is released and the NusA elongation factor assists RNAp to form a stable elongation complex. Elongation of the single stranded nascent RNA proceeds until the RNAp reaches the termination site on the DNA. At the termination site, both the RNA and RNAp are released. The RNAp is then free to initiate another transcription event (Reused with permission from Biocyclopaedia).

The promoter region is the initial target for RNAp binding to DNA. In *E. coli*, there are ~3000 promoters with most promoters having a consensus sequence at -10 (TATAAT) and -35 (TTGACA) positions relative to the TSS (+1, See Fig. 5) (Cho *et al.*, 2009; Harley and Reynolds, 1987; Mendoza-Vargas *et al.*, 2009; O'Neill, 1989).

Its highly conserved nature is attributed to their necessity for RNAp to bind to the promoter, and its subsequent activation to initiate transcription (Von Hippels *et al*., 1984; Wang and Greene, 2011).

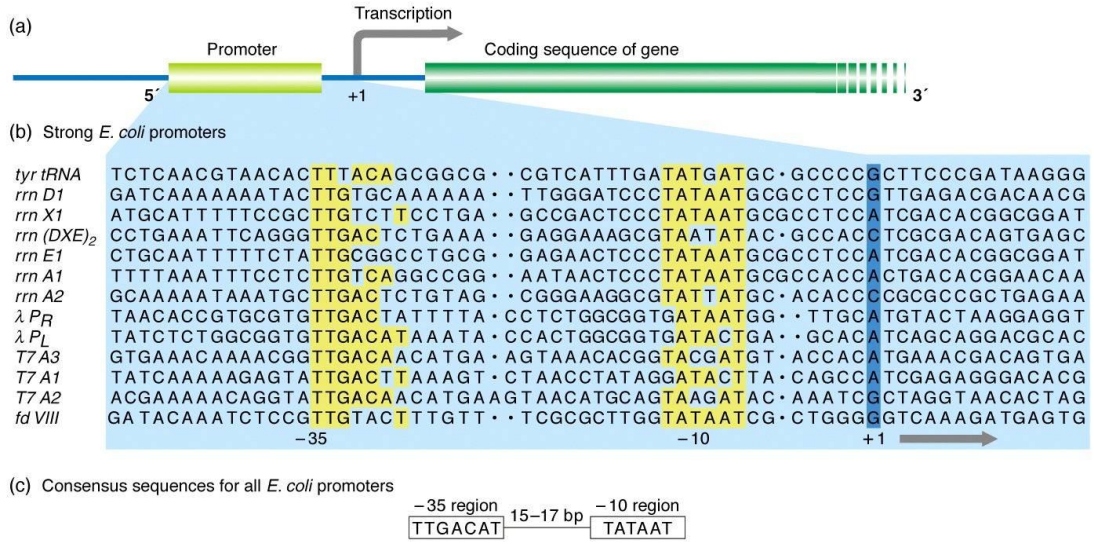

Figure 5: Promoter region of *E. coli* DNA. (a) A coding sequence of gene under the control of a promoter. +1 is the TSS at which transcription is initiated. (b) Examples for strong promoters of *E. coli*. The conserved -10 and -35 hexamers are highlighted with yellow colour. The TSS is highlighted with blue colour. (c) The consensus sequences of -10 and -35 hexamers are shown with the approximate distance between them (From Griffiths *et al*., *An Introduction to Genetic Analysis* (All rights reserved) Reused with permission from Nature education.).

The mechanism by whichRNAp reaches the promoter region is dominated by three-dimensional diffusion (Wang *et al*., 2013). It binds to the promoter as a holoenzyme and unwinds the double stranded DNA (~12 bp) to form the transcription bubble. In Fig.6, the structural interactions of RNAp with the promoter DNA to form open complex formation in transcription initiation are shown (Murakami *et al*., 2002). Then, using the coding strand as the template DNA, RNA chain elongation begins by the addition of NTPs and the formation of phosphodiester bonds between them (Uptain *et al*., 1997). Usually, the first few attempts are not successful, resulting in abortive transcripts of length 11~15 nt (Goldman *et al*., 2009). After a successful escape from the promoter, the σ factor is released from the coreRNAp, which then forms a stable elongation complex (Browning and Busby, 2004).

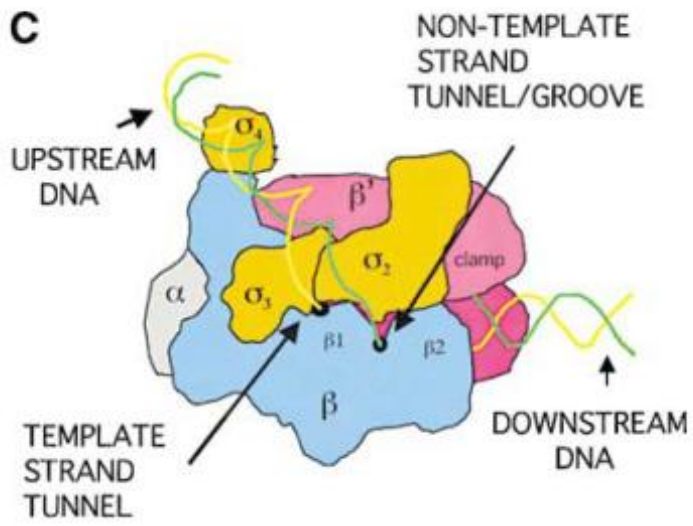

Figure 6: Structure of RNAp holoenzyme interaction with the promoter region during open complex formation in transcription initiation. The promoter DNA is double stranded in the upstream and downstream regions. The open DNA lies across the upstream face of the holoenzyme, with the majority of sequence-specific recognition carried out by the σ subunit (Murakami *et al*., 2002; Young *et al*., 2002). Reprinted from Young *et al*., (2002) with permission from Elsevier.

During elongation, the RNAp slides on the template strand until it reaches the termination site with some pausing and backtracking that may be necessary to ensure the fidelity of the transcription (Bai *et al*., 2004). At this point, the nascent RNA molecule and the RNAp are released from the DNA (Arndt & Chamberlin, 1988). In *E. coli*, termination occurs by intrinsic sequence dependent or Rho-dependent mechanisms. In intrinsic termination, the elongating RNA forms a secondary structure resulting in destabilization of the elongation complex. In Rho-dependent termination, the Rho protein detaches the RNA from the DNA while the RNAp is paused at the rho specific pause site (Nudler & Gottesman, 2002; Ciampi, 2006).

2.3 Intermediate rate-limiting steps in transcription initiation

The numbers of RNA and protein molecules are tightly controlled in a cell under normal conditions. This control appears to be based on the ability to control the frequency with which a gene can be expressed (Bernstein *et al*., 2002), rather than the degradation process. *In vitro* studies and models of transcription suggest that

transcription initiation is a sequential process requiring a series of chemical reactions to occur before RNA chain initiation (McClure, 1985, Bertrand Burggraff *et al*., 1984, Chamberlin, 1976). Regulation of the initiation frequency therefore involves the control of the speed at which these reactions take place (McClure, 1985).

The intermediate steps in initiation are identified mainly by two methods. Namely, the abortive initiation assay, and the *in vitro* transcription assay (Buc & McClure, 1985; McClure *et al*., 1978; McClure, 1980; Lutz *et al*., 2001). The abortive initiation assay is based on the binding of two triphosphates, ATP and UTP in an RNA chain, in the presence of a saturating amount of RNAp and promoter DNA with specific reaction conditions. ATP is always the first nucleotide, followed by UTP as ordered by equilibrium. As a result of a phosphodiester bond between these two nucleotides, pppApU and PPi are produced. In the absence of other NTPs, pppApU and PPi dissociates (thus aborting initiation) rapidly resulting in a steady state. This steady state kinetics is the basis for the abortive initiation assay (McClure *et al*., 1978).

Generally, radioactive phosphates or fluorescently labelled phosphates are used in these assays to detect product formation (McClure, 1980, Bertrand-Burggraf *et al*., 1984). From the measurements of abortive initiation assays, the kinetics of the intermediate steps in initiation was derived. The rate of open complex formation was measured by the delay to reach steady-state production of the abortive product (abortive initiation oligonucleotides) (See Fig. 7) (McClure, 1980). Further work revealed promoter-specific lag times varying from 10 s to several minutes before approaching the steady state.

Similar lag times were observed in *in vitro* transcription reactions (Bertrand-Burggraf *et al*., 1984). Since they are promoter specific, they were interpreted as the time taken for RNAp binding to form the closed complex and isomerization, which forms a catalytically active open complex. The lag time was longer when compared to the time required for elementary steps in most enzyme catalyzed reactions. Therefore, those slow steps preceding RNA chain initiation were considered as “rate-limiting” for an initiation reaction (McClure, 1980; Buc & McClure, 1985; McClure, 1985).

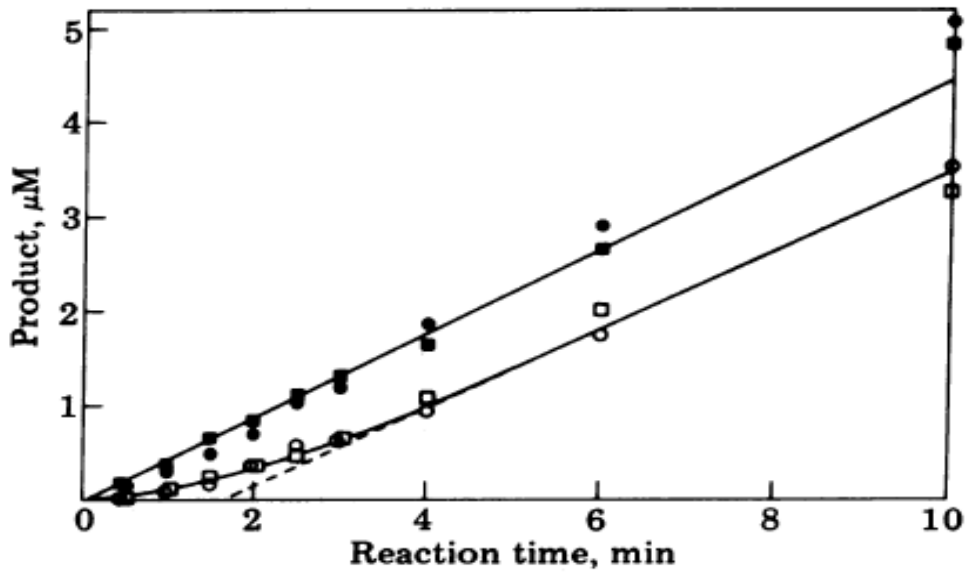

Figure 7: Measurement of open complex formation of the bacteriophage T7 D promoter. The product is pGpUpU. The formation of pGpUpU is delayed by the intermediate steps in transcription initiation, resulting in a lag time to reach steady-state production (McClure, 1980).

Also, the lag times varied with the concentration of RNAp. This dependence of the lag time on the RNAp allows the rate of binding (closed complex formation) to be distinguished from the isomerization steps resulting in the open complex (Buc & McClure, 1985). This was derived from an equation to describe the rate of formation of the active binary complex, based on the two-state model (See equation 1) (Chamberlin, 1974; Walter *et al*., 1967).

$$R+P \leftrightarrow RP_c \leftrightarrow RP_o \qquad (1)$$

In the equation 1, R stands for free RNAp, P stands for free promoter, RPc stands for closed complex, and RPo stands for open complex. The closed complex formation is the binding of the RNAp holoenzyme to the promoter region of the closed double stranded DNA. Direct evidence for the closed complex came from an electron microscope analysis that showed the binding of RNAp to the promoter DNA even at 0 °C, when it is not possible to form an open complex (Williams & Chamberlin, 1977). Open complex formation is the unwinding of the DNA ~13 bp away from the -10 element until the TSS, forming a transcriptionally active open complex. Both the

closed complex and the open complex formations are reversible processes (Buc & McClure, 1985).

In addition, a “tau plot” between the lag times and concentration of RNAp, resulted in a linear relationship (See Fig. 8), as expected if the rate of RPc formation is linearly proportional to the concentration of R. The slope in this plot yields the mean time for closed complex formation and the intercept gives the mean time for isomerization into the open complex. This method was applied to several promoters to understand the relation between lag times and transcription initiation. It was found that the strong promoters (for example, A_2) have short lag times and weak promoters (for example, D) have long lag times. In other words, promoters with high initiation frequency had short lag times. Further, they had low slope and low intercept on tau plots (McClure, 1980). From all the above results, it was concluded that there are intermediate steps in transcription initiation which are rate-limiting in transcription initiation.

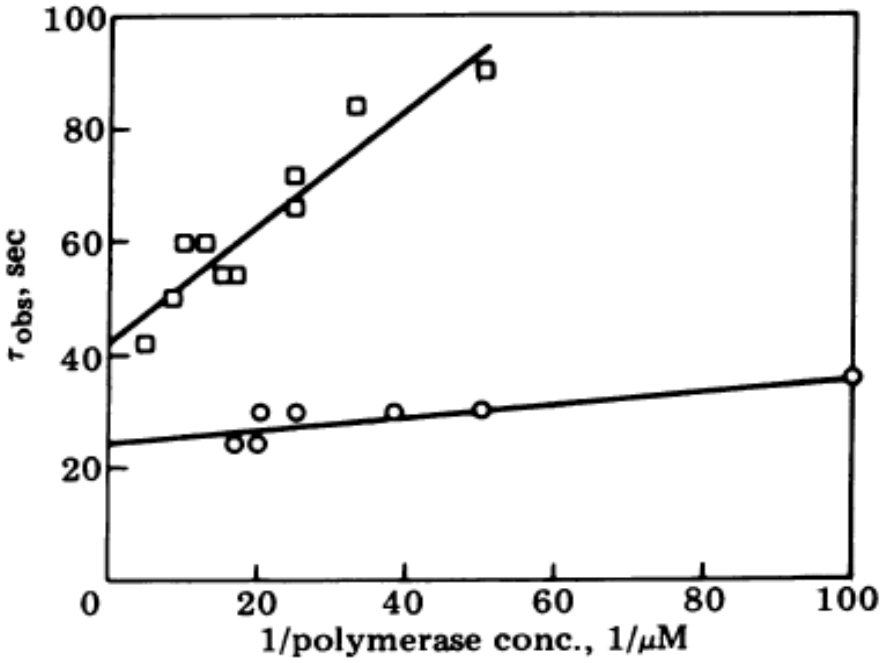

Figure 8: The tau plot for the bacteriophage T7 D and A_2 promoters. The lag times observed (τ_{obs}) for pGpUpU synthesis from the D promoter (squares) and pGpC synthesis from the A_2 promoter (circles) are plotted against the reciprocal of the RNAp concentrations used (McClure, 1980).

A modified minimal model with an additional third rate-limiting step (RP_i for isomerization) in transcription initiation was proposed by Buc & McClure (See equation 2) (Buc & McClure, 1985).

$$R+P \leftrightarrow RP_c \leftrightarrow RP_i \leftrightarrow RP_o \quad (2)$$

This isomerization step is a temperature dependent, rapid step for unstacking of DNA immediately before open complex formation. This step is rate-limiting below 20 °C for the *lacUV5* promoter. Temperature also affects the open complex formation step. Another environmental factor, pH (Bertrand-Burggraf *et al*., 1984) and structural properties such as DNA supercoiling are shown to affect the kinetics of the steps in transcription initiation (Buc & McClure, 1985). In general, there are at least two rate-limiting steps identified in reactions close to physiological conditions with a third step emerging in extreme conditions. Further, different sets of these steps become rate-limiting for different promoters under the same reaction conditions (Buc & McClure, 1985, McClure, 1985).

Further studies on the biochemical kinetics of transcription added more intermediate steps in initiation after the open complex formation. These steps include abortive initiation even in the presence of all four nucleotides, pausing of RNAp at the start site and the release of the σ factor from the holoenzyme, together condensed as the promoter clearance step, which are generally rapid and unstable (See Fig. 9). Therefore, it is difficult to detect these steps with either *in vitro* or *in vivo* methods. For this reason, a simplified model with major rate-limiting steps, such as closed complex formation and isomerization to open complex are used widely to explain the kinetics of initiation (Saecker *et al*., 2011). The following is a model of sequential steps in transcription initiation based on *in vitro* studies of several promoters (See equation 3) (Saeckar *et al*., 2011; Kandhavelu *et al*., 2011):

$$\underset{}{R+P} \overset{\text{Slow}}{\Longleftrightarrow} \underset{}{RP} \overset{\text{Rapid}}{\Longleftrightarrow} \underset{}{RP_c} \overset{\text{Rapid}}{\Longleftrightarrow} \underset{}{I_1} \overset{\text{Slow}}{\Longleftrightarrow} \underset{}{I_2} \overset{\text{Rapid}}{\Longleftrightarrow} \underset{}{I_3} \overset{\text{Rapid}}{\Longleftrightarrow} \underset{}{RP_o} \overset{\text{Rapid}}{\Longleftrightarrow} \underset{}{RP_{init}} \quad (3)$$

I_1 to I_3 stands for intermediate steps in isomerization. The last step RP_{int} in (3) competes with abortive initiation.

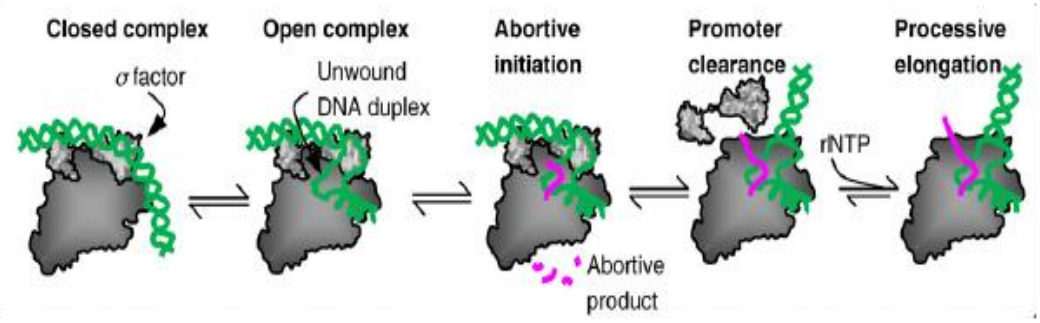

Figure 9: Intermediate steps in transcription initiation**.** The RNAp holoenzyme binds tightly to the promoter and bends the DNA to form the closed complex. After forming the holoenzyme with σ factor, RNAp unwinds the the promoter region and forms the open complex. Abortive RNA products are synthesized initially until RNAp escapes the promoter region. After promoter escape step, elongation takes place. Reprinted from (Wang & Greene, 2011), with permission from Elsevier.

A detailed study of the kinetics of the rate-limiting steps in initiation demonstrated the intrinsic regulation of these steps by the promoter sequence. For this, promoters were derived from the *lac* promoter of *E. coli* by modifying the bases in the highly conserved -10 and -35 hexamers and the distance between them. The CRP binding region was replaced by the AraC binding region to avoid pleiotropic effects of CRP mediated regulation. In addition, promoter strengths were measured from the *in vivo* luciferase activities of cells harbouring plasmids with similar promoter constructs. The promoter strengths of P_{lar} variants differed with range of over 600 fold, when changes are made in the consensus regions of the promoter (Lutz *et al*., 2001).

In the same study, using radioactive labelling and filter binding techniques, closed complex formation between the RNAp and the promoter was measured in the presence and the absence of repressor (lacI) and activator (AraC). Also, open complex formation was measured under the same conditions using the $KMnO_4$ footprinting method. In this method, the intensity of the permanganate reaction signal corresponds to the duration of the open complex state of the promoter. From these methods, the rates of three major rate-limiting steps were derived. In general, in the presence of lacI repression, the rate of closed complex formation is the most rate-limiting step whereas the activator, AraC affects all three major rate-limiting steps in transcription initiation. In particular, it accelerates the stable closed complex formation and the isomerization

from closed complex to open complex state (Lutz *et al*., 2001). A recent study based on a delayed stochastic model of gene expression also suggests that, by regulating the kinetics of the closed and open complex formations, it is possible to regulate both the mean and the fluctuations in RNA numbers (Ribeiro *et al*., 2010).

Overall, the above studies suggest that the mean rate of transcription of a gene is mostly determined by the promoter sequence, and the regulatory molecules of transcription initiation, which accelerate or hinder the steps in transcription initiation. Furthermore, DNA supercoiling and environmental factors such as temperature also influence the kinetics of the steps in transcription initiation.

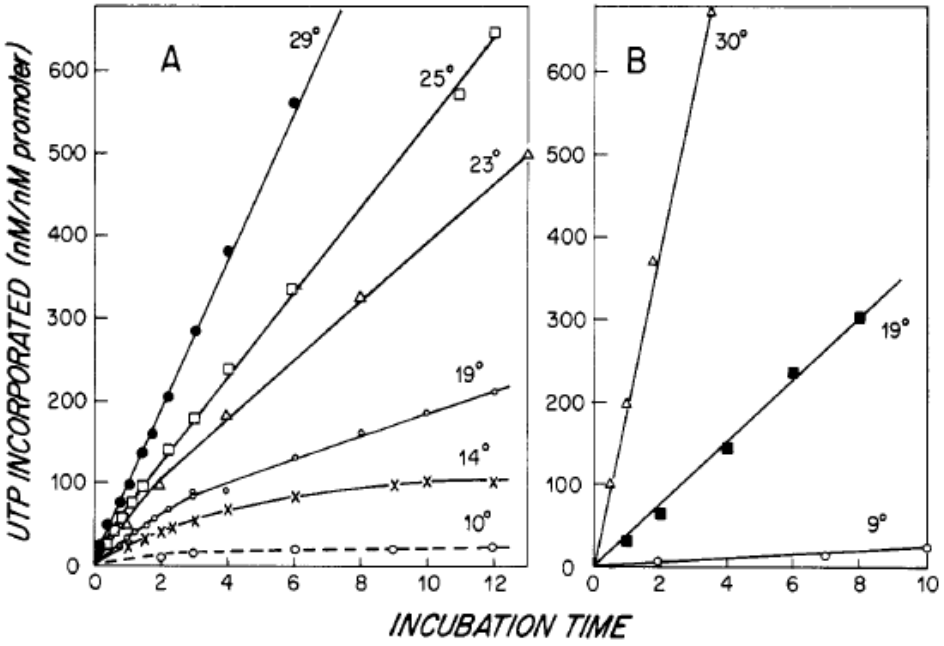

Figure 10: Temperature shift experiments performed with the RNAp-*lacUV5* promoter complex. This figure shows the effect of temperature on the rate of open complex formation by measuring the rate of UTP nucleotide incorporation in an abortive initiation reaction. In addition, it shows the effect of DNA supercoiling on the rate of open complex formation with linear promoter DNA (A), and with the promoter inserted into supercoiled plasmid DNA (B) Reprinted with permission from (Buc & McClure, 1985). Copyright (1985) American Chemical Society.

2.4 Regulation of transcription initiation

All micro-organisms undergo internal changes such as different phases of the cell cycle and also respond to external changes, such as temperature. Therefore, it is

critical for their survival to tightly regulate gene expression in order to ensure appropriate number and type of proteins depending on the internal and external signals (Lopez-Maury *et al*., 2008; McClure, 1985; Von Hippels, 1984).

Although every step in gene expression is regulated, transcription initiation is the major point of regulation (Bertrand-Burgraff *et al*., 1984; Chamberlin, 1976; Pribnow *et al*., 1979; Rosenberg & Court, 1979; Young *et al*., 2002). This is viewed as an economic strategy for an organism where there is tight spatial and dynamic coupling of transcription and translation (Browning & Busby, 2004; Yarchuk *et al*., 1992). There are several mechanisms to regulate the steps in transcription initiation in *E. coli* (See Table 1) (Browning & Busby, 2004; McClure, 1985; Reznikoff *et al*., 1985).

2.4.1 Intrinsic regulation of the promoter region

The promoter sequence determines the affinity of RNAp binding. Thus, it affects the rate of closed complex formation. Also, the DNA conformation is sequence dependent. Therefore, the promoter sequences also have an impact on the rate of open complex formation. Thus, the intrinsic promoter sequence plays a significant role in transcriptional regulation (Von Hippels, 1984).

However, these sequences can provide only a static regulation since they cannot be tuned to changing environmental conditions. Such capability is achieved by the activity of several trans-acting factors such as σ factors, repressors, activators and also small metabolites. Thus, intrinsic regulation by the promoter sequence together with extrinsic regulation by trans-acting factors determines the strength of a promoter (See Table 1) (Browning & Busby, 2004; Von Hippels, 1984).

Regulator	Target	Action
Promoter sequence	RNAp, σ Factor, TFs, ppGpp	Intrinsic regulation by specific interaction and stability
σ Factor	RNAp	Promoter specificity, DNA melting,
guanosine tetra phosphates (ppGpp)	RNAp	Stress associated up or down-regulation
DNA conformation and nucleoid associated proteins	RNAp, TFs	TFs distribution, rate of open complex formation
Repressors	Promoter, upstream and downstream, Activators	Prevents transcription by steric hindrance to RNAp, DNA looping and anti-activator activity
Activators	RNAp	Recruiting RNAp to transcription site and alteration of DNA conformation to favor its binding

Table 1: This table is a summary of the regulatory molecules, their target for regulation and their mode of regulation in transcription initiation in *E. coli* (Browning & Busby, 2004; McClure, 1985; Reznikoff *et al*., 1985).

2.4.2 Regulation by σ factors

Direct interaction σ factors with RNAp are studied using FRET and crystallography methods (Mekler *et al*., 2002; Murakami *et al*., 2002).Further studies showed that σ factor is essential for specific promoter recognition in response to the physiological state of the cell (Hengge-Aronis, 2002; Weber *et al*., 2005) Also, σ factors play role in melting of the DNA double strand to form the open complex (Chamberlin, 1976; Travers & Burgess, 1969).

There are seven types of σ factors that recognize specific promoters based on sequence information. This allows *E. coli* to regulate transcription by using alternate σ factors to initiate transcription of different sets of genes appropriate for different environmental conditions (Hengge-Aronis, 2002). The σ^{70} is the primary σ factor that is responsible for transcription of the majority of genes. Other σ factors recognize promoters which regulate a discrete set of genes encoding proteins needed for rapid adaptation to a specific environmental condition. For example, σ^{32} specifically regulates genes for heat shock response (Hengge-Aronis *et al*., 2001). Another σ factor, σ^{38} or RpoS is the major regulatory factor for global regulation of transcription providing general stress response mechanisms in *E. coli* (Battesti *et al*., 2011).

Interestingly, *in vitro* and *in vivo* studies on the selective recognition of promoters by σ^{38} and σ^{70} revealed that both σ factors recognize the -10 and -35 hexamers containing promoters. However, there is a conserved -13 C residue close to the TSS in the promoters recognized by σ^{38} (Becker & Hengge-Aronis, 2001; Landini *et al*., 2013) and a G-C rich discriminator region in the promoters for σ^{70}. These slight variations in the promoter sequences regulate the promoter selectivity of σ factors.

Recent studies on interaction of σ with RNAp suggest its role other than in transcription initiation. According to these studies, σ factor can influence the steps in elongation by temporarily or permanently associating with the elongation complex (Brodolin *et al*., 2004; Kapanidis *et al*., 2005; Mooney & Landick, 2005).

2.4.3 Regulation by small ligands

Another mechanism by which RNAp is regulated for rapid adaptation is by interaction of small ligands (Browning & Busby, 2004). For instance, under stress

conditions, ppGpp accumulate in large amounts in *E. coli* cells (Batesti *et al*., 2011). This is an alarmone that aids in bacterial adaptation (Zuo *et al*., 2013). Binding of ppGpp to RNAp, results in down-regulation of a subset of genes during the stringent response to stress, for example, genes for ribosome synthesis. Further, it destabilizes the open complex formation of the promoters having GC-rich discriminator region. This region is shown to form unstable open complexes, but this not a requirement for the down regulation of genes by ppGpp (Barker *et al*., 2001; Magnusson *et al*., 2005).

A recent study on the structural interaction of ppGpp with RNAp revealed that ppGpp binding brings the catalytic module and the nucleotide chain pathway together that inhibits the opening of the enzyme cleft, preventing transcription initiation (Ross *et al*., 2013; Zuo *et al*., 2013). Apart from this negative regulation of genes, there are also recent evidences for positive regulation by ppGpp interaction (See Fig. 12). They up-regulate genes with promoters containing AT-rich discriminator region during stress conditions, for example, genes for amino acid biosynthesis (Ross *et al*., 2013)

These findings revealed the association between σ factors and ppGpp in regulating transcription. Promoters recognized by σ^{70}, such as controlling cell growth and proliferation, are negatively regulated by ppGpp whereas promoters recognized by RpoS, controlling stress responsive genes, are positively regulated by ppGpp (Magnusson *et al*., 2005).

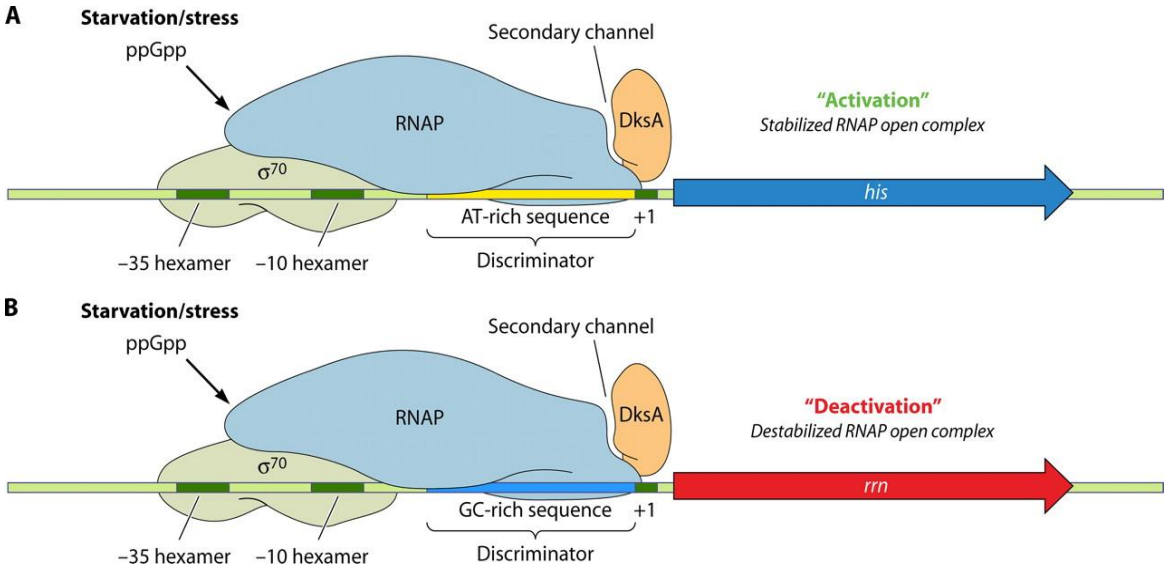

Figure 12: Regulation of RNAp interaction with promoter region by ppGpp. Activated targets such as the *E. coli* promoter for the histidine biosynthetic (his) operon typically have an AT-rich DNA sequence between the −10 hexamer and the +1 TSS, known as the discriminator region. Conversely, repressed targets such as the P1 promoter of rRNA (rrn) operons typically have a GC-rich discriminator sequence. Reprinted from (Zachary *et al*., 2010) with permission from American Society for Microbiology.

2.4.4 Regulation by DNA folding and associated proteins

The supercoiled structure of DNA in *E. coli* also plays role in regulation of transcription initiation. For example, with an increase in the supercoiling, the rate of open complex formation also increases, as DNA distortion is favored by supercoiling. But beyond some level, there is a decrease in the rate, which is due to the loss of conformation that is needed for the RNAp binding (deHaseth and Helmann, 1995). Further, increased supercoiling resulted in 40 fold increase in the lac repressor binding affinity. The degree of supercoiling is also shown to affect the binding of σ factors of and the alteration in the lag time of transcription initiation (Bertarnd Burggraf *et al*., 1984, Von Hippels, 1984).

Also, there are several nucleoid associated proteins that maintain the compact structure of the DNA, namely, the H-NS, HU proteins. The distribution of RNAp on promoters is affected by the folding of DNA by nucleoid associated proteins. For example, the H-NS protein forms extended nucleoprotein structures that completely prevent transcription by the *proU* and *bgl* promoters (McLeod & Johnson, 2001). In some cases, there is also activation of transcription by nucleoid associated proteins, for example, Fis proteins mediated regulation (Gerstell *et al*., 2003).

2.4.5 Regulation by transcription factors

The major molecular machinery that regulates transcription initiation are Transcriptional factors (TFs). There are ~250 TFs in *E. coli*, *which* bind on cis-elements of DNA, For example, lacI, AraC (Babu & Teichman, 2003). Generally a TF has two domains, one for receiving the internal or external signal and the other one for directly interacting with the DNA, resulting in transcriptional modulation. Most of the transcription factors are sequence specific and targeted to specific promoter-operator regions. But there are also TFs which can recognize multiple binding sites and regulate different sets of genes. Similarly, there is co-regulation of a single promoter by

multiple TFs (Balleza *et al*., 2009). Also, there are TFs binding to regions far from the promoter regions. Therefore, the mechanisms by which the TFs regulate transcription are both diverse and complex. They can be broadly categorized as follows:

2.4.5.1 Repression

Repression is a negative regulation of transcription by a repressor protein, which inhibits transcription initiation. There are at least three ways by which repression is executed. One simple way is to cause steric hindrance to the binding of RNAp to the promoter region. For example, the lacI repressor protein of the *lac* operon (See Fig. 3). When there is no lactose in the environment, the repressor protein binds to the promoter of the lac operon. This hinders the binding of RNAp to the promoter thereby preventing transcription initiation. In the presence of lactose, repressors bind to lactose molecules, resulting in reduced binding affinity to the promoter and subsequent release of the repressor protein from the promoter region. This facilitates the binding of the RNAp to the promoter and thus, transcription is initiated (Jacob *et al*., 1960).

In some cases, repressors bind to regions downstream to the promoter in multiple copies. This results in the formation of DNA loops which prevent binding of the RNAp to the DNA. An example of this mechanism is the regulation at the *gal* operon by GalR repressor protein (Aki *et al*., 1996). Another way to repress transcription is the binding of repressor proteins to activators which promote transcription. For example, the cytR repressor molecule interacts with the CRP protein, which acts as an activator for several operons (Van Hijum *et al*., 2009).

2.4.5.2 Activation

Activation is the positive regulation of transcription initiation. There are several mechanisms in *E. coli* to activate transcription. They are categorized as class I, class II and class III activation mechanisms. Class I and class II mechanisms involve direct interaction of activator molecules with the RNAp. An example for class I activation is the CRP protein mediated activation of the *lac* operon. This protein recruits RNAp by directly interacting with the αCTD subunit of the enzyme (Ebright *et al*., 1993). In class II activation, the RNAp is also recruited but in this case it is by interaction with the σ factor (Busby and Ebright, 1997). In Class III activation, multiple copies of

activators bind to the DNA and alter the conformation of the promoter region so as to cause the increase of the binding affinity of the RNAp to the promoter and, thus, promote transcription initiation (Sheridan *et al*., 1998).

Besides repressors and activators, there are also molecules that can act both as a repressor and as an activator. AraC protein is a typical example of such behaviour (See Fig. 13). This protein regulates the *ara* operon coding genes for arabinose metabolism. In the absence of arabinose, the AraC binds to $araI_1$ and $araO_2$ and forms a DNA loop. This looping inhibits the binding of RNAp, preventing transcription. In the presence of arabinose, AraC acts as an activator by forming AraC- arabinose complexes that promotes transcription by recruiting RNAp (Schleif *et al*., 2010).

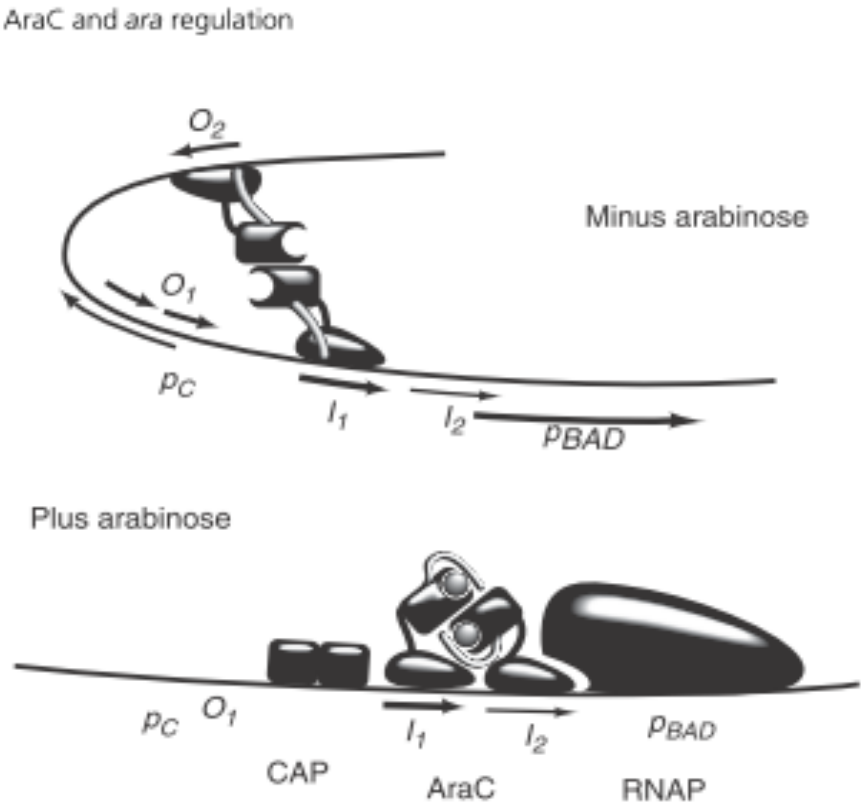

Figure 13: Regulation of transcription initiation of arabinose operon under the control of P_{BAD} promoter. I_1 and I_2 are termed half-sites as only a single subunit of AraC contacts each. O_1 consists of two half-sites and serves as an operator to the P_C promoter that controls the synthesis of AraC, whereas O_2 is a single half-site. The single CAP site regulates the divergently oriented *Pc* and the P_{BAD} promoters. In the absence of arabinose, the protein AraC binds to $araI_1$ and $araO_2$, forming a DNA loop, thereby preventing transcription initiation by RNAp . In the presence of arabinose, AraC acts as an activator by forming AraC- arabinose

complexes that promotes transcription by recruiting RNAp. Reprinted from (Schleif *et al*., 2010) with permission from John Wiley and Sons.

2.4.6 Global regulation

Transcription factors in *E. coli* can act both specifically and globally. In *E. coli*, the majority of the promoters are regulated by at least two transcription factors, one responding to the specific signal and the other for the global signal according to the cellular physiology. For example, let us consider the *lac* operon. The repressor regulates the *lac* operon based on the interaction with the specific metabolic signal, lactose whereas the CRP regulates according to the global signal. Thus the two TFs independently regulate transcription initiation so as to tune gene expression to balance both specific and global environment (Busby & Kolb *et al*., 1996). Recent studies with microarray revealed that six to seven TFs regulate 50% of the genes and around sixty TFs regulate a single promoter (Martinez-Antonio & Collado-Vides, 2003). These studies emphasize the role of transcription factors in tuning gene expression for physiological needs in a specific environment, which is necessary for adaptability and evolvability (Lopez-Maury *et al*., 2008).

2.5 Transcriptional noise and phenotypic plasticity

Despite tight regulation, transcription is a noisy process. Many of the molecules involved in transcription are present in the cell in few copies (Maheshri *et al*., 2007, Xie *et al*., 2008). Thus, even small fluctuations introduced by random biochemical events can generate significant noise in this process (Kaern *et al*., 2005). This noise in gene expression is a source of phenotypic diversity in a population of genetically identical cells (McAdams & Arkin, 1997). Stochastic chemical transformations contribute to intrinsic noise, while variation in external factors such as the number of ribosomes or transcription factors contributes as correlated or extrinsic noise (Swain *et al*., 2002). One single-cell level study on gene expression in *E. coli* measured the intrinsic and extrinsic noise levels from cells expressing dual fluorescent reporters under the control of the *lac* promoter. The intrinsic noise was measured from the differences between two fluorescent proteins within each cell. Extrinsic noise was measured from the degree of correlation between these two fluorescent protein levels in individual cells.

In strains with constitutive promoters lacking the *lac* repressor and feedback mechanism, low noise and low cell-to-cell variation was observed. In wild type lac promoters with *lac* repressor, both intrinsic and extrinsic noise increased (See Fig. 14). When the repressor activity is inhibited by adding IPTG, the low noise levels were restored. From these results it was concluded that the transcription rate and the regulatory molecules of transcription initiation can control the amplitude of noise (Elowitz *et al*., 2002).

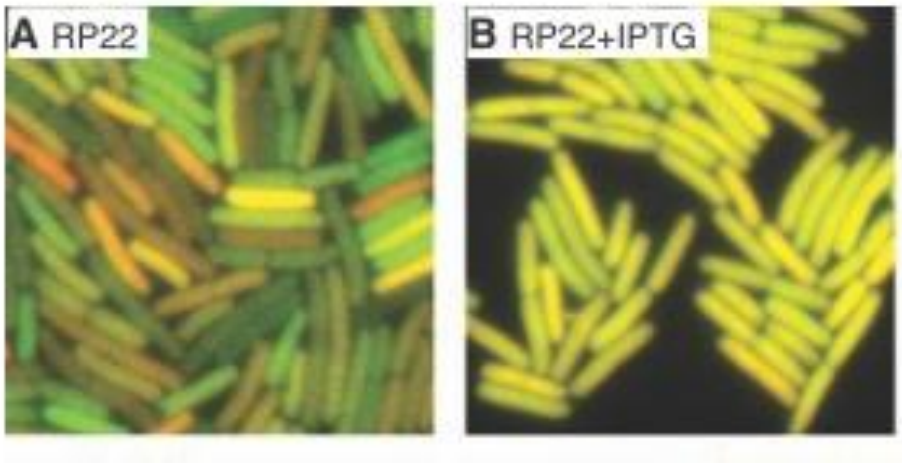

Figure 14: Noise in transcription. A & B are the fluorescent microscope images of cells expressing YFP (yellow) and CFP (green) proteins. Cells with the same amount of each protein appear yellow, whereas cells expressing more of one fluorescent protein than the other appear red or green. (**A**) In strain RP22, with promoters repressed by the wild-type *lacI* gene, red and green indicate significant amounts of intrinsic noise, (**B**) RP22 grown in the presence of lac inducer, IPTG. Both fluorescent proteins are expressed at higher levels and the cells exhibit less noise. From (Elowitz *et al*., 2002). Reprinted with permission from AAAS..

The phage λ lysis-lysogeny decision circuit of *E. coli* is frequently used as a model system to understand the coupling between noise in gene expression and phenotype selection in an organism (Arkin *et al*., 1998). *E. coli* cells infected with phage λ particles undergo either the lytic or the lysogenic cycle. This decision depends on the nutritional status of the cell, and the average phage input (API), the amount of infecting phage particles at the time of infection. These factors influence the concentration of the Cro and the cII regulatory proteins which control transcription of the genes for the lysis and lysogeny processes (See Fig. 15). Sufficient levels of Cro proteins favour lysis whereas sufficient levels of cII proteins favour lysogeny. These

pathways are mutually exclusive and this is ensured by mutual repression of these two genes.

A kinetic model of the decision circuit based on the stochastic formulation of chemical kinetics (Gillespie, 1976) was used to predict the lysis-lysogeny decision outcome. This model contains stochastic mechanisms of gene expression and a statistical-thermodynamic model of promoter regulation necessary for predicting the dynamic behaviour of the circuit. The results suggested that in genetically identical cells of a population, different developmental pathways are possible due to the randomness in the intracellular concentration of regulatory molecules, in this case, the Cro and the cII proteins. Moreover, from the association of the API/cell to phenotype selection, it was clear that randomness in the external environmental signals also influences the decision made by this regulatory circuit.

Overall, this study suggests that stochasticity in the production kinetics of transcription factors can cause phenotype switching. Further, environmental factors can influence not only the mean behaviours, but also the variability in the outcomes. Finally, noise in transcription initiation is a key determinant of the frequency of stochastic phenotype switching.

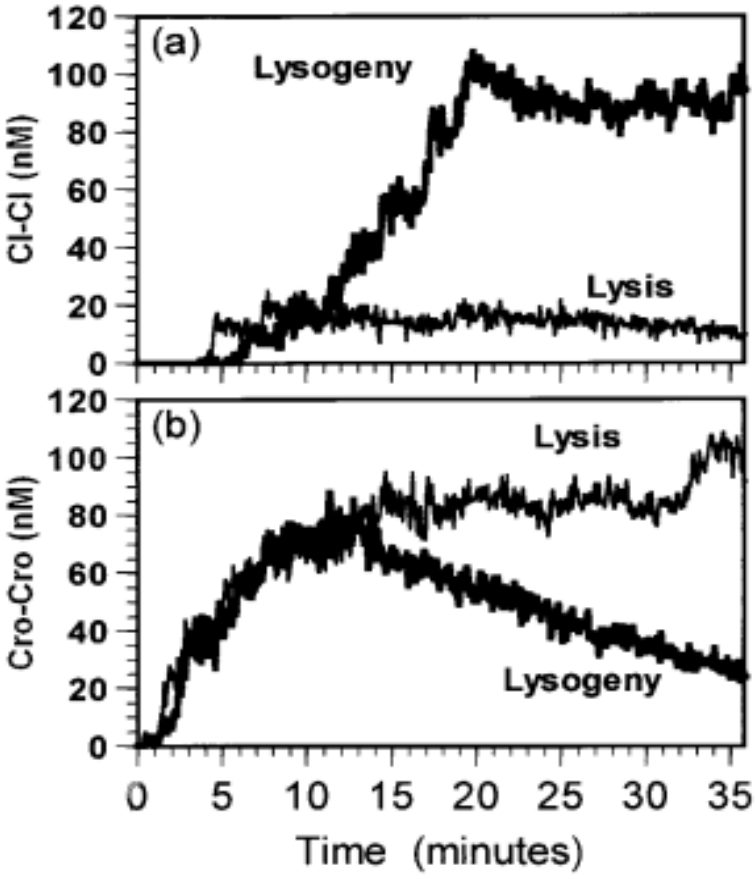

Figure: 15: Examples for simulation runs of time evolution of Cro and CI dimer concentrations. **(a)** Lysogenic case. High CII concentration after 6 min results in the accumulation of CII2 and inhibits Cro production which is needed for lytic cycle (b) Lytic case, in contrast, Cro2 accumulation increases immediately after infection and continues until it reaches the concentration to establish a feedback loop. From (Arkin *et al*., 1998). Reprinted with permission from Genetics Society of America.

Several studies on phenotype switching have demonstrated that the stochastic switching between phenotypes can be advantageous to a population of cells. This has been studied in the soil bacterium *Bacillus subtilis*, where noise in the *comK* gene's transcription determines the number of competent cells in a population (See Fig. 16) (Maamar *et al*., 2007). Competence is necessary for the uptake of DNA from the environment during the entry into stationary phase, to increase their fitness. Hence, transcription of the *comK* gene is temporally regulated such that cells have higher levels of *comK* proteins in the stationary phase.

Upon entry into stationary phase only a few randomly chosen cells of a population undergo the transition to the competent state. Measurements of the mean

and variance of mRNA levels with single-molecule FISH (Fluorescent *in situ* hybridization), showed that there is increased noise in the transcription of *comK* in these cells. When the transcription rate of *comK* is reduced, the transition to the competent state is also reduced. From these results it was concluded that noise properties are subject to evolutionary selection pressures.

The reasons why it is advantageous to have a random subset of the population in the competent state at a given time are not yet clear. However, one hypothesis is that such a stochastic switching strategy can be used to maintain part of a population in all phenotypes, so that not all cells will be at a disadvantage should the environment change (Acar *et al*., 2008). Thus, the stochastic switching of phenotype can be advantageous for microorganisms needing to cope with rapidly and unpredictably fluctuating environments.

In this thesis, to understand the plasticity in transcription kinetics, in publication I, we studied the regulation of mean and noise of the *in vivo* kinetics of transcription under the *lac/ara-1* promoter by transcription factors in *E. coli*. In publication II we studied the effect of temperature on the dynamics of transcription under the *tetA* promoter that governs tetracycline resistance in *E. coli*. Finally, we studied the transcriptional kinetics of a non-stress responsive gene controlled by *lac/ara-1* promoter. These studies shed new light on the underlying mechanisms responsible for the plasticity of transcription initiation in *E. coli*.

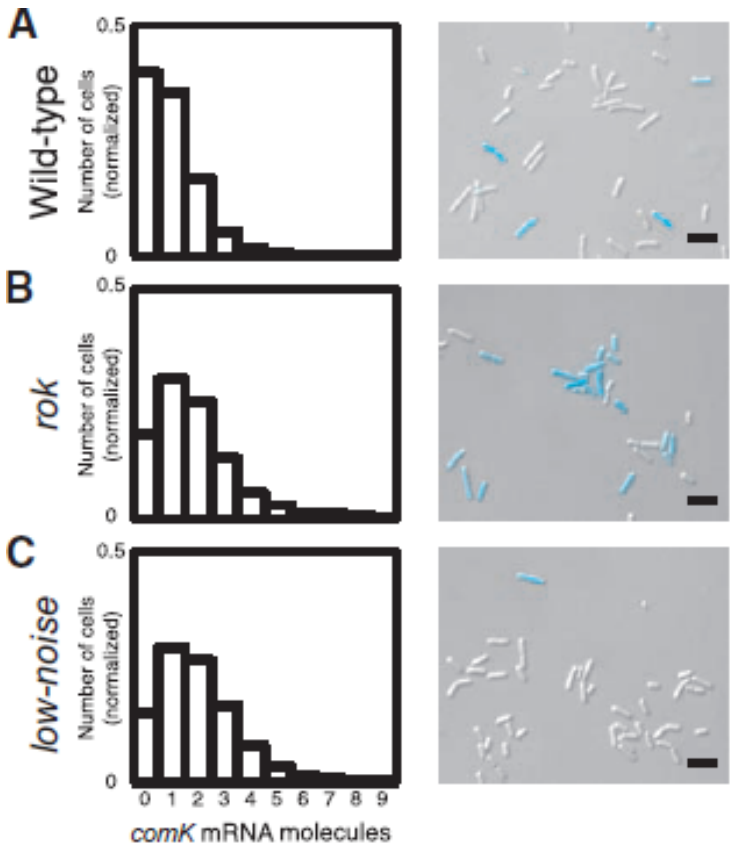

Figure 16: Noise in *comK* gene transcription and number of competent cells in a population. The histograms on the left in A - C show the comK mRNA distributions predicted by the model for the wild-type strain, *rok* strain that transcribe twice the level of *comk* mRNA than the wild-type strain and the strain with low noise in *comk* transcription. The images on the right in A – C show the number of competent cells in the same three strains from the CFP fluorescence images. From (Maamar *et al.*, 2007). Reprinted with permission from AAAS.

Chapter 3

Molecular methods for studying the dynamics of transcription

This chapter gives an overview of the methods commonly used for studying transcriptional dynamics and describes, in detail, those used in this thesis for studying the *in vivo* kinetics of transcription. These range from single-molecule approaches of fluorescent tagging, to microscopy, to methods for independent validation of the measurements from live cell imaging.

3.1 Single-molecule approaches for understanding the dynamics of transcription

The transcriptional machinery has been studied using traditional biochemical and biophysical methods. X-ray crystallography and FRET have been used to study RNAp structure and the structure of complexes formed during transcription (Mekler *et al*., 2002; Murakami *et al*., 2002). A footprinting method based on gel electrophoresis allows the detection of RNAp binding regions on the DNA and also sequence dependent interaction of TFs on DNA (Craig *et al*., 1995, deHaseth *et al*., 1998, Larson *et al*., 2011). As discussed in Chapter 2, abortive initiation assays and *in vitro* transcription assays have identified and characterized the intermediate rate-limiting steps in transcription initiation (Bertrand-Burggraf *et al*., 1984; McClure, 1985). Although these studies have significantly contributed to the understanding of transcription, they provide a mostly static picture of a dynamic process.

In particular, the kinetics of the intermediate steps in transcription initiation is the major point of regulation to tune gene expression based on changing environmental conditions (McClure, 1985; Young *et al*., 2002). Nevertheless, in prokaryotes there is usually only a single copy of a gene. Therefore, stochastic molecular fluctuations in any of the steps therefore can cause large cell-to-cell variability in mRNA levels (Elowitz *et al*., 2002). Since the mRNA levels affect the protein levels, this noise can propagate through translation to the protein numbers, thus

affecting the phenotype (Kaern *et al*., 2005). It is therefore necessary to study this dynamic process with great detail, with single-molecule sensitivity, in order to understand the underlying mechanisms responsible for the plasticity of transcription and how cells cope with or take advantage of molecular fluctuations (Zhang *et al*., 2014). Furthermore, these techniques eliminate the need for ensemble averaging as they allow the detection of rapid, transient, intermediate states in the reactions (Larson *et al*., 2012).

In vitro single-molecule techniques allow the detection of individual molecules, their assembly, conformational changes and biochemical interactions under the microscope by means of either fluorescent probes or beads tagged to molecules (See Fig. 1) (Larson *et al*., 2011; Wang & Greene, 2011). In tethered-bead techniques one end of a nucleic acid molecule is attached to a glass surface, while the other end is attached to a bead. This allows the measurement of changes in the properties of DNA from the change in the length of the tethered DNA. For example, conformational changes during lac repressor interactions were investigated using this method (Vanzi *et al*., 2006). This method yields a lower spatial resolution than that achieved with other force-based techniques, but can be used to measure the activities of hundreds of molecules in parallel (Larson *et al*., 2011).

Force spectroscopy methods such as optical or magnetic tweezing offers high resolution, but only one molecule at a time can be studied (See Fig. 1) (Larson *et al*., 2011). In these methods, both ends are commonly tagged with beads, and one of the beads is focused by a laser beam in optical trapping methods and by a magnetic field in magnetic tweezing methods, to generate a pulling or twisting force. This force is used to introduce changes in the conformation. For example, superhelical turns were introduced in DNA with *lacCONS* promoter with magnetic tweezers and the effect of supercoiling on the open complex formation was investigated. The results showed that the open complex formation is stable on negatively supercoiled DNA. In addition, the kinetics of promoter clearance and DNA scrunching during abortive initiation has been studied (Revyakin *et al*., 2006). Similarly, pausing and backtracking of RNAp during elongation have been studied with an optical tweezing method (Herbert *et al*., 2006, Neuman *et al*., 2003).

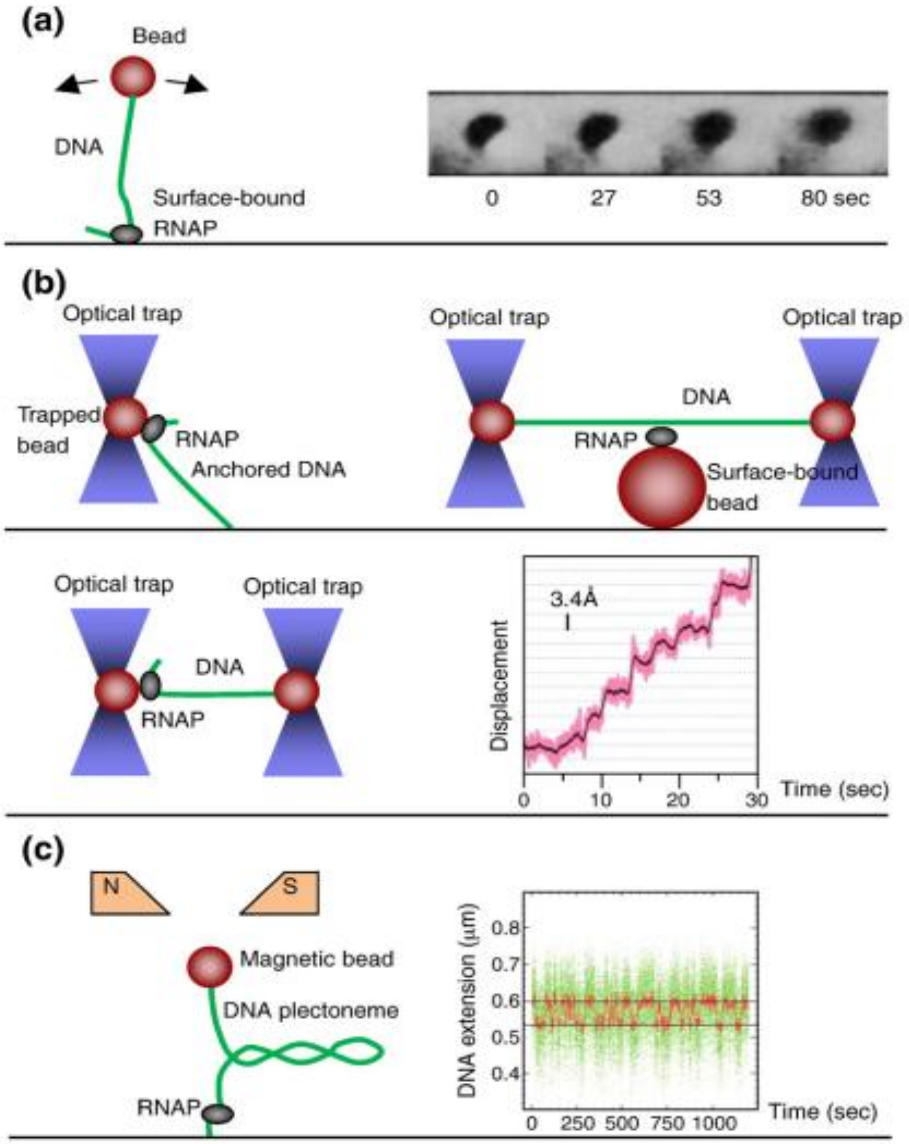

Figure 1: (a) Tethered bead technique on the left and the tracking of changes in the bead movement on the right. (b) Optical trapping method. In a one-bead optical trap, the RNAp is bound to a bead and the DNA is anchored to a surface. In a two-bead optical trap, the transcription complex is lifted from the surface by two optical traps to minimize vibration. In a three-bead optical trap, the RNAp is bound to a surface-attached bead. On the right, RNAp translocation traces using a two-bead optical trap is shown. (c) Magnetic tweezing of DNA with the tethered bead method. The plot is the DNA length with respect to the number of superhelical turns to study the effect of DNA supercoiling on promoter unwinding. Reprinted from (Wang & Greene, 2011) with permission from Elsevier.

FRET is another commonly used *in vitro* single-molecule method, based on energy transfer between two chromophores. In this method, the conformation changes are studied by labelling the domains of the structure (Young *et al.*, 2002). For

example, RNAp clamp conformation at each step in transcription initiation and elongation was investigated by monitoring the distance between a fluorescent probe, serving as a donor, incorporated at the tip of the clamp and a fluorescent probe, serving as acceptor, incorporated at the tip of the β pincer (See Fig. 2). The clamp is predominantly open in free RNAp and early intermediates in transcription initiation, but closes upon formation of a catalytically competent transcription initiation complex and remains closed during initial transcription and transcription elongation (Chakraborty *et al*., 2012).

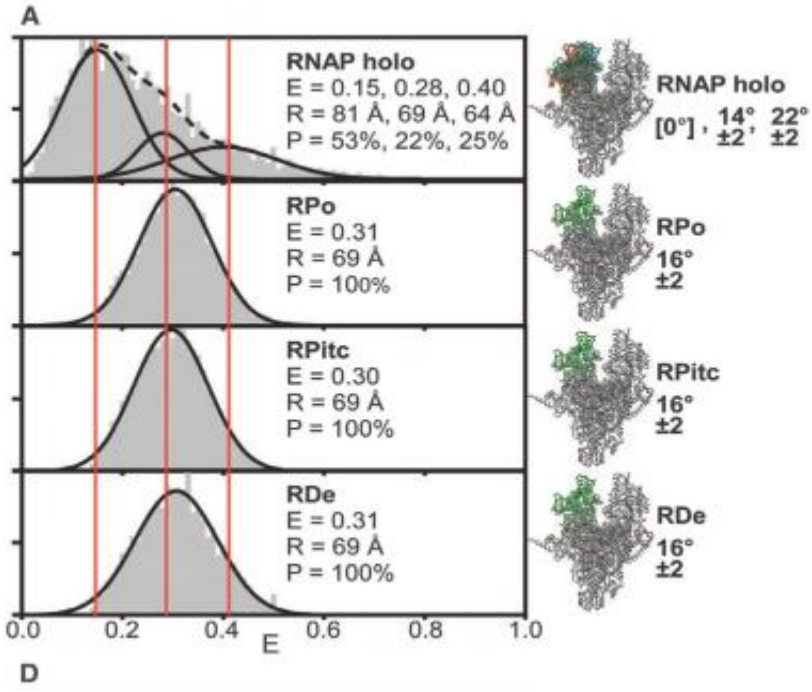

Figure 2: Determination of RNAp clamp conformation by single-molecule FRET between fluorescent probes incorporated at the tips of the RNAp β′ pincer (clamp) and the RNAp β pincer. RNAP clamp conformation in RNAP holoenzyme, RPo, RPitc (Transcription initiation complex) (4 nt of RNA), and RDe (Elongation complex) (14 nt of RNA). From (Chakraborty *et al*., 2012). Reprinted with permission from AAAS.

FISH, a cytogenetic method based on complementary nucleic acid hybridization has been recently adapted for single-molecule fluorescence probing (Raj *et al*., 2008). The advantage of this method is simultaneous detection of RNA transcripts from several genes by using probes with complementary sequences. Using this method, mRNA levels in *E. coli* can be measured under different growth

conditions. Measuring mRNA levels in a population of cells allow studies on cell to-cell variability, transcription burst frequency and burst size (See Fig. 3) (Skinner *et al*., 2013).

Though FISH provides single-molecule sensitivity in individual cells, it lacks spatial and temporal resolution. In particular, this method is performed by fixing the cells, and the probes hybridize several orders of magnitude slower than diffusion under physiological conditions. Furthermore, it is not possible to monitor individual transcription events, which is essential to understand the plasticity in the transcription process.

Cell-to cell diversity measurements using this method showed high diversity, consistent with a super-Poissonian process of RNA production. It was hypothesized that the cause is the existence of periods of activity and inactivity of the promoter (Skinner *et al*., 2013). It is noted that the variability exhibited by the cells in these measurements is not affected solely by the stochasticity in transcription, as it is also affected by RNA degradation kinetics.

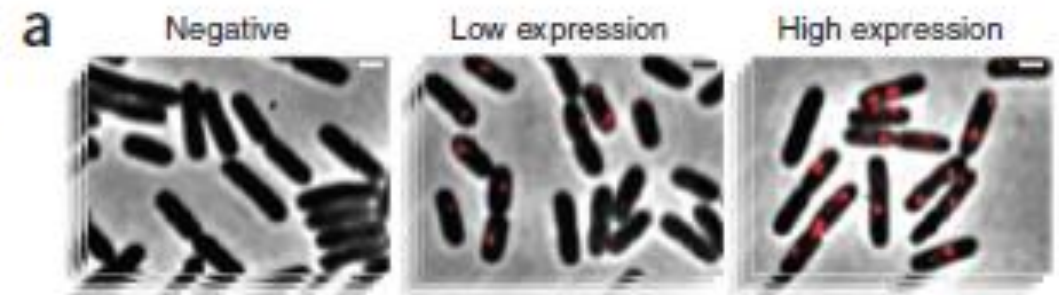

Figure 3: (**a**) The left-most panel shows a Phase-contrast image of *E. coli* cells with no fluorescence detected. The middle and right-most panels show cells detected by FISH probes with low and high expression levels, respectively. From (Skinner *et al*., 2013). Reused with permission from Nature Publishing Group.

Although the above methods can provide insights into the various steps in transcription, real-time *in vivo* single-molecule studies are required to understand the mechanisms of transcription (Xie *et al*., 2008). These methods allow the dissection of the steps in transcription, to create a detailed picture of every step in the process. A living cell is a dynamic system with a large number of coupled reactions resulting in a large, complex network of interactions. *In vivo* single-molecule studies enable the

understanding of the dynamic response of gene transcription in a physiological context and the mechanisms of tight regulation of the steps in this process (Xie *et al*., 2008; Friedman and Gelles, 2012). Further, they provide additional insights into the heterogeneity in the transcriptional activities among a population of genetically identical cells, and into the significance of this heterogeneity for enhancing fitness for adaptation (Larson *et al*., 2012).

The integration of fluorescent probing with time-lapse microscopy for single-molecule live cell imaging allows the study of the *in vivo* kinetics of transcription in individual cells (Kandhavelu *et al*., 2011). In this thesis, this strategy is used to study the *in vivo* kinetics of transcription initiation. In the following sections, this strategy is discussed in detail.

3.2 Fluorescent proteins in live cell imaging

The use of fluorescent proteins became indispensable for *in vivo* single-molecule measurements as they provide non-invasive, highly specific and sensitive methods for real time monitoring (Xie *et al*., 2008). In other standard methods for observing molecular processes, there is a trade-off between specificity and spatiotemporal resolution. For example, a Northern blot can specifically estimate mean RNA numbers, but it cannot provide spatiotemporal or 'single-cell' information. On the other hand, an electron microscopy method provides spatial information with high resolution but lacks temporal information due to fixation of the cells (Muzzey and Van Oudenaarden, 2009). Other labelling techniques such as organic dyes and nanoparticles remain challenging for specific labelling in live cells. The use of genetically encoded fluorescent proteins combined with microscopy provides an effective tool to overcome these limitations and makes possible *in vivo* investigation of spatial and temporal dynamics at the single- molecule level in individual cells (Xie *et al*., 2008).

Fluorescent proteins are non-toxic to cells in the concentration limit for detection. They do not require any cofactors except oxygen, which is necessary for the chromophore generation (Giepmans, 2006). Nevertheless, they have several limiting factors, such as brightness, photostability and lack of ability to overcome cellular autofluorescence. Therefore, it is necessary to choose the appropriate fluorescent proteins for specific studies and to optimize the imaging parameters.

Cellular autofluorescence at short wavelength excitation in bacteria is another challenge to image with single-molecule sensitivity. For example, Green Fluorescent Protein (GFP) with short wavelength is not appropriate for single fluorophore detection due to the autofluorescence. Nevertheless, it can be used to detect a tandem of repeats, above the fluorescent background of freely diffusing fluorescent proteins (Xie *et al*., 2008). For example, a gene with 96 tandem repeats of an RNA hairpin binding sequence was constructed to allow specific binding of many MS2-GFP fusion proteins that allows the detection of individual mRNA molecules in bacteria (Golding *et al*., 2005). In publications I, II and study III, we used this method (Golding *et al*., 2005) for studying the *in vivo* dynamics of transcription. In Publication IV, we present a tool that we developed for automatic tracking of cell lineages and extraction of the spatiotemporal distribution of such fluorescently tagged molecules in individual cells.

3.3 MS2-GFP tagging method

Currently, one of the most sensitive real-time *in vivo* single-molecule methods to study the *in vivo* kinetics of transcription is the MS2-GFP tagging method, originally developed for eukaryotic mRNA visualization (Bertrand *et al*., 1998). With some modifications, this method was adapted for imaging single RNA molecules in *E. coli* (Golding and Cox, 2004, Golding *et al*., 2005). It allows *in vivo* detection of individual transcription events to be detectable shortly after production. The method is based on two genetic constructs: a fluorescent protein fused to the RNA bacteriophage MS2 coat protein and a reporter RNA containing tandemly repeated MS2 binding sites (See Fig. 4).

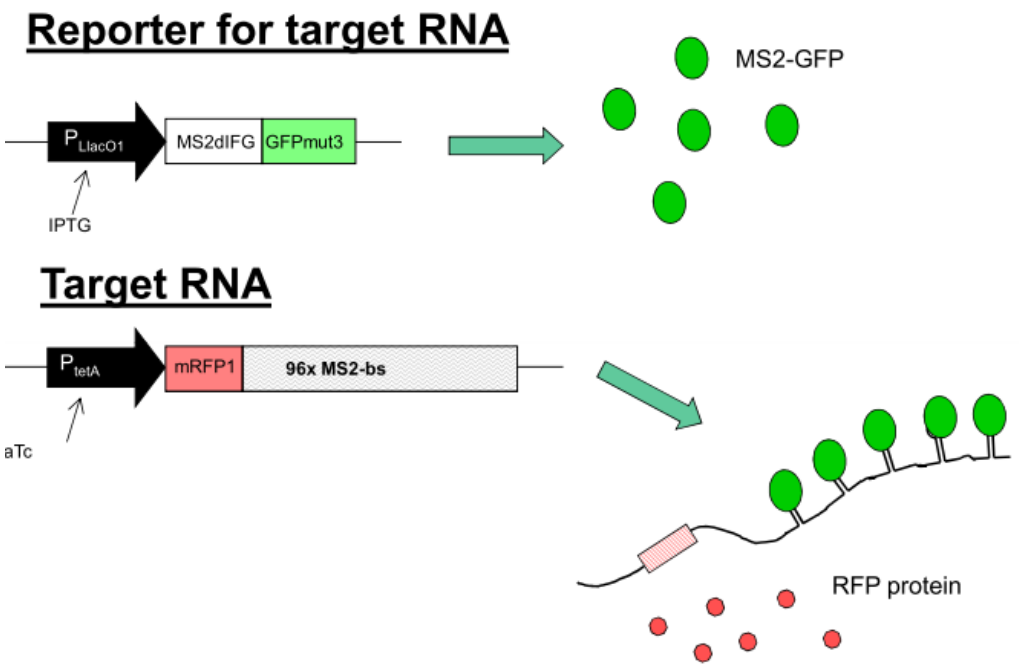

Figure 4: A schematic description of the constructs used for MS2-GFP tagging of RNA molecules in publication II of this thesis. The constructs are similar to the original constructs created by Golding and co-workers (Golding *et al*., 2005). Only the promoters are modified. The *MS2d* dimer and GFP variant, *GFPmut3* fusion gene (MS2-GFP) is under the control of the *LlacO1* promoter. In the presence of IPTG, MS2-GFP proteins are expressed. The gene for target RNA, namely, the mRFP1 coding region followed by 96 binding sites for MS2-GFP is under the control of the *tetA* promoter. Induction with anhydrotetracycline (aTc) results in RNA transcription. The regions for binding sites are immediately tagged with MS2-GFP proteins previously accumulated in the cytoplasm. The mRFP1 region is translated into proteins with red fluorescence (RFP).

The MS2 coat protein interacts with a stem–loop structure in viral RNA to repress translation of the replicase and to encapsulate the viral genome. This property was exploited and the MS2–GFP fusion proteins were constructed to tag a reporter RNA containing tandemly repeated MS2 binding sites. The *MS2d* dimer was fused to the N terminus of *GFPmut3* (GFP variant) and placed under the control of the tetracycline promoter $P_{LtetO-1}$ in a medium copy number plasmid. After inducing with aTc, cells appear bright green with thousands of copies of MS2-GFP proteins in the cytoplasm. These proteins are functional *in vivo* and are resistant to perturbations. The high copy number of MS2d-GFP proteins is necessary for tagging the tandem repeats of each RNA with 50 -100 molecules (in case of 96 binding sites) of MS2-GFP, to

form a bright spot that can be separated from the background fluorescence (Golding and Cox, 2004; Golding *et al*., 2005).

The gene for the target RNA was placed in a single copy F-plasmid based vector under the control of the *lac/ara-1* promoter. The gene contains a coding region for mRFP1, a monomeric red fluorescent protein (Campbell *et al*., 2002). This region is followed by 96 binding sites for MS2d-GFP proteins. To increase the stability of the construct, random sequences were inserted between the binding sites. Transcription of target RNA under the control of *lac/ara-1* promoter is induced by adding IPTG and arabinose, after accumulating sufficient quantity of MS2d-GFP proteins. The individual RNA molecules tagged by MS2d-GFP appear as bright spots soon after transcription has occurred. This method is generally refered as the MS2-GFP tagging method (Golding *et al*., 2005). Samples of cells, after induction under specific culture conditions, are imaged with fluorescence microscopy.

3.4 Time-lapse microscopy

The RNA molecules can be detected with the green channel of a fluorescence microscope with single-molecule sensitivity. The corresponding mRFP1 proteins can be visualized with the red channel, but as they diffuse in the cytoplasm, only total fluorescence intensity can be measured (see Fig. 5). Further, the weak signal from mRFP1 is a limitation for time-lapse observation of these proteins. The RNA spots move freely in the cytoplasm and most fluorescent spots first appear near the centre or the quarter points of the cell, where F-plasmids are localized. The RNA molecules bound to MS2-GFP are immortalized, due to reduced or complete prevention of degradation. Thus, this allows a quantitative study of transcription initiation kinetics without contamination by RNA degradation (Golding *et al*., 2005).

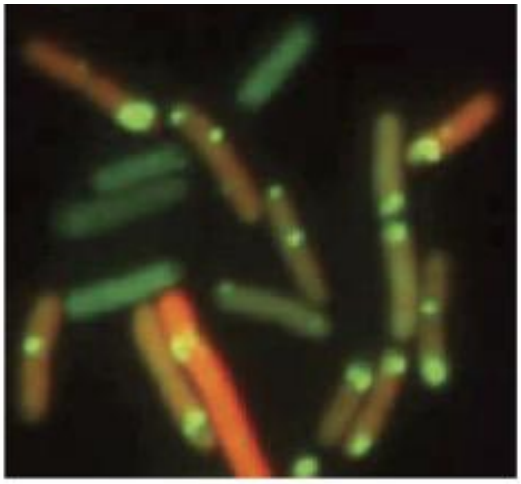

Figure 5: Detection of mRNA and protein in living cells. The picture is a false-coloured overlay of the green and red channels. The green cytoplasmic background is from the diffusion of MS2d-GFP proteins. The red background fluorescence is from mRFP1 proteins diffusing in the cytoplasm. The bright spots are tagged RNA molecules. From (Golding *et al*., 2005), Copyright (2005). Reprinted with permission from Elsevier.

For time-lapse microscopy, the cells are generally placed between an agarose gel pad on the surface of a microscopy slide, and a coverslip. The agarose gel pad contains the nutrients to support growth of the cells and inducers for gene expression (Golding *et al*., 2005; Kandhavelu *et al*., 2011). This setup facilitates long term observation of cells. In the case of temperature-dependent studies, a temperature-regulated chamber ensures a specific temperature during measurement.

Additionally, using a peristaltic pump allows measurements of transcription production from the moment of induction, since this is then possible to perform under the microscope. This has an additional advantage in that the medium can be refreshed, allowing steady-state growth for several hours under the microscope (Choi *et al*., 2008). Using this approach we studied, in study III, the changes in the kinetics of transcription activation and subsequent RNA production of a non-stress related gene, in *E. coli* cells under oxidative stress and acidic shift. In all studies included in this thesis, the images were taken with a fluorescent confocal microscope with a specific time interval. Computational tools were then used for image processing and data extraction. These tools and methods are discussed in Chapter 4.

3.5 Methods for independent validation of live cell imaging measurements

Since these newer *in vivo* single-molecule techniques are more informative than any previous method, at present there is no independent method of validation for all the information extracted. There are, however, methods that can be used to partially validate the information. For example, quantitative polymerase chain reaction (qPCR) measures mean numbers of RNA in a population. Errors in RNA counting from our method should lead to either an overestimation or an underestimation of RNA numbers in all cells, but not both, so in that sense, if the mean numbers match with the qPCR results, we can be confident on the numbers at the single-cell level, to some extent.

qPCR is a standard method for measuring RNA levels at population levels (Bustin and Nolan, 2004; Schmittgen and Livak, 2008). In this method, the total RNA from the cells is extracted and converted to cDNA with reverse-transcriptase enzyme. Using specific primers for the core region of the target, product length of 100 to150 bp are generally amplified, using a protocol similar to a general PCR reaction cycle with some modifications.

In addition to amplification, the amplified double stranded DNA is detected in real-time by SYBR Green, a fluorescent dye which selectively binds to double stranded DNA. The cycle values (Cq) are determined by finding the number of PCR cycles until the fluorescence level crosses a baseline threshold. From the Cq values, the mRNA fold changes in expression levels can be quantified by several methods (See Fig. 6) (Livak *et al*., 2001; Pfaffl *et al*., 2001).

In this thesis, in Publication II, the relative changes in the levels of target mRNA under the control of *tetA* promoter, with varying induction levels and temperature was quantified using relative quantification method (See Fig. 6). In this method, genes which have homogenous expression levels in the experimental conditions are used as internal control for the relative quantification of the target mRNA levels. To avoid any misinterpretation, it is necessary to calibrate the results with multiple reference genes. Also, in this method of quantification, PCR efficiencies should be similar for the reference and the target genes (Livak *et al*., 2001).

In study III, absolute quantification with external calibration method was employed to study the effect of stress conditions on the target mRNA expression

levels. For this, a standard curve is generated with target DNA for which the copy numbers are known. This standard curve is then used for calibration of copy number in unknown samples. This method needs that the external standard and the target sample are amplified in the same kinetic run to avoid misinterpretation due to PCR efficiencies. Further, the external standard is different from the sample cDNA , which is generally synthesized from total RNA. Therefore, it is necessary to provide similar background to the external standards (Pfaffl *et al*., 2001; Pfaffl, 2004).

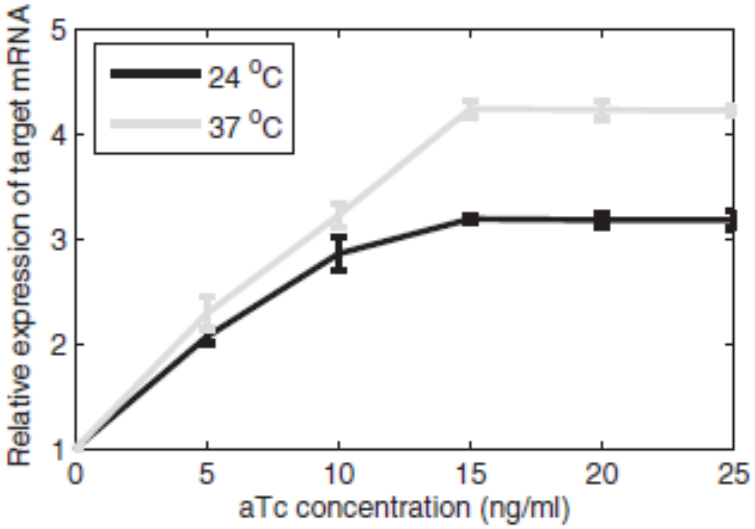

Figure 6: An example of qPCR measurements of RNA levels. The relative expression level of the target mRNA induced with different concentrations of aTc (ng/ml) at 24 °C and 37 °C is shown, quantified by qPCR using the *16 S rRNA* gene as reference. The standard deviation bars are from three independent experiments (Muthukrishnan *et al*., 2012).

Chapter 4

Computational tools for the quantification of *in vivo* dynamics of transcription in individual cells from empirical data by time-lapse microscopy

This chapter is an overview of the computational tools that were used in the studies included in this thesis. These include cells and spot segmentation of microscopy images, quantification of RNA by intensity-jump detection method, models and statistics for *in vivo* kinetics of transcription initiation.

4.1 Cells and spots segmentation

In a single time-lapse microscopy experiment, hundreds of images are taken. Each image contains several hundreds of cells. Further, the images are affected by factors such as focus drift. Also the nature of the data to be extracted from the images for studying *in vivo* transcription kinetics is complex. Therefore, it is very difficult and time consuming to do a manual analysis of images. Consequently, computational tools are used for image analysis and data extraction (Meijering, 2012).

The first step in image processing is segmentation of cells. In Publication I, II and study III, the cells are segmented using a semi-automatic method (Wang *et al*., 2010). For this, the cells in the images are manually masked and principal component analysis is used to find the dimensions and the orientations of the cells. After segmenting the cells, the segmentation of RNA spots is performed by a kernel density estimation method with a Gaussian kernel (Chowdhury *et al*., 2013; Ruusuvuori *et al*., 2010). An example of the results of the cells and spots segmentation is shown in Fig. 1.

To fully automate the above steps, CellAging tool was developed as a part of this thesis and this work is included as publication IV in this thesis. This tool performs cell segmentation of *E. coli* cells in bright field images automatically and aligns the fluorescent images with this segmentation result. Further, it establishes a temporal relationship between the cells. This allows the extraction of information on the spatial and temporal distribution of fluorescence intensity in cell lineages. Hence this tool can

be used in *in vivo* studies such as phenotypic adaptation and aging mechanisms that involves observation of several generations of *E. coli*.

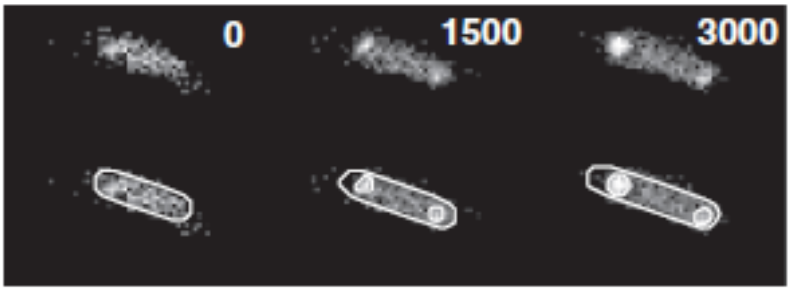

Figure 1: MS2-GFP tagged target RNAs in *E. coli* cells. Unprocessed frames and segmented cells and RNA spots. The moments when images were taken are shown for each frame (Muthukrishnan *et al.*, 2012).

4.2 Quantification of *in vivo* kinetics of transcription and cell-to-cell diversity

To detect the individual target RNA spots, the total spot intensity in each cell, at each moment, is obtained by summing the background-corrected intensities of all spots. This intensity should follow a monotonically increasing piecewise-constant function, since the tagged RNAs are immortalized (Golding *et al*., 2005). Jumps in the intensity correspond to the appearance of new target RNAs in the cell (Kandhavelu *et al*., 2012). An example of scaled spot intensity levels from individual cells using jump detection method is shown in Fig. 2.

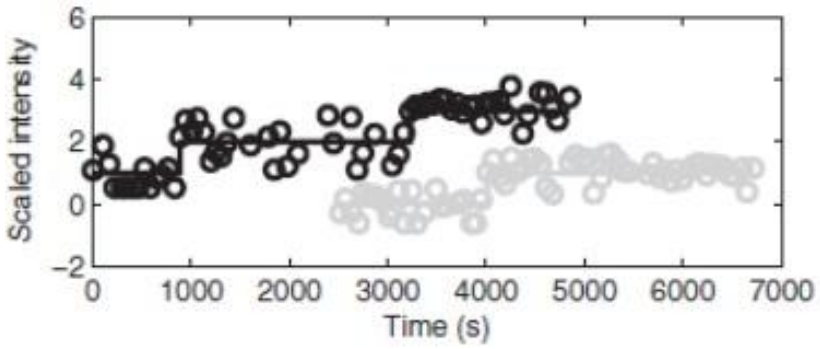

Figure 2: Examples of time series of scaled spot intensity levels from individual cells (circles) and the corresponding estimated RNA numbers (solid lines) (Muthukrishnan *et al*., 2012).

Quantification of single RNA in individual cells, the moment it appears, allows studying the *in vivo* kinetics of transcription. First, the time taken for transcription activation can be measured by extracting the time, t_0, for the first target RNA to appear in each of the cells present at the start of the imaging (See Fig. 3). This activation time includes both the time for a cell to intake inducers from the medium and the time to produce the first target RNA, once the promoter is activated by the inducers (Makela *et al*., 2013). In study III, this measurement is used to study the effects of stress on the transcription activation time of a non-stress related probe gene.

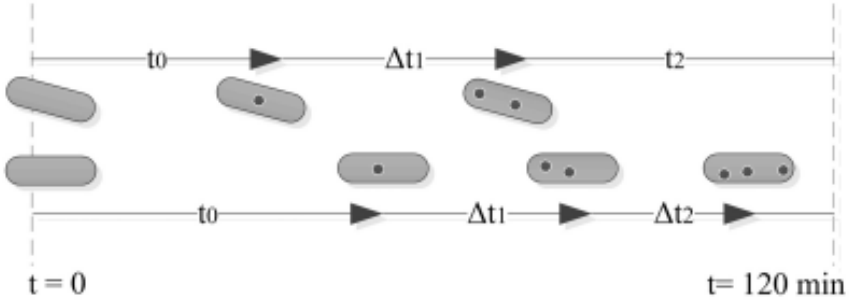

Figure 3: Schematic description of the waiting (activation) time for the first RNA production (t_0) and intervals between subsequent productions (Δt), in *E. coli* cells. Images are taken once per minute for 2 h. From (Makela *et al*., 2013). Reprinted with permission from Oxford University Press.

In addition, the distributions of time intervals between the consecutive transcription events in individual cells can be obtained. From these distributions of time intervals, the number and duration of the intermediate rate-limiting steps in transcription initiation can be inferred (Mannerstrom *et al*., 2011). For this, a sequential d-step model with the duration of each step following an independent exponential distribution is assumed, with one, two and three steps. The parameters are estimated using maximum-likelihood estimation. A likelihood-ratio test between different models can be used to reject a lower-degree model in favour of a higher-degree one. Kolmogrov-Smirnov (KS) test can be used to test if the preferred model explains the data (Kandhavelu *et al*., 2011).

. The inference method assumes that the measured kinetics of intervals between RNA productions is not affected by elongation. This is based on certain facts. First, the in this method, the tagged RNA molecules are detected soon after transcription

starts and once tagged, they are immortalized (Golding *et al*., 2005). Also, the mean interval duration of transcription initiation from this method and *in vitro* studies is in the order of 500 s or higher (Bertrand-Burggraf *et al*., 1984; Kandhavelu *et al*., 2011; McClure, 1980). In contrast, elongation is a rapid step in transcription, which takes only tens of seconds (Grieve *et al*., 2005, Herbert *et al*., 2006). Therefore, it is plausible to assume that the distributions are shaped by transcription initiation, which includes steps such as the closed complex formation, isomerization and open complex formation.

The kinetics of rate-limiting steps in transcription initiation controlled by *lac/ara-1* promoter was studied. This promoter is regulated by lac repressor and AraC activator molecules, thus inducible by IPTG and arabinose respectively (Lutz *et al*., 2001) (See Fig. 4). The mean interval duration is 2233 s for weak induction case and 1433 s for the medium induction case (See Fig. 5). Thus, the kinetics of the mean interval duration between the consecutive transcription events is affected by the inducer concentration (Kandhavelu *et al*., 2011).

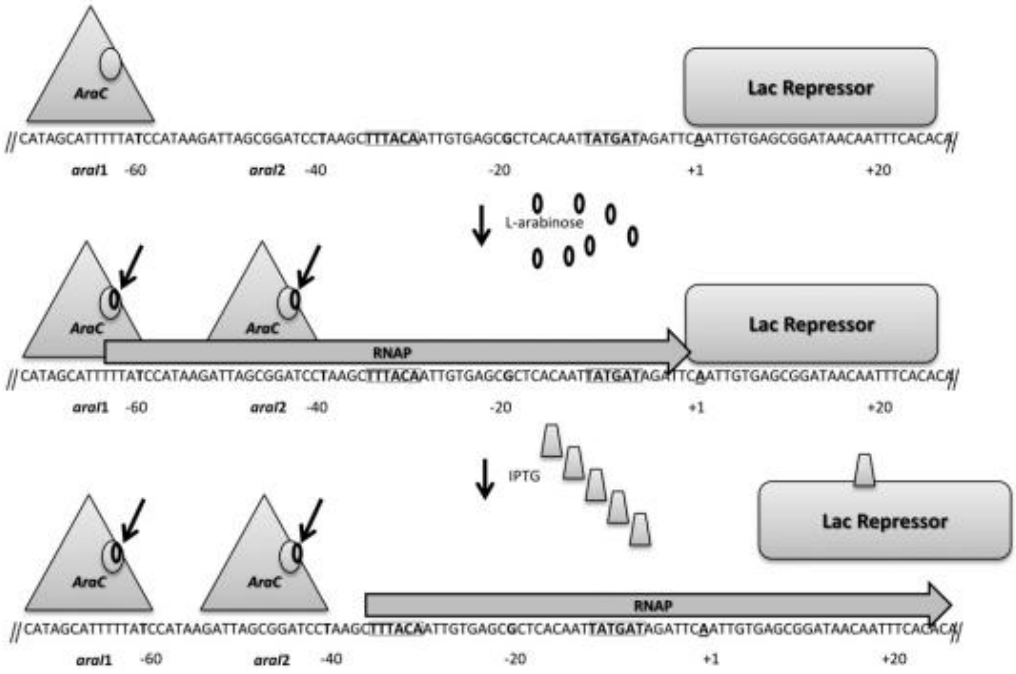

Figure 4: Schematic representation of the induction of the *lac/ara-1* promoter by arabinose and IPTG. In the presence of arabinose, AraC binds to *araI1* and *araI2* regions on the DNA and recruits RNAp to the promoter region (conserved hexamers are boxed). But RNAp cannot continue transcription since the lac repressor blocks the region spanning the TSS (+1). In the presence of IPTG, the lac repressor is released from this region and transcription by RNAp occurs. From (Kandhavelu *et al*., 2012).

There are two identified rate-limiting steps in the transcription initiation controlled by $P_{lac/ara-1}$, in both weak and medium induction cases (Kandhavelu *et al*., 2011). The kinetics of both the steps are affected by the inducer concentration and these steps are assumed as closed and open complex formation, since the lac repressor and AraC activator regulates these steps in transcription initiation (Browning and Busby, 2004). Further, the durations of the two steps differ from the *in vitro* measurements under similar conditions, but have the same order of magnitude (Buc & McClure, 1985; Lutz *et al*., 2001). As discussed earlier, one can expect this difference owing to the complexity of a living cell. One limitation of this method is that it is not able to detect the temporal order of the steps in transcription initiation.

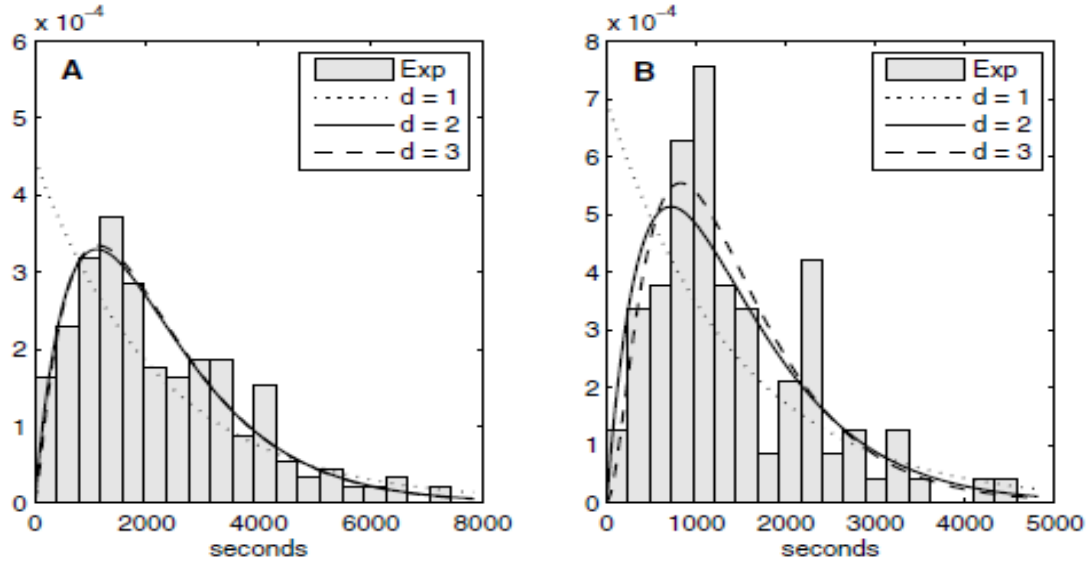

Figure 5: Histogram of the measured intervals superimposed with the probability density functions of the models. Distributions of intervals between consecutive transcription events for the weak (left) and medium (right) inductions. Each bar is 180 s. Measurement time is 2 hours (measured every 60 s). (A) mean of measured intervals is 2233 s and standard deviation is 1506 s (data from 233 intervals extracted from 283 cells). (B) mean of measured intervals is 1433 s and standard deviation is 1243 s (data from 99 intervals extracted from 40 cells). The histograms of measured intervals are superimposed with probability density functions of models with 1, 2 and 3 steps that best fit the data. Dotted line: 1-step model, solid line: 2-step model and dashed line: 3-step model (partially covered by solid line). From (Kandhavelu *et al*., 2011).

The measured intervals not only change in the mean of their interval durations, but also in their standard deviations (See Fig. 3 legend). The measurements of the mean and standard deviation are significant as they allow measuring the transcriptional noise and cell-cell variability. One can understand the variability of the transcription process by observing the shape of the distributions. For example, in Fig. 3, the shapes of the distributions are not exponential-like, indicating that the process is not a Poisson process. To quantify the noise and consequent cell-to-cell variability, statistical measures can be used.

Noise in gene expression can be measured by the square of the coefficient of variance (CV^2) (Tao *et al*., 2007). CV is defined as the standard deviation divided by the mean. For a Poisson process, CV^2 equals to one. In Figure 3, the standard deviations of the distributions are smaller than their means resulting in CV^2 less than one. Thus, RNA production by *lac/ara-1* promoter has less variability than a Poisson process with an equal rate (here called sub-Poissonian) (Kandhavelu *et al*., 2012, Muthukrishnan *et al*., 2012). Interestingly, studies of models of gene expression show that two exponentially distributed rate-limiting steps in initiation will lead to smaller fluctuations in RNA numbers than when there is only one rate-limiting step (Ribeiro *et al*., 2010). In publication II and study III of this thesis, the CV^2 has been used to describe the multi-step nature of transcription initiation.

Transcriptional noise levels can determine the extent of phenotypic diversity (Kaern *et al*., 2005). Hence, it is important to quantify noise in transcription to understand the adaptability of cells in changing environments. A sensitive statistical measure for studying the cell-to-cell variability in molecule numbers is the Fano factor (Thattai and Van Oudenaarden, 2001). It is defined as the variance divided by the mean. For a Poisson process, where the variance equals the mean, Fano factor equals to one. In publication II and study III, Fano factor has been used to measure the cell-to-cell diversity in RNA numbers.

Overall, using the computational analysis of *in vivo* measurements of the time intervals between transcription events, one can study the changes in transcription activation time, kinetics of RNA production, the kinetics of underlying rate-limiting steps of transcription initiation, noise in RNA production and consequent cell-to-cell diversity. In this thesis, in publication I, II and study III, the above computational methods are used to study the plasticity in the kinetics of the rate-limiting steps in transcription initiation.

Chapter 5

Conclusions and Discussion

The plasticity of transcription in *Escherichia coli* has been studied in this thesis using a combination of several methods. The main method of this study was the analysis of the *in vivo* kinetics of RNA production and the extraction of the underlying rate-limiting steps responsible for the observed kinetics of RNA production. From these, this study additionally focused on the regulation of noise in transcription and on the cell-to-cell diversity in RNA numbers. An *in vivo* single-molecule detection method was used to detect the productions of individual transcript as these occur in live, individual cells (Golding *et al*., 2005). Other experimental methods were used for validating the results from this method, such as qPCR and plate reading. Advanced signal processing methods were employed for consistent data analysis. Models of transcription initiation, including a sequential model of initiation, were confronted with the experimental data, for understanding of the nature and number of underlying events that affect the *in vivo* kinetics of RNA production. Finally, a software tool was developed that includes the methods developed and facilitated the extraction of the information from the data.

In Publication I, a study was presented of the individual effects of two regulatory molecules, IPTG and L-arabinose, on the *in vivo* kinetics of transcription initiation under the control of *lac/ara-1* promoter (Lutz *et al*., 2001). For this, the distributions of time intervals in the presence of each of these molecules and in the presence of both molecules in different concentrations were obtained. The mean interval duration and the degree of transcriptional noise in RNA production was found to differ between conditions, in a graded fashion. Interestingly, in all conditions, the squared coefficient of variation, CV^2, is smaller than one, suggesting that RNA production is a Sub-Poissonian process.

Previous studies have suggested that RNA production is a Super-Poissonian process, based on measurements of cell-to-cell diversity in RNA numbers (Golding *et al*., 2005; So *et al*., 2011; Taniguchi *et al*., 2010). However, unlike in our

measurements, the results reporting high values of CV^2 of RNA numbers were affected by events other than RNA production, such as RNA degradation (Yarchuk *et al*., 1992) and partitioning of tagged RNA molecules in cell division (Huh and Paulsson, 2011; Lloyd-Price *et al*., 2011).

From the distributions of time intervals, the number and the kinetics of the rate-limiting steps in transcription initiation was inferred for different concentrations of the inducers, IPTG and Arabinose. In all conditions, there are at least two rate-limiting steps in the process of transcription initiation of the *lac/ara-1* promoter. Also, the duration of these steps is independently regulated by the concentration of the two regulatory molecules. IPTG significantly affects solely the longest step, whereas arabinose affects at least the shorter step. The independent regulation of the duration of these steps suggests that the stochasticity in RNA production from this promoter is affected by environmental factors, namely the extracellular concentrations of the inducers.

The independent regulation by its two inducers provides plasticity to the production kinetics of the *lac/ara-1* promoter. Interestingly, in *E. coli*, most promoters are regulated by more than one transcription initiation factor (Browning and Busby, 2004). We suggest that one possible selective advantage of this is the likely consequent increase in plasticity, which may be needed to tune the noise of transcript production to rapidly adapt to fluctuating environments.

At the moment, it is not possible to identify the order of steps using our methods. However, based on previous *in vitro* studies (Buc & McClure, 1985; McClure, 1980; Lutz *et al*., 2001), we hypothesize that the longest step found corresponds to the open complex formation while the shorter step corresponds to the closed complex formation.

In Publication II, the effects of induction and temperature on the *in vivo* kinetics of RNA production under the control of the *tetA* promoter were studied. The methods were similar to those used in the previous publication, aside some improvements in the qPCR tests, usage of plate reading, and usage of additional variables to characterize the dynamics of RNA production. First, it was found that the Fano factor of RNA numbers in individual cells in both optimal and sub-optimal conditions was smaller than one, indicating a degree of cell-to-cell diversity in RNA numbers that is expected if the process of production is sub-Poissonian. Further, the CV^2 of the distributions of time intervals between RNA productions was found to be

smaller than one in all conditions, explaining the low cell-to-cell diversity in RNA numbers, and supporting the sub-Poissonian model of RNA production. In the same study, it was also reported that induction with aTc of the *tetA* promoter decreases the mean interval duration but, surprisingly, it increases the noise in RNA production. In contrast, compared to the dynamics in optimal conditions (37 °C), it was found that lowering temperature to sub-optimal levels (24 °C) causes an increase in both the mean interval duration, as well as in the noise level. Given these results, it is possible to conclude that noise and mean production rate can be, to some extent, regulated independently.

Assuming the sequential model of transcription initiation, the number and durations of the underlying rate-limiting steps responsible for shaping the interval distributions were inferred. There are at least three rate-limiting steps in optimal temperature and full induction conditions. Nevertheless, induction was found to have an effect on the rate-limiting steps. In particular, under full induction all the three rate-limiting steps become more rapid, two with similar duration and the third step with much shorter duration. Meanwhile, temperature is also an important factor. At a low temperature, two of the steps become more rate-limiting with increased duration, while the third step does not seem to be affected. In particular, the third step, becomes barely detectable, suggesting that it maintains a short duration at sub-optimal temperatures. Given all of the above, it was concluded that the *in vivo* kinetics of the intermediate steps in transcription initiation of the *tetA* promoter is sensitive to both temperature and induction conditions, although the number of steps is not affected. This sensitivity is the cause for the observed changes in cell-to-cell diversity in tagged RNA numbers with induction and temperature.

In study III, we investigated the *in vivo* transcription dynamics in cells under sub-lethal acidic shift and oxidative stress. For this, we used the same probe gene as in publication I, namely the gene encoding for the RNA target for MS2-GFP under the control of the *lac/ara-1* promoter. It is well known that the responses of cells to stress conditions involve, among other, global metabolic and transcriptomic changes that affect most, if not all, genes' expression dynamics. Because our probe gene is composed of an artificial promoter that is induced solely by external inducers, IPTG and Arabinose, it allows assessing the effects of these stress conditions on genes that are not involved in stress-response pathways. We started by verifying that, in fact, the probe gene was not directly activated by the stress-response mechanisms. Following

this, we studied changes in the dynamics of RNA production of fully induced promoters due to the application of stress conditions. Using a peristaltic pump to supply the media components responsible for stress induction and the inducers of the target gene to cells under observation, we observed the kinetics of transcription prior and following the activation of stress conditions. In both stress conditions, we observed that the transcription activation time increases. This activation time includes both the time for a cell to intake inducers from the medium and the time to produce the first RNA (Makela *et al*., 2013). By showing that the kinetics of RNA production in the first hour of the stressed cells does not differ significantly from the kinetics in cells under optimal conditions, we concluded that the delay in transcription activation under stress conditions is likely due to changes in the kinetics of the intake of inducers, as a result of alterations in membrane permeability, a known effect of stress (Farr *et al*., 1988; Yuk and Marshall, 2004).

Furthermore, we observed that the mean transcription rate is reduced gradually, and sufficiently, for it to differ significantly between the first and the second hour following the application of stress. Also, the effect is more prominent in the case of oxidative stress, suggesting that the cells have global, stress-specific responses. In support, we observed that the transcriptional noise is unaffected by acidic shift, but decreases under oxidative stress. Further, there are at least two rate-limiting steps in acidic shift, similar to optimal conditions. Nevertheless, the duration of both steps is increased by the same relative amount, causing the mean transcription rate to decrease while noise is unaffected. Interestingly, in the oxidative stress case, a third step in transcription becomes rate-limiting. This is the likely cause for both the mean rate and the noise in RNA production to decrease. A similar observation on the number of rate-limiting steps was reported in a previous study on the dynamics of *lac-UV5*, a *lac*-based promoter, where a third step became rate-limiting step when temperature was reduced to values below 20 °C (Buc & McClure, 1985).

In Publication IV, we published a software tool that was developed during and for the execution of the other Publications. In particular, this tool is capable of automatically extracting the relevant data from time-lapse fluorescent microscopy images using corresponding bright-field images. Also, this tool allows the automatic data analysis of *in vivo* single-molecule measurements over several generations, such that gene expression dynamics can be studied, and possible temporal correlations in the cell lineages can be detected. As a side note, this tool can also be used for studies

related to aging, as it can detect aging-related features such as polar segregation and partitioning in division of aggregates in *E. coli*.

From the above studies included in this thesis on the *in vivo* kinetics of RNA production in *E. coli*, it is possible to draw the following conclusions. *In vivo*, the process of transcription initiation in *E. coli* is multi-stepped. This imposes limits on the cell-to-cell diversity in RNA numbers introduced by the process of RNA production. Specifically, RNA production must be Sub-Poissonian. The mean transcription rate is, as expected, promoter-specific. In promoters responsive to two inducers, the mean transcription rate and the noise levels in transcript production can be regulated independently by the two inducers. This regulation appears to be achieved by an independent regulation of the individual steps involved in transcription initiation, by the different regulatory molecules. Further, in general, the number and duration of the rate-limiting steps in transcription initiation are affected by changing environmental conditions, with different steps becoming rate-limiting in different environmental conditions. From all the above, it is possible to conclude that the phenotypic plasticity of *E. coli* arises, at least in part, from the plasticity of the kinetics of the rate-limiting steps in transcription initiation.

Furthermore, using the methods developed during the course of these studies, one can study the *in vivo* regulation of the kinetics of the mean transcription rate and noise levels in RNA production of any promoter, by making use of the MS2-GFP tagging method. One limitation of our methodology is that it is not possible from it to directly identify the mechanism behind each rate-limiting step. However, it is possible to gather circumstantial evidence for this, e.g. based on the gradual changes in the duration of a particular step in response to gradual changes in the numbers of a regulatory molecule, and complementing this with existing knowledge on the possible roles of that molecule.

Bibliography

Acar, M., Mettetal, J. T. and Van Oudenaarden, A. (2008) Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.*, 40, 471-475, doi: 10.1038/ng.110

Aki, T., Choy, H. E. and Adhya, S. (1996) Histone-like protein HU as a specific transcriptional regulator: co-factor role in repression of *gal* transcription by GAL repressor. *Genes Cells*, 1 (2), 179–188.

Arkin, A., Ross, J. and McAdams, H. H. (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics*, 149, 1633-1648.

Arndt, K. M., Chamberlin, M. J. (1988) Transcription termination in *Escherichia coli*: Measurement of the rate of enzyme release from rho-independent terminators. *J. Mol. Biol.*, 202, 271–285.

Babu, M. M. and Teichmann, S. A. (2003) Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res.*, 31 (4), 1234–1244.

Bai, L., Santangelo, T. J. and Wang, M. D. (2006). Single-molecule analysis of RNA polymerase transcription. *Annu. Rev. Biophys. Biomol. Struct.*, 35, 343–360.

Bai, L., Shundrovsky, A. and Wang, M. D. (2004). Sequence-dependent kinetic model for transcription elongation by RNA polymerase. *J. Mol. Biol.*, 344 (2), 335–349.

Balleza, E., López-Bojorquez, L. N., Martínez-Antonio, A., Resendis-Antonio, O., Lozada-Chávez, I., Balderas-Martínez, Y. I., Encarnación, S. and Collado-Vides, J. (2009). Regulation by transcription factors in bacteria: beyond description. *FEMS Microbiol. Rev.*, 33 (1), 133–151.

Battesti, A., Majdalani, N. and Gottesman, S. (2011) The RpoS-mediated general stress response in *Escherichia coli*. *Annu. Rev. Microbiol.*, 65, 189–213.

Becker, G. and Hengge-Aronis, R. (2001) What makes an *Escherichia coli* promoter σ^S dependent? Role of the -13/-14 nucleotide promoter positions and region 2.5 of σ^S. *Mol. Microbiol.*, 39, 1153–1165.

Bernstein, J.A., Khodursky, A.B., Lin, P.H., Lin-Chao S. and Cohen, S.N. (2002) Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. U.S.A.*, 99 (15), 9697-9702.

Bertrand-Burggraf, E., Lefèvre, J.F. and Daune, M. (1984) A new experimental approach for studying the association between RNA polymerase and the *tet* promoter of pBR322. *Nucleic Acids* Res., 12 (3), 1697-1706.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H. and Long, R. M. (1998) Localization of ASH1 mRNA particles in living yeast. *Mol. Cell, 2* (4), 437–445.

Blattner *et al*., (1997) The Complete Genome Sequence of *Escherichia coli* K-12. *Science*, 277 (5331), 1453–1462.

Brodolin, K., Zenkin, N., Mustaev, A., Mamaeva, D., & Heumann, H. (2004) The sigma70 subunit of RNA polymerase induces *lacUV5* promoter-proximal pausing of transcription. *Nat. Struct. Mol. Biol.*, 11 (6), 551–557.

Brown, J. R. and Doolittle, W. F. (1997) Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.*, 61 (4), 456–502.

Browning, D. F. and Busby, S. J. (2004) The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.*, 2(1), 57-65.

Buc, H. and 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.

Busby, S. and Kolb, A. (1996) in *Regulation of Gene Expression in E.coli*. (Lin, E. C. C. & Lynch, A. S., eds) 255–279, New York.

Busby, S. and Ebright, R. H. (1997) Transcription activation at Class II CAP-dependent promoters. *Mol. Microbiol.* 23 (5), 853–859.

Bustin, S. A. and T. Nolan (2004). Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J. Biomol. Tech.*, 15 (3): 155-166.

Campbell, R. E., Tour, O., Palmer, A. E., Steinbach, P. A., Baird, G. S., Zacharias, D. A. and Tsien, R. Y. (2002) A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci.* USA, 99 (12), 7877-7882.

Causton, H. C. *et al*., (2001) Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell*, 12 (2), 323–337.

Chakraborty, A., Wang, D., Ebright, Y.W., Korlann, Y., Kortkhonjia, E., Kim, T., Chowdhury, S., Wigneshweraraj, S., Irschik, H., Jansen, R., *et al*., (2012). Opening and closing of the bacterial RNA polymerase clamp. *Science* 337 (6094), 591–595.

Chamberlin, M. J. (1974) The selectivity of transcription. *Annu. Rev. Biochem.*, 43, 721 -775.

Chamberlin, M. J. (1976) in RNA Polymerase. (Losick, R. and Chamberlin, M., eds) 17-67 and 159-191, Cold Spring Harbor Laboratory, New York.

Chen, D. *et al*., (2003) Global transcriptional responses of fission yeast to environmental stress. *Mol. Biol. Cell*, 14, 214–229.

Cho, B.-K., Zengler, K., Qiu, Y., Park, Y. S., Knight, E. M., Barrett, C. L., and Palsson, B. (2009) The transcription unit architecture of the *Escherichia coli* genome. *Nat. Biotechnol.*, *27* (11), 1043–1049.

Choi, P. J., Cai, L., Frieda, K., & Xie, X. S. (2008). A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*, 442–446.

Chowdhury, S., Kandhavelu, M., Yli-Harja, O. and Ribeiro, A. S. (2013) Cell segmentation by multi-resolution analysis and maximum likelihood estimation (MAMLE). *BMC Bioinformatics*, 14 (Suppl 10), S8.

Ciampi, M. S. (2006). Rho-dependent terminators and transcription termination. Microbiology, 152 (9), 2515–2528.

Craig, M. L., Suh, W. C., and Record, M. T. (1995). HO. and DNase I probing of E sigma 70 RNA polymerase--*lambda PR* promoter open complexes: Mg^{2+} binding and its structural consequences at the transcription start site. *Biochemistry*, 34 (48), 15624–15632.

Crick, F. (1970) Central dogma of molecular biology. *Nature*, 227 (5258), 561–563.

Dalebroux, Z. D., Svensson, S. L., Gaynor, E. C., and Swanson, M. S. (2010) ppGpp conjures bacterial virulence. *Microbiol. Mol. Biol. Rev.*, 74 (2), 171–199.

deHaseth, P. L. and Helmann, J. D. (1995) Micro Review Open complex formation by *Escherichia coli* RNA polymerase the mechanism of polymerase-induced strand separation of double helical DNA. *Mol. Microbiol.*, 16 (5), 817-824.

deHaseth, P. L., Zupancic, M. L. and Record, M. T. (1998) RNA Polymerase-Promoter interactions: the comings and goings of RNA Polymerase. *J. Bacteriol*., 180 (12), 3019-3025.

Ebright, R. H. (1993) Transcription activation at Class I CAP dependent promoters. *Mol. Microbiol*., 8, 797–802.

Eliasson, A., Bernander, R., Dasgupta, S., and Nordström, K. (1992). Direct visualization of plasmid DNA in bacterial cells. *Mol. Microbiol.*, 6 (2), 165–170.

Elowitz, M. B., Levine, A. J., Siggia, E. D. and Swain P. S. (2002) Stochastic gene expression in a single cell. *Science*, 297 (5584), 1183-1186.

Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G.and Gardner, T. S. (2007) Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.*, 5 (1), e8.

Farr, S. B., Touati, D. and Kogoma, T. (1988) Effects of oxygen stress on membrane functions in *Escherichia coli*: role of HPI catalase. *J. Bacteriol.*, 170, 1837–1842.

Francois, J., David, P., Carmen, S. and Jacques, M. (1960) The Operon: A group of genes whose Expression is coordinated by an operator. *Science,* 250, 1727-1729.

Friedman, L. J., and Gelles, J. (2012). Mechanism of transcription initiation at an activator-dependent promoter defined by single-molecule observation. *Cell*, 148 (4), 679–689.

Garrett, R. (1999) Mechanics of the ribosome. *Nature*, 400 (6747), 811–812.

Gerstel, U., Park, C. and Romling, U. (2003) Complex regulation of *csgD* promoter activity by global regulatory proteins. *Mol. Microbiol.*, 49, 639–654.

Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, *22*, 403–434.

Giepmans, B. N. G., Adams, S. R., Ellisman, M. H., and Tsien, R. Y. (2006). The fluorescent toolbox for assessing protein location and function. *Science*, 312 (5771), 217–224.

Golding, I., Paulsson, J. J., Zawilski, S. M. and Cox, E. C. (2005) Real-time kinetics of gene activity in individual bacteria. *Cell*, 123 (6), 1025-1036.

Golding, I. and Cox, E. (2004) RNA dynamics in live Escherichia coli cells**.** *Proc. Natl. Acad. Sci*. USA, 101**,** 11310-11315.

Goldman, S. R., Ebright, R. H., and Nickels, B. E. (2009) Direct detection of abortive transcripts *in vivo*. *Science*, 324 (5929), 927–928.

Greive, S. J. and von Hippel, P. H. (2005) Thinking quantitatively about transcriptional regulation. *Nat. Rev. Mol. Cell Biol.*, 6, 221–232.

Gunasekera, T. S., Csonka, L. N. and Paliy, O. (2008) Genome-wide transcriptional responses of *Escherichia coli* K-12 to continuous osmotic and heat stresses. *J. Bacteriol.*, 190, 3712–3720.

Harley, C. B., and Reynolds, R. P. (1987) Analysis of *E. coli* promoter sequences. *Nucleic Acids Res.*, 15 (5), 2343–2361.

Hengge-Aronis, R. (1999) Interplay of global regulators and cell physiology in the general stress response of *Escherichia coli*. *Curr. Opin. Microbiol.*, 2(2), 148–152.

Hengge-Aronis, R. (2002) Recent insights into the general stress response regulatory network in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.*, 4, 341–346.

Herbert, K. M., La Porta, A., Wong, B. J., Mooney, R., Neuman, K. C., Landick, R., and Block, S. M. (2006) Sequence-resolved detection of pausing by single RNA polymerase molecules. *Cell*, 125 (6), 1083–1094.

Huh, D. and Paulsson, J. (2011) Random partitioning of molecules at cell division. *Proc. Natl. Acad. Sci.* USA, 108 (36), 15004-15009.

Jozefczuk, S., Klie, S., Catchpole, G., Szymanski, J., Cuadros-Inostroza, A., Steinhauser, D., Selbig, J. and Willmitzer, L. (2010) Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol. Syst. Biol.*, 6, 364.

Kaern, M., Elston, T.R., Blake, W.J., and Collins, J.J. (2005) Stochasticity in gene expression from theories to phenotypes. *Nat. Rev. Genet.*, 6 (6), 451-464.

Kandhavelu, M., Mannerström, H., Gupta, A., Häkkinen, A., Lloyd-Price, J., Yli-Harja, O. and 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 Syst. Biol.*, 5, 149.

Kandhavelu, M., Lloyd-Price, J., Gupta, A., Muthukrishnan, A.-B., Yli-Harja, O. and 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 Lett.*, 586, 3870–3875.

Kannan, G., Wilks, J. C., Fitzgerald, D. M., Jones, B. D., Bondurant, S. S., & Slonczewski, J. L. (2008) Rapid acid treatment of *Escherichia coli*: transcriptomic response and recovery. *BMC Microbiol.*, 8-37.

Kapanidis, A. N., Margeat, E., Laurence, T. A, Doose, S., Ho, S. O., Mukhopadhyay, J. and Weiss, S. (2005) Retention of transcription initiation factor sigma70 in transcription elongation: single-molecule analysis. *Mol. Cell*, 20 (3), 347–356.

Landini, P., Egli, T., Wolf, J. and Lacour, S. (2013) sigmaS, a major player in the response to environmental stresses in *Escherichia coli*: role, regulation and mechanisms of promoter recognition. *Environ. Microbiol. Rep.*, 6 (1), 1-13.

Larson, M. H., Landick, R., & Block, S. M. (2011). Single-molecule studies of RNA polymerase: one singular sensation, every little step it takes. *Mol. Cell*, 41 (3), 249-262.

Lee, P. S., & Lee, K. H. (2003) *Escherichia coli*--a model system that benefits from and contributes to the evolution of proteomics. *Biotechnol. Bioeng.*, 84 (7), 801–814.

Livak, K. J. and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{(-\text{Delta Delta C(T)})}$ Method. *Methods*, 25, 402–408.

López-Maury, L., Marguerat, S., & Bähler, J. (2008) Tuning gene expression to changing environments from rapid responses to evolutionary adaptation. *Nat. Rev. Genet.*, 9 (8), 583–593.

Lloyd-Price, J., Lehtivaara, M., Kandhavelu, M., Chowdhury, S., Muthukrishnan, A. B., Yli-Harja, O. and Ribeiro, A. S. (2011) Probabilistic RNA partitioning generates transient increases in the normalized variance of RNA numbers in synchronized populations of *Escherichia coli*. *Mol. Biosyst*., 8(2), 565-571.

Lutz, R., Lozinski, T., Ellinger, T., and Bujard, H. (2001) Dissecting the functional program of *Escherichia coli* promoters the combined mode of action of Lac repressor and AraC activator. *Nucleic Acids Res.*, 29 (18), 3873-3881.

Neuman, K. C., Abbondanzieri, E. A., Landick, R., Gelles, J. and Block, S. M. (2003). *Cell*, 115 (4), 437–447.

Mannerstrom, H., Yli-Harja, O. and Ribeiro, A.S. (2011) Inference of kinetic parameters of delayed stochastic models of gene expression using a markov chain approximation. *EURASIP J. Bioinform. Syst. Biol.*, 2011, Article ID: 572876.

Maamar, H., Raj, A., and Dubnau, D. (2007). Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science,* 317 (5837), 526–529.

Magnusson, L. U., Farewell, A., and Nyström, T. (2005) ppGpp: a global regulator in *Escherichia coli*. *Trends Microbiol.*, 13 (5), 236–242.

Maheshri, N., & O'Shea, E. K. (2007) Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu. Rev. Biophys. Biomol. Struct.*, 36, 413–434.

Martinez-Antonio, A. & Collado-Vides, J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol.*, 6, 482–489.

Ma, S. and Bohnert, H. J. (2007) Integration of *Arabidopsis thaliana* stress-related transcript profiles, promoter structures, and cell-specific expression. *Genome Biol*., 8, R49.

Mäkelä, J., Kandhavelu, M., Oliveira, S.M.D., Chandraseelan, J.G., Lloyd-Price, J., Peltonen, J., Yli-Harja, O. and Ribeiro, A.S. (2013) *In vivo* single-molecule kinetics of activation and subsequent activity of the arabinose promoter. *Nucleic Acids Res.*, 41, 6544–6552.

McAdams, H. and Arkin, A. (1997) Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci.* USA, 94 (3), 814-819.

McClure, W. R., Cech, C. L., and Johnston, D. E. (1978). A steady state assay for the RNA polymerase initiation reaction. *J. Biol. Chem.*, 253 (24), 8941–8948.

McClure, W.R. (1980) Rate-limiting steps in RNA chain initiation. *Proc. Natl. Acad. Sci.* USA, 77 (10), 5634-5638.

McClure, W. R. (1985) Mechanism and control of transcription initiation in prokaryotes, *Annu. Rev. Biochem.*, 54, 171-120.

McLeod, S. M., and R. C. Johnson. 2001. Control of transcription by nucleoid proteins. *Curr. Opin. Microbiol.* 4, 152–159.

Meijering, E. (2012) Cell Segmentation: 50 Years Down the Road, *IEEE Signal Processing Mag.*, 29 (5), 140–145.

Mendoza-Vargas, A., Olvera, L., Olvera, M., Grande, R., Vega-Alvarado, L., Taboada, B. and Morett, E. (2009). Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PloS One*, 4 (10), e7526.

Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A.N., Niu, W., Ebright, Y.W., Levy, R., and Ebright, R.H. (2002) Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell.*, 108, 599–614.

Miller, O. L., Hamkalo, B. A, and Thomas, C. A. (1970). Visualization of bacterial genes in action. *Science*, 169 (3943), 392–395.

Mooney, R. A., Darst, S. A, and Landick, R. (2005). Sigma and RNA polymerase: an on-again, off-again relationship? *Mol. Cell*, 20 (3), 335–345.

Mott, M. L., & Berger, J. M. (2007). DNA replication initiation: mechanisms and regulation in bacteria. *Nat. Rev. Microbiol.*, 5 (5), 343–354.

Murakami, K. S., Masuda, S., and Darst, S. A. (2002) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. *Science*, 296 (5571), 1280–1284.

Murray, J. I. *et al*., (2004) Diverse and specific gene expression responses to stresses in cultured human cells. *Mol. Biol. Cell*, 15, 2361–2374.

Muthukrishnan, A. B., Kandhavelu, M., Lloyd-Price, J., Kudasov, F., Chowdhury, S., Yli-Harja, O. and Ribeiro, A.S. (2012) Dynamics of transcription driven by the *tetA* promoter, one event at a time, in live *Escherichia coli* cells. *Nucleic Acids Res.*, 40, 8472–8483.

Muzzey, D., Van Oudenaarden, A. (2009) Quantitative time-lapse fluorescence microscopy in single cells. *Annu. Rev. Cell. Dev. Biol.* 25, 301-327.

O'Neill, M. C. (1989). *Escherichia coli* Promoters. *J. biol. Chem.*, 264 (10), 5522–5530.

Nudler, E., & Gottesman, M. E. (2002). Transcription termination and anti-termination in *E. coli*. *Genes Cells*, 7 (8), 755–768.

Pfaffl, M. W. (2001). Development and validation of an externally standardised quantitative Insulin-like growth factor-l RT-PCR using LightCycler SYBR Green I Technology. In: (Meuer, S., Wittwer, C., and Nakagawara, K., eds.), Rapid Cycle Real-time PCR, Methods and Applications Springer Press, Heidelberg, ISBN 3-540-66736-9, pp 281-191.

Pfaffl, M. W. (2004) Quantification strategies in real-time PCR. In A–Z of Quantitative PCR. (Bustin, S.A., ed.), La Jolla: International University Line, USA.

Pribnow, D. (1979) in Biological Regulation and Development (Goldberger, R.F., ed.) vol. 1, 219-269, *Plenum Press, New York.*

Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. and Tyagi, S. (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods*., 5 (10), 877-879.

Ramos, J. L., Gallegos, M. T., Marqués, S., Ramos-González, M. I., Espinosa-Urgel, M., and Segura, A. (2001) Responses of Gram-negative bacteria to certain environmental stressors. *Curr. Opin. Microbiol.*, 4 (2), 166–171.

Revyakin, A., Liu, C., Ebright, R. H., and Strick, T. R. (2006) Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science*, 314 (5802), 1139–43.

Reznikoff, W. S., Siegele, D. A, Cowing, D. W., and Gross, C. A. (1985) The regulation of transcription initiation in bacteria. *Ann. Rev. Genet.*, 19, 355–387.

Ribeiro, A. S. (2010) Stochastic and delayed stochastic models of gene expression and regulation. *Math. Biosci.*, 223 (1), 1–11.

Ribeiro, A. S., Häkkinen, A., Mannerström, H., Lloyd-Price, J., & Yli-Harja, O. (2010) Effects of the promoter open complex formation on gene expression dynamics. *Physical Review E*, 81 (1), 011912.

Ross, W., Vrentas, C. E., Sanchez-Vazquez, P., Gaal, T., & Gourse, R. L. (2013) The magic spot: a ppGpp binding site on *E. coli* RNA polymerase responsible for regulation of transcription initiation. *Mol. Cell*, 50 (3), 420–429.

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

Rosenberg, M., Court, D. (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. *Ann. Rev. Genet*., 13, 319-353

Roszak, D. B., and Colwell. R. R. (1987) Survival strategies of bacteria in the natural environment. *Microbiol. Rev.*, 51(3), 365-379.

Roussel, M. and Zhu, R. (2006) Validation of an algorithm for delay stochastic simulation of transcription and translation in gene expression. *Phys. Biol.*, 3 (4), 274-284.

Saecker, R. M., Record, M. T. and deHaseth, P. L. (2011) Mechanism of bacterial transcription initiation: RNA polymerase-promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J. Mol. Biol.*, 412 (5), 754-771.

Schleif, R. (2010) AraC protein, regulation of the L-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiol. Rev.*, 34 (5), 779–796.

Schmittgen, T. D. and Livak K. J. (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nat.Protoc.*, 3 (6), 1101-1108.

Sheridan, S. D., Benham, C. J. and Hatfield, G. W. (1998) Activation of gene expression by a novel DNA structural transmission mechanism that requires supercoiling-induced DNA duplex destabilization in an upstream activating sequence. *J. Biol. Chem.*, 273, 21298–21308.

Skinner, S. O., Sepúlveda, L. A, Xu, H., and Golding, I. (2013). Measuring mRNA copy number in individual *Escherichia coli* cells using single-molecule fluorescent *in situ* hybridization. *Nat. Protoc.*, 8 (6), 1100–1113.

Sleator, R. D. and Hill, C. (2002) Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. *FEMS Microbiol. Rev.*, 26 (1), 49–71.

So, L. H., Ghosh, A., Zong, C., Sepúlveda, L. A., Segev, R. and Golding, I. (2011) General properties of transcriptional time series in *Escherichia coli*. *Nat. Genet*., 43 (6), 554-560.

Stelling, J., Sauer, U., Szallasi, Z. and Doyle, F.J. (2004) Robustness of cellular functions. *Cell*, 118 (6), 675–685.

Stoebel, D. M., Hokamp, K., Last, M. S., & Dorman, C. J. (2009) Compensatory evolution of gene regulation in response to stress by *Escherichia coli* lacking RpoS. *PLoS Genet.*, 5 (10), e1000671.

Swain, P. S., Elowitz, M. B., and Siggia, E. D. (2002) Intrinsic and extrinsic Contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci.* USA, 99 (20), 12795-12800.

Taniguchi,Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A. and Xie, X. S. (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329 (5991), 533–538.

Tao, Y., Zheng, X. and Sun Y. (2007). Effect of feedback regulation on stochastic gene expression. *J. Theor. Biol.*, 4(21), 827-836.

Thattai, M. and van Oudenaarden, A. (2001) Intrinsic noise in gene regulayory networks. *Proc. Natl. Acad. Sci.* USA, 98 (15), 8614-8619.

Travers, A. A. and Burgess, R. R. (1969) Cyclic re-use of the RNA polymerase sigma factor. *Nature*., 222 537–540.

Uptain, S. M., Kane, C. M. and Chamberlin, M. J. (1997) Basic mechanisms of transcript elongation and its regulation. *Annu. Rev. Biochem.*, 66, 117–172.

Van den Broek, B., Vanzi, F., Normanno, D., Pavone, F. S., and Wuite, G. J. L. (2006). Real-time observation of DNA looping dynamics of Type IIE restriction enzymes NaeI and NarI. *Nucleic Acids Res*., 34 (1), 167–174.

Van Hijum, S. A.F. T., Medema, M. H., and Kuipers, O. P. (2009) Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiol. Mol. Biol. Rev.*, 73 (3), 481–509.

Von Hippel, P., Bear, D., Morgan, W. And McSwiggen, J. (1984) Protein-nucleic acid interactions in transcription. A molecular analysis. *Ann. Rev. Biochem.*, 53, 389-446.

Walter, G., Zillig, W., Palm, P., Fuchs, E. (1967) Initiation of DNA – Dependent RNA synthesis and the effect of Heparin on RNA Polymerase. *Eur. J. Biochem*., 3, 194-201

Wang, Q., Niemi, J., Tan, C.-M., You, L., & West, M. (2010) Image segmentation and dynamic lineage analysis in single-cell fluorescence microscopy. *Cytometry. Part A : The J. of the Int. Society for Anal. Cytol.*, 77 (1), 101–110.

Wang, F., & Greene, E. C. (2011) Single-molecule studies of transcription: from one RNA polymerase at a time to the gene expression profile of a cell. *J. Mol. Biol.*, 412 (5), 814–831.

Wang, F., Redding, S., Finkelstein, I. J., Gorman, J., Reichman, D. R., and Greene, E. C. (2013) The promoter-search mechanism of *Escherichia coli* RNA polymerase is dominated by three-dimensional diffusion. *Nat. Struct. Mol. Biol.*, 20 (2), 174–181.

Weber, H., Polen, T., Heuveling, J., Wendisch, V.F. and Hengge, R. (2005) Genome-wide analysis of the general stress response network in *Escherichia coli*: σ^S-dependent genes, Promoters, and Sigma Factor Selectivity. *J. Bacteriol*., 187, 1591–1603.

Xie, X. S., Choi, P. J., Li, G.-W., Lee, N. K. and Lia, G. (2008) Single-molecule approach to molecular biology in living bacterial Cells. *Annu. Rev. Biophys*., 37, 417–444.

Yarchuk, O., Jacques, N., Guillerez, J. and Dreyfus, M. (1992) Interdependence of translation, transcription and mRNA degradation in the *lacZ* gene. *J. Mol. Biol.,* 5 (226), 581–596.

Yuk, H. and Marshall, D.L. (2004) Adaptation of *Escherichia coli* O157 : H7 to pH alters membrane lipid composition , Verotoxin Secretion , and resistance to simulated gastric fluid. *Appl. Environ. Microbiol.*, 70, 3500–3505

Young, B. A., Gruber, T. M., Gross, C. A., and Francisco, S. (2002) Views of transcription initiaion. *Cell,* 109, 417–420.

Zhuravel, D., Fraser, D., St-Pierre, S., Tepliakova, L., Pang, W. L., Hasty, J., & Kærn, M. (2010). Phenotypic impact of regulatory noise in cellular stress-response pathways. *Syst. Synth. Biol.*, 4 (2), 105–116.

Zuo, Y., Wang, Y., and Steitz, T. A. (2013). The mechanism of *E. coli* RNA polymerase regulation by ppGpp is suggested by the structure of their complex. *Mol. Cell,* 50 (3), 430–436.

Publications

Publication I

I. M. Kandhavelu, J. Lloyd-Price, A. Gupta,, A-B. Muthukrishnan, O. Yli-Harja, and A.S. Ribeiro, *Regulation of mean and noise of the in vivo kinetics of transcription under the control of the lac/ara-1 promoter*, FEBS Lett. 586, 3870–3875, 2012.

Publication II

II. A-B. Muthukrishnan, M. Kandhavelu, J. Lloyd-Price, F. Kudasov, S. Chowdhury, O. Yli-Harja, and A.S. Ribeiro, *Dynamics of transcription driven by the tetA promoter, one event at a time, in live Escherichia coli cells*, Nucleic Acids Res. 40(17), 8472-8483. DOI:10.1093/nar/gks583, 2012.

Supplementary for publication II

Study III

III. A.-B. Muthukrishnan, A. Martikainen, R. Neeli Venkata, and A.S. Ribeiro, *In vivo single-molecule probing of the transcriptional dynamics of non-stress-responsive genes in stressed Escherichia coli cells*, FEBS Journal, under review.

Publication IV

IV. A. Häkkinen, A.-B. Muthukrishnan, A. Mora, J.M. Fonseca, and A.S. Ribeiro, *CellAging: A tool to study segregation and partitioning in division in cell lineages of Escherichia coli*, Bioinformatics 29 (13), 1708-1709. DOI: 10.1093/bioinformatics/btt194, 2013.

Supplementary for Publication IV

Tampereen teknillinen yliopisto
PL 527
33101 Tampere

Tampere University of Technology
P.O.B. 527
FI-33101 Tampere, Finland

ISBN 978-952-15-3344-0
ISSN 1459-2045