DELAYED STOCHASTIC MODELLING OF PROKARYOTIC TRANSCRIPTION WITH ABORTIVE INITIATION

SILKY SHARMA
Bachelor of Technology, Uttar Pradesh Technical University, India, 2013

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfillment of the
Requirements for the Degree

MASTER OF SCIENCE

Department of Chemistry and Biochemistry
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

DELAYED STOCHASTIC MODELLING OF PROKARYOTIC TRANSCRIPTION WITH ABORTIVE INITIATION

SILKY SHARMA

Date of Defense: January 20, 2017

Dr. Marc R. Roussel Supervisor	Professor	Ph.D.
Dr. Robert Laird Committee Member	Associate Professor	Ph.D.
Dr. Stacey Wetmore Committee Member	Professor	Ph.D.
Dr. Dennis Salahub External examiner	Professor	Ph.D.
Dr. Nehal Thakor Examination Committee Chair	Assistant Professor	Ph.D.

Dedication

To my parents,

Mr. and Mrs. Sharma,

and my beloved brother Sahil,

whose hard-work, love, encouragement and prayers by day and night make me able to get such success and honour.

Along with my supervisor, Dr. Marc R. Roussel,

who always picked me up on time and has been a constant source of knowledge and inspiration for me.

Abstract

A quantitative model for gene transcription should treat two main features: stochastic fluctuations due to low copy numbers, and time delays resulting from the lengthy sequence of reactions that form the transcription process. Bratsun et al. (PNAS, 102(41):14593-14598, 2005) proposed an algorithm by which we can include both of these features. I have studied a delay stochastic model for RNA transcription including abortive initiation. This study focused on the parametric dependence of the mean stationary RNA concentration and of the rise time, i.e. the characteristic time taken to achieve the stationary concentrations. The mean RNA concentration depends on initiation, elongation, polymerase arrest and RNA degradation rates including delays associated with clearance of promoter and polymerase arrest. Rise time depends on initiation, polymerase arrest, termination and RNA degradation rates. I have developed an analytic theory of rise time based on the delayed mass-action formalism to validate the simulation results.

Acknowledgments

This journey has been so special to me because of a few special people who helped me to get through this adventure. My deep gratitude goes to my supervisor Dr. Marc R. Roussel, *thank you*. He gave me this wonderful opportunity to pursue my Master's under his supervision. This thesis would not have been possible without his guidance and support.

I also wish to thank my other committee members at the University of Lethbridge, Dr. Robert Laird and Dr. Stacey Wetmore, for their substantial feedback and guidance.

Last but not the least, I would like to thank all the members of the Roussel lab for helping me every time I needed and making this period unforgettable.

Contents

Contents vi

List of Tables viii

List of Figures ix

1 Introduction 1
- 1.1 Transcription . . . 1
  - 1.1.1 Stages of transcription . . . 2
- 1.2 Stalling of RNA polymerase . . . 4
  - 1.2.1 Reasons for RNA polymerase stalling . . . 5
- 1.3 RNA degradation . . . 6
- 1.4 Stochasticity . . . 6
  - 1.4.1 Stochastic simulation . . . 7
  - 1.4.2 Pipe model for transcription . . . 8
  - 1.4.3 Delay stochastic simulation method . . . 9
- 1.5 Plan of the thesis . . . 11

2 Model description 13
- 2.1 Detailed transcription model . . . 13
- 2.2 Simple RNA expression model . . . 15
  - 2.2.1 Initiation . . . 15
  - 2.2.2 Elongation . . . 16
  - 2.2.3 Termination . . . 17
  - 2.2.4 RNA degradation . . . 18
  - 2.2.5 Parameter Estimates . . . 19
- 2.3 Model of transcription with abortive initiation . . . 21
  - 2.3.1 Parameter estimates . . . 23

3 Stochastic simulation 25
- 3.1 Simple RNA expression model . . . 26
  - 3.1.1 Analysis of the simple RNA expression model . . . 26
  - 3.1.2 Dependence of RNA levels and rise time on kinetic parameters . . . 27
- 3.2 Delay stochastic simulation model with arrest . . . 38
  - 3.2.1 Analysis of delay stochastic simulation model with arrest . . . 38
  - 3.2.2 Dependence of RNA levels and rise time on kinetic parameters . . . 41
- 3.3 Summary . . . 47

4 Mathematical analysis **48**
4.1 Analytic theory of mean stationary RNA number 49
4.2 Analytic theory of rise time 54
4.2.1 At small values of k_8 54
4.2.2 At small values of k_6 59

5 Conclusions **62**
5.1 Summary 62
5.2 Conclusions 63
5.3 Future perspectives 64

Bibliography **66**

Appendices **71**

A **71**

List of Tables

2.1 Default kinetic parameters used in the model 20
2.2 Initial conditions 20
2.3 Default kinetic parameters used in model with abortive initiation 24

3.1 Tables showing the values of k_2 and k_{arrest} used to generate figure 3.17 . . . 40
3.2 Table showing the value of arrest probability corresponding to the value of k_{arrest} 43

List of Figures

1.1 Schematic diagram showing the fates of transcribing RNA polymerase during a transcription process 4
1.2 Pipe model of the transcription process 9
1.3 Algorithm for the delay stochastic simulation algorithm as given by Roussel and Zhu 10
1.4 Schematic diagram of RNA polymerase arrest in the model 11

2.1 Initiation 15
2.2 Elongation 16
2.3 Termination 18
2.4 Schematic diagram of RNA polymerase arrest in the model showing the definitions of τ_2 and τ_{as} 23

3.1 Curves showing the number of RNA molecules produced with respect to time in a single simulation of the simple RNA expression model 27
3.2 R_{ss} versus $\log_{10} k_x$ 28
3.3 Rise time versus $\log_{10} k_x$ 28
3.4 R_{ss} versus $\log_{10} k_2$ 30
3.5 Rise time versus $\log_{10} k_2$ 30
3.6 R_{ss} versus $\log_{10} k_6$ 31
3.7 Rise time versus $\log_{10} k_6$ 32
3.8 R_{ss} versus $\log_{10} k_7$ 33
3.9 Rise time versus $\log_{10} k_7$ 33
3.10 R_{ss} versus $\log_{10} k_8$ 34
3.11 Rise time versus $\log_{10} k_8$ 35
3.12 R_{ss} versus $\log_{10} \tau_1$ 36
3.13 Rise time versus $\log_{10} \tau_1$ 36
3.14 R_{ss} versus $\log_{10} \tau_2$ 37
3.15 Rise time versus $\log_{10} \tau_2$ 37
3.16 Curves showing the number of RNA molecules produced with respect to time by simulating the simple RNA expression model and by simulating the delay stochastic simulation model with abortive initiation 39
3.17 Curves showing the number of RNA molecules produced with respect to time by simulating the simple RNA model, the model with abortive initiation, and the simple RNA model with k_2 increased to 100 s^{-1} and the model with abortive initiation and k_2 increased to 100 s^{-1} 40
3.18 R_{ss} versus $\log_{10} k_{arrest}$ 42
3.19 Rise time versus $\log_{10} k_{arrest}$ 43

3.20 Rise time versus $\log_{10} k_{\text{arrest}}$ at several values of k_6 45
3.21 R_{ss} versus $\log_{10} \tau_3$ 46
3.22 Rise time versus $\log_{10} \tau_3$ 46

4.1 Analytical curve and simulation results for R_{ss} versus $\log_{10} k_x$ 54
4.2 Analytical curve and simulation results for R_{ss} versus $\log_{10} k_2$ 55
4.3 Analytical curve and simulation results for R_{ss} versus $\log_{10} k_8$ 55
4.4 Analytical curve and simulation results for R_{ss} versus $\log_{10} \tau_3$ 56
4.5 Analytical curve and simulation results for R_{ss} versus $\log_{10} k_{\text{arrest}}$ 56
4.6 Analytical curve and simulation results for the rise time versus $\log_{10} k_8$ . . . 58
4.7 Analytical curve and simulation results for the rise time versus $\log_{10} k_6$ . . . 60

Chapter 1

Introduction

In this chapter, I am going to discuss the transcription process and stochasticity. Later on, several stochastic simulation method will also be explained. At the end, I will talk about a delay stochastic simulation method to analyze the effects of delays in the process of prokaryotic transcription, and about my research plan.

1.1 Transcription

According to the central dogma, transcription is a stage in the gene expression process in which the information present in a DNA sequence is copied into an RNA molecule. The mature RNA molecule further forms protein with the help of ribosomes by a process called translation. The process of transcription is catalyzed by an enzyme called RNA polymerase [1]. The bacterial RNA polymerase is a multi-subunit enzyme consisting of five subunits: two α subunits, one β subunit, one β' subunit, and one ω subunit. During transcription, RNA polymerase reads the sequence from a template DNA strand and converts this into a complementary RNA strand. The process of transcription starts when the RNA polymerase binds with the transcription factor σ which then together bind with the template DNA strand, separating the double strands of the DNA helix. This process requires the breakage of the hydrogen bonds between the nucleotides [2]. RNA polymerase forms a complementary RNA strand by catalyzing the formation of an RNA sugar-phosphate backbone. During transcription, the RNA-DNA hybrid is base paired over a stretch of 8 or 9 nucleotides [3]. The RNA gets extruded through an exit channel in the polymerase. The

transcript is only released at the end after the events of termination.

I am specifically dealing with the prokaryotes in my research due to the easy availability of the relevant parameters as they are most studied and well-understood. This chapter will provide the necessary background information on transcription, leaving aside translation. There are three major steps in prokaryotic transcription as described in the following subsections.

1.1.1 Stages of transcription

Initiation

In bacterial transcription, the RNA polymerase binds with a transcription factor known as σ factor (to form the holoenzyme) at a specific site on the template strand of the DNA called the promoter [4]. The initial polymerase-DNA complex formed is known as the closed complex because at this point the DNA is double stranded. This complex will then unwind the DNA to expose a single stranded DNA sequence and the complex of separated DNA with the polymerase is then called the transcription bubble.

This bubble, which includes the RNA polymerase, the promoter and σ transcription factor, will then move forward, releasing the promoter. The nucleotides between the promoter region (+1 site) to the terminator site are transcribed into a complementary RNA sequence with the formation of phosphodiester bonds between the nucleotides to bind together the transcript [2]. Eventually, transcription factors leave and elongation proceeds. The RNA polymerase has the tendency to dissociate from the template strand after a few bonds are synthesized and generate truncated RNA transcripts, which is known as abortive initiation [5].

Abortive initiation

The process of abortive initiation is observed at the early stages of transcription. It is found that the RNA polymerase takes several attempts to escape a promoter and thus forms a very short nucleotide transcript of a length of about 9-32 nucleotides [5–7]. Several

conformational changes have also been observed within an RNA polymerase during the process of abortive initiation [8]. In addition to this, it has been found that approximately 75 percent of all the total initiations lead to the formation of abortive transcripts [7].

In prokaryotes, abortive initiation has been studied by a single molecule experiment but the exact mechanism is still unknown [5]. In the single molecule experiment, the technique of total-internal-reflection fluorescence microscopy was used to study the mechanism of abortive initiation in detail. The authors measured the distances between a fluorescent probe incorporated in RNA polymerase and a fluorescent probe incorporated in DNA with the help of fluorescence resonance energy transfer. It has been found that this process of abortive initiation is observed when the RNA polymerase was translocating forward on the template DNA strand and the release of the abortive product is considered to be a rate limiting step [5]. The abortive product is degraded by the ribonucleases present in the cell. After dissociation of the transcription complex, the promoter will be available for the next initiation event. If the RNA transcript does not abort during the initiation phase then it starts the elongation phase of transcription.

Elongation

In this phase, RNA polymerase moves on the template strand all the way from the 3' end to the 5' end and the newly formed RNA transcript is formed in the direction from 5' to 3'. This newly formed RNA transcript has the same sequence as that of the coding DNA strand except that the thymine (T) is replaced by uracil (U) [9].

Termination

There are two known mechanisms by which the prokaryotic transcription process terminates, known as the Rho independent termination process (also called intrinsic termination), and the Rho dependent process (also called extrinsic termination). In the Rho independent process, termination happens due to the formation of a G-C rich hairpin loop structure (from a palindromic sequence) followed by several uracils [2]. These hairpin loop structures are

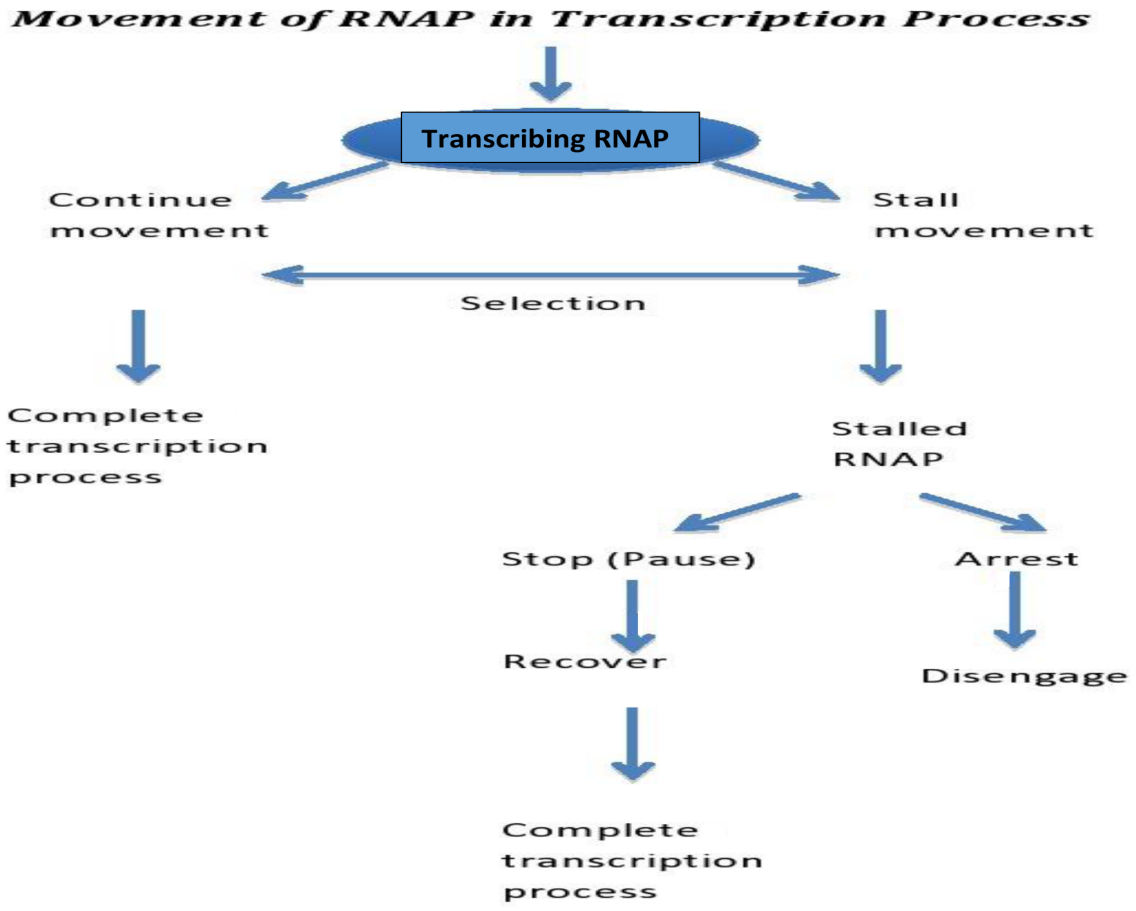

Figure 1.1: Schematic diagram showing the fates of transcribing RNA polymerase during a transcription process

very strong. This formation of the loop structure creates a mechanical stress, breaking the hydrogen bonds between the RNA-DNA hybrid and causing RNA polymerase to leave [9]. In the Rho dependent termination mechanism, the protein factor Rho binds with the nascent RNA transcript (which is still attached to the RNA polymerase) and this complex will then decrease the affinity between the RNA polymerase and the template DNA strand, causing the RNA polymerase to detach [10].

1.2 Stalling of RNA polymerase

Many biological processes like transcription do not always proceed from start to finish without interruption. This issue arises with processive enzymes, that is enzymes that work

on a polymer one monomer at a time without releasing the polymer [11]. These processive enzymes include RNA polymerase which carries out the transcription process. Stalling is a halt in the transcription process occurring when the RNA polymerase stops traversing the DNA strand. It can have two special cases, pausing and arrest [12] as shown in figure 1.1. As the name suggests, pausing is the reversible form of stalling, from which the polymerase can recover after some time leading to a normal transcription product later on. In contrast, arrest in the transcription process is irreversible and is usually not resolved by resumption of the transcription process, instead leading to the degradation of both the polymerase and the incomplete RNA transcript [12].

In the subsection below, I will give several reasons for stalling that can lead to the arrest of RNA polymerase during the elongation phase in a transcription process.

1.2.1 Reasons for RNA polymerase stalling

There are many reasons that can make an RNA polymerase stall. They are as follows:

- Sequence-dependent stalling: some sequences (e.g. homopolymeric-oligo(T) tracts) on the template strand are pre-determined to make RNA polymerases stall (pause or arrest) on them [13–15].
- The polymerase may attempt to read a damaged template, blocking its progression and causing arrest [12].
- The polymerase itself can be damaged or incorrectly assembled, again causing arrest [12].
- Barriers on templates, such as proteins, can also impede the progress of polymerases, resulting in pauses [12].
- DNA supercoiling may also lead to pausing [16].
- Low nucleotide pools cause polymerase pausing [12].

- RNA polymerases do not pass each other or displace one another, so that one stalled polymerase can cause upstream polymerases to pause [17].
- When two RNA polymerases collide in attempting to search for and then bind to the same promoter, one or both of them will arrest [17].
- Sometimes, there are promoters on the opposite strands that direct transcription in opposite directions, called convergent promoters. The transcription bubble can not accommodate polymerases traveling in opposite directions, and that leads to their arrest [17].

1.3 RNA degradation

Prokaryotic RNA is prone to degradation due to the presence of several ribonucleases in the cytoplasm. The rate of RNA degradation is studied on the basis of its half-life where half-life is directly related with the stability of RNA in a cellular environment. In *Escherichia coli*, the half-life of a nascent RNA varies over a wide range which is from 0.6 minutes to 36 minutes [18].

1.4 Stochasticity

A stochastic process is a random process that evolves with time [19]. These processes can evolve in more than one way as even if we know the initial conditions the process can have several different outcomes because they depend on the sequence of random variables.

At the microscopic level, chemical reactions occur at random times when, for instance, two reactants happen to collide in an appropriate orientation and with appropriate energy. The random nature of reactive events is a form of stochasticity. Biochemical processes (including transcription and translation) are stochastic in nature. In order to deal with these kinds of systems, we need a stochastic simulation method that mimics the reactions occurring in the system. These simulations use random variables; data are collected from several

realizations, each using different sets of random numbers. As a result, running multiple realizations of a stochastic process allows us to approximate the distributions of the random variables of the model.

Several stochastic simulation methods are given to deal with these stochastic processes, including the Gillespie algorithm [20], the Gibson and Bruck stochastic simulation algorithm [21], and the delay stochastic simulation method by Roussel and Zhu [22].

1.4.1 Stochastic simulation

In the year 1977, Gillespie presented an algorithm to simulate stochastic systems. In this method, several chemical and biochemical reactions are simulated to generate a trajectory or a possible solution. It is a discrete and stochastic simulation method that can handle complex reactions as long as they are decomposed into elementary reactions for which the products are obtained at the same time as the reactants are consumed [20]. As a result of this simulation, a trajectory is obtained that is an exact sample from the probability mass function that we would get by solving the chemical master equation. Each reaction in a stochastic model has a reaction propensity which gives the probability of the occurrence of a reaction per unit time. This reaction propensity function can be calculated with the help of a stochastic rate constant and the molecular populations of the reactants. It provides the rate at which the probability is transferred from one state to another [20].

For second-order reactions, for example, the collision of the molecules is required to obtain the products but the collision is successful only if the molecules are colliding with sufficient energy and in a correct orientation. So, this simulation method generates random numbers for choosing the sequence and timing of reaction events. Several simulations (or realizations) are run as each one will give different trajectories because of the random aspect of the dynamics. The data from several simulations can be analyzed to obtain statistical quantities of interest.

The Gillespie method is very expensive in terms of computational costs depending on

the number of reactions. Moreover, kinetic factors can affect the cost of the simulations. The probability of picking a reaction and the waiting time directly depend on the number of reacting molecules and the stochastic rate constants governing the reactions. For example, if we have a rapid equilibrium (with very large rate constants and reasonably large populations of the two species) say between species A and B, we could spend a lot of time converting A to B and B to A before a more significant reaction occurred [23].

We could build in detailed models of transcription and translation to a model for a gene network, but these models would become exceedingly complex and slow. By using delays to represent the transcription and translation steps, we take into account the time these processes take. In Gillespie's algorithm, all the reaction steps in a biological model are instantaneous and delayed effects, in which there is a time lag between the consumption of reactants and the appearance of products, are not included. In order to overcome these drawbacks, Gibson and Bruck gave a Next Reaction Method to make the Gillespie simulation method more time efficient [20] in the year 2000. Also, this new method allowed the effects of delays in a stochastic system to be included. In this method, a waiting time is generated for each possible reaction by generating a single random number per iteration and the reaction to occur next is the one with the smallest waiting time [21].

Furthermore, in order to deal with large reaction systems with a large number of delays, Bratsun et al. proposed an algorithm [24]. They modified the Gillespie algorithm [20] to account for the delay in kinetics which then could be used to simulate a stochastic model corresponding to delay differential equations [24].

1.4.2 Pipe model for transcription

The delays in gene expression (or other biochemical) models can be thought of as a pipe for example as shown in figure 1.2. In the pipe model, promoter and RNA polymerase enter the metaphorical pipe which is signifying the initiation of the transcription process here. Then, τ_1 time units after entering the pipe, the promoter is cleared, coming out of the

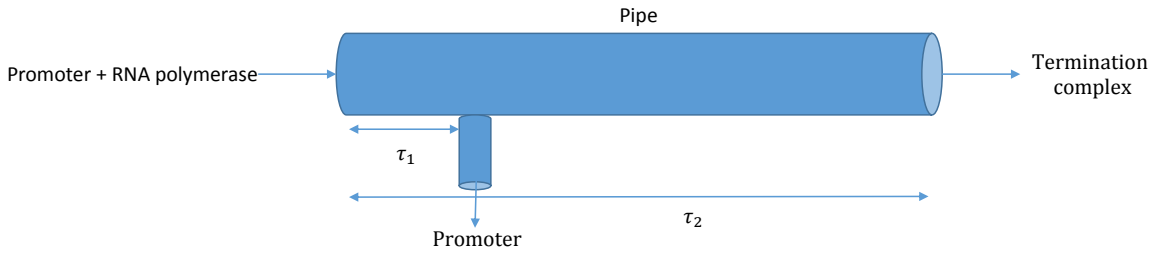

Figure 1.2: Pipe model of the transcription process.

pipe and becoming available for another binding event. At the end, the termination complex comes out of the pipe as a final product τ_2 time units after the process started. We cannot add anything to or take anything out of the pipe once the process starts as assumed in the model explained above.

In the conventional simulation algorithm, once a reaction is 'in the pipe', we cannot easily change it, which is what we need to do if there are eventualities that can occur after initiation of a reaction with delayed effects. The arrest of RNA polymerase is an example of the kind of thing that is difficult to deal with in the conventional algorithm (as shown in the pipe model) because it happens after transcription initiation (or translation initiation) and can result in different outcomes (abort or continue after a pause).

1.4.3 Delay stochastic simulation method

Here, I will present the algorithm used in my work from the paper presented by Roussel and Zhu [22], based on the work of Bratsun et al. [24]. Bratsun et al.'s paper appeared about five years after Gibson and Bruck's pioneering paper [21]. In the algorithm given by Roussel and Zhu [22], the presence of more than one delayed output in a single reaction step is considered unlike in the algorithm given by Bratsun et al. [24], where they assumed a single delayed event. Also, the version of the algorithm given by Roussel and Zhu, allows for delays drawn from any desired distribution, and not just fixed delays [22].

Figure 1.3 shows the algorithm given by Roussel and Zhu [22]. The delayed reactions are separated here into a set of reacting events (involving consumption of reactants) and

1. Initialize: set initial numbers of molecules; set $t \leftarrow 0$; form a group of reacting events and a separate group of generating events from the list of reactions; create an empty wait list L for delayed generating events.

2. Get the next reacting event R_1 and the corresponding occurrence time t_1 using the standard Gillespie stochastic simulation algorithm.

3. Compare t_1 with the least time, τ_{min}, in L.

   - If $t_1 \leq \tau_{\text{min}}$, get the generating event(s) corresponding to the reacting event R_1 and generate delays τ_{j} drawn from appropriate distributions for each generating event G_j. Then,
     (a) set $t \leftarrow t + t_1$;
     (b) update the number of molecules by performing R_1;
     (c) for each j such that $\tau_{\text{j}} = 0$, do G_j ;
     (d) decrement all the delays in L by t_1;
     (e) if $\tau_{\text{j}} \neq 0$, add G_j τ into L.
   - If $t_1 > \tau_{\text{min}}$,
     (a) set $t \leftarrow t + \tau_{\text{min}}$;
     (b) update the number of molecules by performing the generating event G_{min}, which is associated with τ_{min};
     (c) delete G_{min} τ_{min} from L;
     (d) decrement all the delays in L by τ_{min}.

4. Go to step 2.

Figure 1.3: Algorithm for the delay stochastic simulation algorithm as given by Roussel and Zhu [22].

Figure 1.4: Schematic diagram of RNA polymerase arrest in the model: On the DNA template strand, RNA polymerase is traversing from the promoter to the end. The stall site X is the pre-determined arrest site present on the template DNA strand as assumed in our model.

generating events (appearance of products). The algorithm in this model involves the use of a standard Gillespie trial by generating a set of random numbers to determine the next reacting event. The ordinary reactions don't pose any special problems and can be treated as in the ordinary Gillespie algorithm. The delayed generating events are stored in a list L, to be executed at the appropriate time in the future [22]. These generating events are denoted by G_j in the algorithm.

1.5 Plan of the thesis

In order to clearly understand cellular dynamics, we have to understand the process of transcription in detail [22]. Transcription is stochastic due to the low copy numbers of the molecules involved in this process [22]. In order to deal with the stochasticity in the transcription process, a stochastic simulation method is used that mimics the random timings of the reactions [20]. In order to further reduce the number of steps to be modelled in a stochastic process like transcription, delays are added to concentrate on inputs and outputs only [22]. For this, a delay stochastic simulation method was given by Roussel and Zhu [22]. In this thesis, a delay-stochastic model of transcription including the process of abortive initiation was developed. Data from this model and from a simplified version without abortive initiation were collected over many realizations and the results were analysed to understand the relationship between several parameters and model responses. The expanded model of transcription and the analytical theory of rise time (a measure of the time required to reach the stationary level of, in this case, the RNA population) can be employed

to analyse the stochastic processes in gene expression. Our results derived for the model with abortive initiation provide insight into how different stochastic reaction rates and delays associated with a transcriptional process can be used to predict the mean stationary RNA levels and rise time in gene expression. The response time has potential implications for understanding the response to environmental perturbations as it can help us to predict how fast gene expression changes when conditions change.

The very first challenge in this work was to modify and expand a previously given model [22] to account for abortive initiation. This is done in chapter 2. In this model, I assumed that the arrest of RNA polymerase happens during the elongation phase and the polymerase undergoes irreversible arrest as shown in figure 1.4. Here, the arrest is assumed to happen at an early stage even before the clearance of the promoter, which is called abortive initiation as explained earlier in this chapter [5]. This irreversible arrest results in the detachment of the polymerase from the template DNA strand which leads to the clearance of the promoter for a new initiation event. For that, I have added an extra step competing with productive elongation in the model given by Roussel and Zhu [22].

I compared results from models with and without arrest of RNA polymerase. In particular, the study of the stationary RNA level and the time taken to reach the stationary RNA level considering the effect of RNA polymerase arrest were studied in detail. This model will also help us to get some insight into the study of the dependence of the RNA levels on the abortive initiation rate, the time for which RNA polymerase stays in the arrested state, and the stochastic rate constants related to this step. Furthermore, we can also find a relationship between the rise time and the stochastic rate constants for the initiation, elongation and termination steps.

Chapter 2

Model description

The transcription of DNA in prokaryotes is a complex multi-step process. In the following section, a simple RNA expression model for bacterial transcription, which is a modified version of the model given by Marc R. Roussel and Rui Zhu [22], will be discussed. In this extended model, degradation of RNA is also taken into consideration as RNA produced at the end of the termination process is prone to degradation [25, 26]. This model includes four major processes: initiation, elongation, termination, and RNA degradation. The first three of these processes are parts of transcription. Several assumptions are made in this model and I am going to specifically look for parameters appropriate to *Escherichia coli*. I will then consider an extension that includes abortive initiation.

2.1 Detailed transcription model

In the model given by Marc R. Roussel and Rui Zhu [22, 27], they have considered a very detailed transcription process. They have included three states of each site on a template DNA strand which include unoccupied (U_i), occupied (O_i) and activated (A_i) states, with subscripts denoting sites on the template (nucleotide positions). In the reaction steps given below, Pro stands for promoter, RNAP stands for RNA polymerase, C_1 and C_2 are the two termination complexes, and R is the nascent RNA transcript. In the unoccupied state, it was assumed that there is no RNA polymerase present on this site. In the occupied state, there is an RNA polymerase present on the site. Finally, in the activated state there is a nucleoside triphosphate bound and the polymerase is ready to perform its catalytic action. All

these reaction steps including the elongation and termination in the end are controlled by the stochastic rate constants k_n, where n varies from 1 to 7. The three major steps involved in this previous model are given below:

- Initiation

$$\mathrm{Pro} + \mathrm{RNAP} + \mathrm{U_1} \xrightarrow{k_1} \mathrm{O_1} \tag{2.1}$$

$$\mathrm{O_1} \xrightarrow{k_2} \mathrm{A_1} \tag{2.2}$$

- Elongation

$$\mathrm{A_i} + \mathrm{U_{i+1}} \xrightarrow{k_3} \mathrm{O_{i+1}} + \mathrm{U_i}, \qquad i = 1,2,\ldots,n-1 \tag{2.3}$$

$$\mathrm{O_i} \xrightarrow{k_4} \mathrm{A_i}, \qquad i = 2,3,\ldots,n \tag{2.4}$$

- Promoter clearance

  It was assumed that the promoter region is only available to bind a new RNAP when the polymerase has moved to site m + 1. So, the reaction (2.3) is replaced by the reaction given below when $i = m$.

$$\mathrm{A}_m + \mathrm{U}_{m+1} \xrightarrow{k_3} \mathrm{O}_{m+1} + \mathrm{U}_m + \mathrm{Pro} \tag{2.5}$$

- Termination

$$\mathrm{A}_n \xrightarrow{k_5} \mathrm{C_1} \tag{2.6}$$

$$\mathrm{C_1} \xrightarrow{k_6} \mathrm{C_2} \tag{2.7}$$

$$\mathrm{C_2} \xrightarrow{k_7} \mathrm{RNAP} + \mathrm{R} + \mathrm{U}_n \tag{2.8}$$

On the other hand, I created a simplified model with delays, which is the basis for the work of this thesis and which will be described in more detail below. In the detailed model, each nucleotide can be in one of three states (U, O, A). However, in the delay model, the

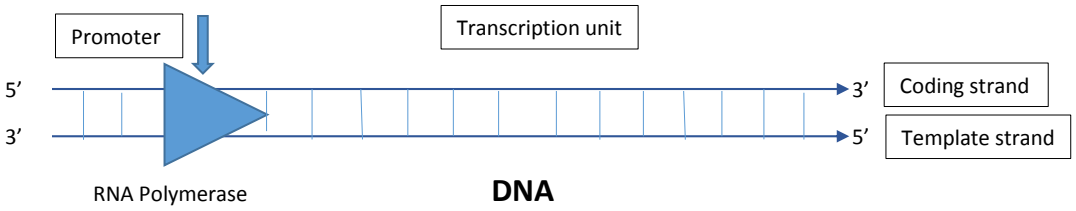

Figure 2.1: Initiation

entire elongation compartment is replaced by a delay, that is reactions (2.2), (2.3), (2.4) and (2.6) in the previous model are replaced by a single delayed process. In terms of the pipe metaphor of section 1.4, we can think of reaction (2.2) as 'entering the pipe' and of reaction (2.6) as 'exiting the pipe'. Once the reactants are in the pipe, we cannot alter anything before we get the product out from the other end of the pipe. I have also dealt with reaction (2.5) by considering a branch in the pipe from which the promoter comes out after a certain time delay. This decreases the overall number of steps in this model as compared to the previous model. The detailed model has 2N + 4 steps, where N is the number of nucleotides in the template. The average gene in prokaryotes has about 1000 nucleotides, so that's about 2000 steps [28]. As we will see in section 2.2, the corresponding parts of the delay model include just 4 steps.

2.2 Simple RNA expression model

2.2.1 Initiation

$$\text{Pro} + \text{RNAP} \xrightarrow{k_1} \text{ElnC} \tag{2.9}$$

The initiation phase of prokaryotic transcription involves the binding of RNA polymerase to the promoter as sketched in figure 2.1. The binding of RNA polymerase to the promoter forms an elongation complex (ElnC). This is basically the same as reaction (2.1) if we make the association of ElnC with O_1. Reaction (2.9) describes initiation as a single reaction, which is an approximation: initiation involves more than one reaction step [29].

Initiation is the slowest event in transcription [30].

The pool of free polymerases is expected to be more or less constant due to the large number of initiation and termination events going on in a cell at any given time [31, 32]. Hence, several kinetic studies of the binding of RNA polymerase to promoter have represented this process by a single pseudo-first-order stochastic rate constant, k_x, where,

$$k_x = k_1[\text{RNAP}] \tag{2.10}$$

So, the simplified equation would be:

$$\text{Pro} \xrightarrow{k_x} \text{ElnC} \tag{2.11}$$

The waiting time for transcription initiation as shown in equation (2.11) starting from a free promoter is controlled by k_x. Specifically, the waiting time is exponentially distributed with mean $1/k_x$. Note that the stochastic rate constants have units of s^{-1}.

2.2.2 Elongation

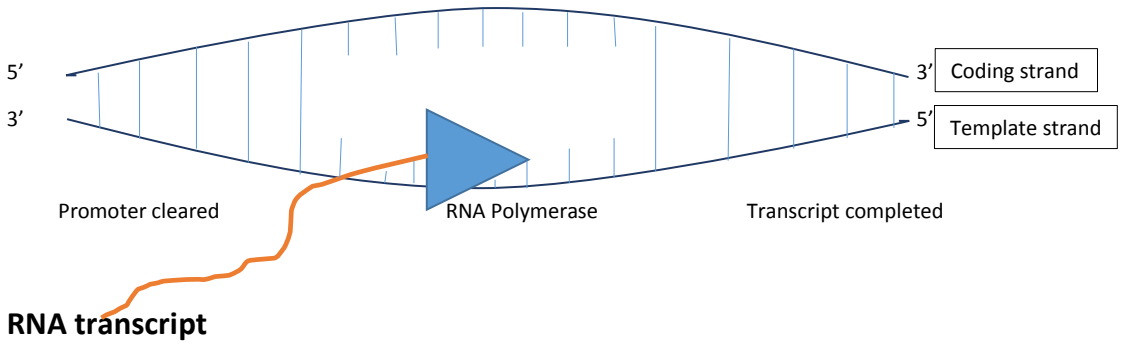

Figure 2.2: Elongation

$$\text{ElnC}(t) \xrightarrow{k_2} \text{Pro}(t+\tau_1) + \text{C}_1(t+\tau_2) \tag{2.12}$$

Initially, the elongation complex is formed as a result of the previous reaction (2.11). At some time after it has been formed, elongation starts. Call the time at which elongation initiates t. In addition to this, the promoter is cleared after time delay τ_1, making it available for another RNA polymerase to start another round of transcription as shown in figure 2.2. This single reaction step is corresponding to the reactions (2.2), (2.3), (2.4) and (2.6) given in the detailed model. The rate at which elongation complexes enter the pipe is controlled by rate constant k_2. This reaction step is same as reaction (2.2) and therefore has the same stochastic rate constant, k_2. It is important to note that the elongation phase in this model does not consider any other alternative pathways such as backtracking, RNA polymerase pausing or transcription arrest [12, 17, 33]. However, a particular form of arrest, abortive initiation, will be considered in section 2.3.

The mean waiting time between the formation of the elongation complex and initiation of elongation, as represented in reaction (2.12), i.e. the mean time it takes for an elongation complex to start transcription, is $1/k_2$. The promoter is cleared after time τ_1 and the termination complex C_1 is formed after time τ_2. Note that τ_2 is greater than τ_1 where the typical value of τ_2 is 20 s and τ_1 is 1 s [22]. To be noted, elongation is the only step in this simple RNA expression model that includes delays.

2.2.3 Termination

$$C_1 \xrightarrow{k_6} C_2 \tag{2.13}$$

$$C_2 \xrightarrow{k_7} R \tag{2.14}$$

Similarly to the detailed transcription model, the process of termination in this model involves two steps. They include the conversion of C_1 into a more stable state C_2 and then the release of RNA polymerase and RNA transcript from C_2. The two reactions (2.13) and (2.14) have the same stochastic rate constants as reactions (2.7) and (2.8) respectively.

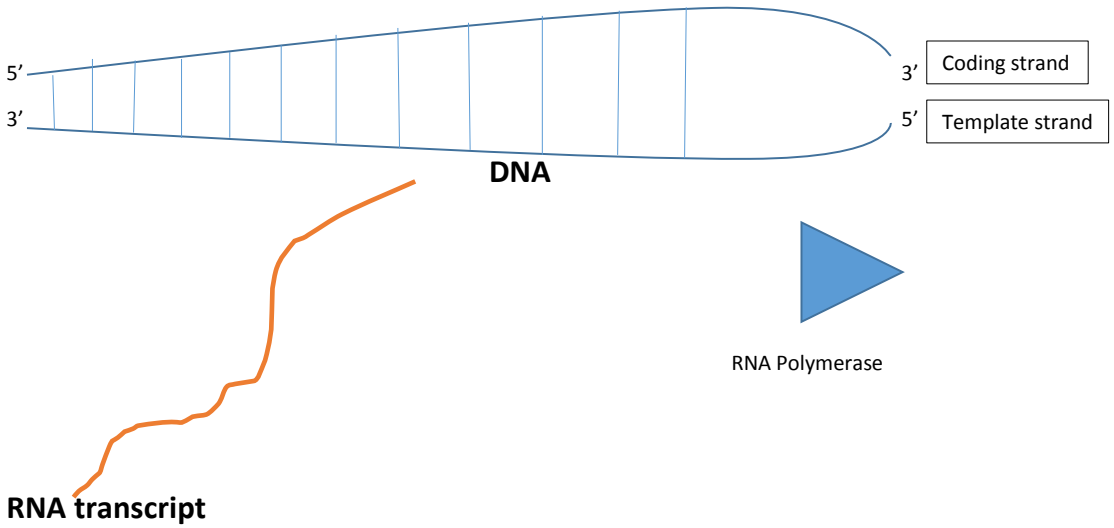

Figure 2.3: Termination

The two steps in the termination process in our model correspond to Rho-independent termination. Here equation (2.13) corresponds to the formation of a GC rich hairpin loop structure to form C_2, and equation (2.14) corresponds to the creation of a mechanical stress and breaking of the hydrogen bonds of the RNA-DNA heterodimer that releases RNA polymerase (not shown in equation (2.14) as a constant pool of RNA polymerase is assumed) and stops the transcription process.

The two reactions could also be understood to correspond to steps in Rho-dependent termination. However, the values of the rate constants would differ. In this termination mechanism, the protein factor Rho binds with the nascent RNA transcript (which is still attached to the RNA polymerase) and this complex will then weaken the interaction between the RNA polymerase and the template DNA strand, causing the RNA polymerase to detach [10]. Hence equation (2.13) could correspond to binding of Rho, and then equation (2.14) could correspond to dissociation of the Rho-polymerase-RNA-DNA complex.

2.2.4 RNA degradation

The newly synthesized RNA transcripts have short half-lives because they are prone to degradation by ribonucleases [25, 26]. Assume the degradation of RNA follows first-order

reaction kinetics,

$$\mathrm{R} \xrightarrow{k_8} \text{waste-product} \tag{2.15}$$

$$k_8 = \ln(2)/t_{\text{half}}$$

where t_{half} is the half-life of an RNA molecule in a prokaryotic cell. The complete simple RNA expression model consists of reactions from (2.11) to (2.15). I have added this reaction to the original model in order to allow for homeostasis in the RNA level.

2.2.5 Parameter Estimates

A dynamic process like transcription depends on the stochastic rate constants and the delays associated with it. These can be estimated from experimentally measured values. Their values and the conditions for their selection are given below. I am using parameters from reference [22] for consistency with this earlier study and the parameter values in this section provide a set of default parameters, but I will vary some parameter values to see their effects.

The pool of free RNA polymerases is roughly constant when the cells are grown in constant conditions. Therefore, the value of free RNA polymerases here is considered as 144 molecules which is taken from the previous literature as reviewed in [34] and [35]. And the value of k_1 in this model is assumed to be 10 s^{-1} as provided in [22]. Hence on putting the values of k_1 and the number of RNA polymerases in equation (2.10), the value of k_x would be equal to 1440 s^{-1}.

The value of k_2 is less than the value of k_1 because the transition from closed promoter complex to open promoter complex is a rate limiting step [36, 37]. Hence, I have adopted the value of the stochastic rate constant k_2 for the transition to elongation in the simple RNA expression model which is taken as 1 s^{-1} from [22]. In addition to this, the stochastic rate

Table 2.1: Default kinetic parameters used in the model

Parameter	Value	Source
k_x	1440 s^{-1}	[22] [34] [35]
k_2	1 s^{-1}	[22]
k_6	10 s^{-1}	[22]
k_7	1 s^{-1}	[22]
t_{half}	870 s	[25] [26]
k_8	$7.96 \times 10^{-4}\,s^{-1}$	[25] [26]
τ_1	1 s	[22]
τ_2	20 s	[22]

Table 2.2: Initial conditions

Variable	Value
Pro	1
ElnC	0
C_1	0
C_2	0
R	0

constant for the degradation of RNA transcript, k_8 is calculated on the basis of a typical half-life of RNA (14.5 minutes or 870 seconds) [25, 26]; this half-life corresponds to a k_8 value of $7.96 \times 10^{-4}\,s^{-1}$.

The delays in equation (2.12) are fixed in order to simplify the model. τ_1 is the time delay that an activated transcription complex spends before it releases the promoter for another binding event. The value of τ_1 used here is 1 s as in [22]. τ_2 is the time delay before the termination complex 1 is released and its average value is found to be 20 s [22]. Note that the value of τ_2 is always greater than the value of τ_1. The parameters used in this model are given in table 2.1.

The initial conditions used in this model are given in table 2.2. I assumed the presence of a single gene promoter in the model and at the beginning of the transcription process in our model. The populations of elongation complex, termination complexes and RNA are considered to be equal to zero.

Although this model is very time efficient and computationally less expensive than the

simulation of a detailed transcription model, it assumes ideal conditions for the transcription process by considering that the movement of RNA polymerase on the DNA template is smooth and continuous. It overlooks the fact that the movement of RNA polymerase is not always continuous and that the stalling of RNA polymerase is a very common phenomenon in the transcription process both in the case of prokaryotes and eukaryotes [12, 38, 39].

2.3 Model of transcription with abortive initiation

In order to take into account the effects of the stalling of RNA polymerase, I formulated an expanded version of this model that takes the abortive initiation into consideration. In this model, I assumed that the movement of RNA polymerase stops during the elongation step for an indefinite period of time and never resumes the movement again on the DNA template strand. To be specific, I am assuming the arrest of RNA polymerase at some of the early sites on the template strand in which the polymerase gets arrested even before the promoter is cleared. This type of initiation process in which arrest of RNA polymerase occurs before the clearance of the promoter is known as abortive initiation.

In abortive initiation, the halt in the movement of RNA polymerase doesn't result in a successful transcription product. After a certain time of RNA polymerase arrest, the arrested RNA polymerase detaches from the template strand. In this model, there is an arrest process competing with the elongation step. I have decided to study abortive initiation in $E.\ coli$ as it is a very simple organism and there is a plenty of literature available in order to have easy access to the parameters to be used.

In the section below, I will explain the reaction steps for the elongation process in which the RNAP gets arrested, and present the kinetic parameters used for this model. All other steps in this model are the same as given in the previous model. Here, I have assumed two alternative fates for the elongation complex out of which only one will be chosen as there will be a competition between these two [7].

In the first scenario, if the continuous/unstalled movement of RNA polymerase occurs

as shown in equation (2.12), then the elongation complex produced in equation (2.11) will continue the elongation and proceeds to completion as in the simple RNA expression model. This process results in a successful RNA transcript at the end of the transcription cycle.

In the second scenario, the elongation complex produced in equation (2.11) starts the elongation process but gets halted due to the arrest of RNA polymerase on the template DNA strand as shown in equation (2.17).

$$\mathrm{ElnC}(t) \xrightarrow{k_{\mathrm{arrest}}} \mathrm{Pro}(t + \tau_{\mathrm{as}} + \tau_{\mathrm{arrest}}) \tag{2.16}$$

Here, I have assumed that the arrest of RNA polymerase happens at a very early stage at about 9–32 nucleotides downstream from the start site [5–7]. This kind of arrest is a special case that results in abortive initiation. Here, I assume that the arrest happens at about 20 nucleotides downstream. That means that, after the elongation starts and RNA polymerase travels about 20 nucleotides, as shown in figure 2.4, it gets arrested at a site I label 'X'. It takes an amount of time τ_{as} to reach the arrest site 'X', and the polymerase remains arrested there for time τ_{arrest}. So τ_{arrest} time units after the arrest happened (at time (t + τ_{as})), the RNA polymerase gets detached from the template strand and the promoter is released at the same time. This step doesn't result in any successful transcription product. The stochastic rate constant for this step is k_{arrest}. Here, I have assumed a term τ_3 which is considered as the sum of τ_{as} and τ_{arrest}. So, equation (2.16) could be rewritten as (2.17) by replacing the sum of τ_{as} and τ_{arrest} as τ_3.

$$\mathrm{ElnC}(t) \xrightarrow{k_{\mathrm{arrest}}} \mathrm{Pro}(t + \tau_3) \tag{2.17}$$

If the polymerase engages in productive elongation according to equation (2.12), then the termination steps that follow are as before, and RNA degradation is still assumed to occur.

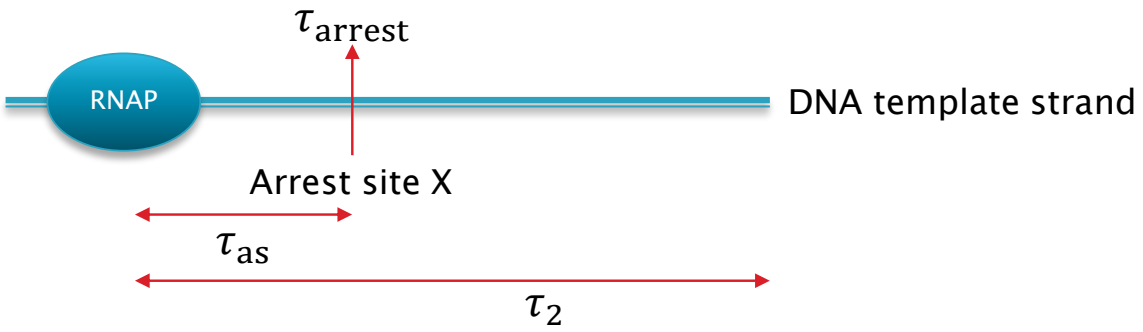

Figure 2.4: Schematic diagram of RNA polymerase arrest in the model: On the DNA template strand, RNA polymerase is traversing from 3' end to the 5' end. It takes τ_{as} time units to reach the arrest site 'X' where it stays arrested for τ_{arrest} time units and gets detached from the template strand after time (τ_{as} + τ_{arrest}) after the start of transcription. After the detachment of the RNA polymerase, the promoter is cleared. If the arrest of RNA polymerase doesn't happen then, termination complex C_1 is released τ_2 time units later, eventually resulting in a normal transcription product. Note, $\tau_2 > \tau_{as}$.

2.3.1 Parameter estimates

The delay stochastic simulation model with abortive initiation is an expansion of the simple RNA expression model that was explained earlier. So, the majority of the parameters used in this model are the same as used in the simple RNA expression model, except a few that are newly introduced in the model with arrest. These newly introduced parameters are k_{arrest} and τ_3.

The stochastic rate constant for the elongation step with arrested RNA polymerase, i.e. k_{arrest}, was calculated using the probability of arrest (Pr_{arrest}), as shown in equation (2.18).

$$Pr_{arrest} = k_{arrest}/(k_{arrest} + k_2) \tag{2.18}$$

Here, k_{arrest} + k_2 is the total rate of initiation. The rate of initiations leading to arrest (in initiations per second) divided by the total rate of initiations (same units) gives the probability of arrest.

The probability of arrest was reviewed from the literature and it was found that on average there is a thirty to seventy percent probability for an RNA polymerase to undergo abortive initiation during the elongation process [16]. The value of k_2 is 1 s^{-1} based on

Table 2.3: Default kinetic parameters used in model with abortive initiation

Parameter	Value	Source
k_1	10 s^{-1}	[22]
k_2	1 s^{-1}	[22]
k_{arrest}	0.428 s^{-1}	[22] [16]
k_6	10 s^{-1}	[22]
k_7	1 s^{-1}	[22]
t_{half}	870 s	[25] [26]
k_8	7.96×10^{-4} s^{-1}	[25] [26]
τ_1	1 s	[22]
τ_2	20 s	[22]
τ_{as}	0.5 s	[7] [40]
τ_{arrest}	1 s	Assumed

the concentrations and dissociation constants used in [22]. Hence, by putting the values of Pr_{arrest} which is 0.3 [16] and k_2 in equation (2.18) the value of k_{arrest} is found to be 0.428 s^{-1}. In order to analyze the effect of arrest probability on other parameters, I have varied this value of arrest probability in the next chapter in section 3.2.2.

In addition to this, the other parameter τ_{as}, which is the time taken by the RNA polymerase to reach to the arrest site 'X', was also estimated on the basis of the transcribing speed of the RNA polymerase and location of the arrested site 'X' on the DNA template strand. In this model, I have considered that the RNA polymerase transcribes at a speed of about 40 nucleotides per second [41] and the arrest site 'X' is present about 20 nucleotides away from the initiation site [7]. So,

$$\tau_{as} = 20\,\text{nt} / 40\,\text{nt}\,\text{s}^{-1} = 0.5\,\text{s} \tag{2.19}$$

Finally, the value of the parameter τ_{arrest} was assumed to be 1 s as there are some relevant data given in the literature [40] that confirm that the arrest leading to abortive initiation is not particularly long of the order of seconds. Table 2.3 gives the values for all the parameters that are used in this model.

Chapter 3

Stochastic simulation

Small molecular populations present in biochemical systems give rise to stochasticity [42]. In order to deal with these small molecular populations, we need a stochastic simulation method, such as the delay stochastic simulation method given by Roussel and Zhu in 2006 [22], to model these systems correctly. This delay stochastic simulation method helps us to determine the interdependence of the biochemical reactions on each other and to study their effects on the overall biochemical process. Roussel and Zhu considered a delay-stochastic model of transcription in their paper [22]. This model was very similar to the simple model that I described in chapter 2. However, no cases of RNA polymerase arrest were considered in the paper presented by Roussel and Zhu [22]. The simulations were done for the continuous movement of RNA polymerase from initiation to termination and then RNA degradation. Also, they didn't look at transient effects in this paper including rise time, where rise time is a measure of the speed at which the system responds to gene activation.

In order to take into account the effects of the stalling of RNA polymerase, I formulated an expanded version of this model that takes the abortive initiation into consideration. This chapter will present my results on the simple RNA expression model [22], followed by the modified version of this model that accounts for the abortive initiation of an RNA polymerase. Finally, I will discuss the simulation results for the modified model and study the dependence of the variables on several parameters. This study focuses on mean stationary RNA concentration (R_{ss}) and rise time.

3.1 Simple RNA expression model

In complex biochemical reactions like transcription where several delayed reactions are occurring at any time, it is computationally quite expensive to consider the state of each site of a DNA template [24,43]. To deal with this, Roussel and Zhu presented a model for simple RNA expression studied using a delay stochastic simulation algorithm [22]. This algorithm can handle processes with several delayed outputs which have stochasticity associated with them. It mimics the random timings of individual biochemical reactions at relatively little computational cost. It also allows us to take into account the non-negligible time it takes to transcribe a gene without having to simulate these processes in detail. The steps of the transcription process were modeled as described in section 2.1.

3.1.1 Analysis of the simple RNA expression model

The simple RNA expression model that was described in section 2.1 was simulated using the delay stochastic simulation method. The simulation code can be found in Appendix A. In this section, analysis of the simple RNA expression model using the delay stochastic simulation method is discussed in detail. Here, I will discuss the simulation results and the dependence of rise time and R_{ss} on several parameters used in the simple RNA expression model. The simulations were carried out using a set of parameters given in table 2.1. Figure 3.1 shows the number of RNA molecules produced with respect to time in a typical simulation of the simple RNA expression model. The fluctuations around the mean level in figure 3.1 are due to the degradation and synthesis of the RNA, simultaneously. At the stationary state, the average rate of synthesis of RNA is equal to the average rate of degradation of RNA. In order to take an average, because of the fluctuations, the simulations were run 10,000 times and then I have taken the average for all the stationary RNA numbers to get the mean stationary RNA number or R_{ss}. The rise time is defined as the time required for the RNA population to reach 90% of R_{ss} [44, 45]. Specifically, it is the first passage time at 90% of the R_{ss} level [44, 45]. Different studies use different thresholds for the rise time

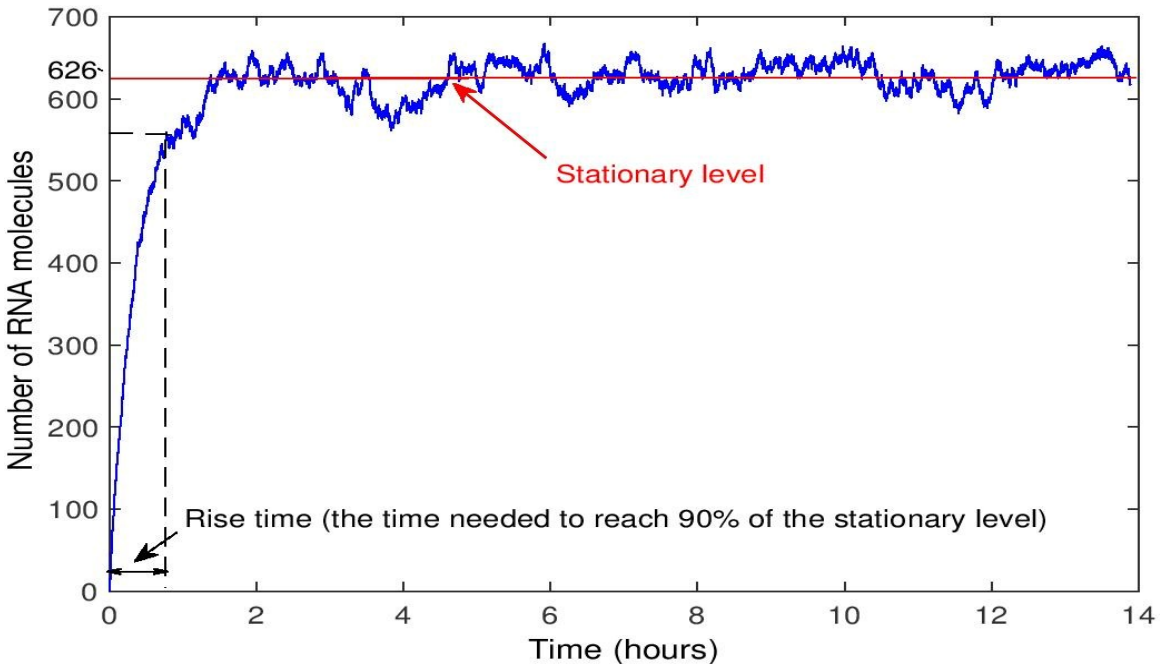

Figure 3.1: Curve showing the number of RNA molecules produced with respect to time in a single simulation of the simple RNA expression model. The simulation result is shown in blue with a mean stationary level of 626 molecules shown in red. This simulation was run for the rate constants of Table 2.1. The rise time is estimated here as the time taken to reach ninety percent of the stationary RNA level, 563 RNA molecules in this case.

ranging from 10% to 90% [46]. The 90% level was chosen as an arbitrary value here in order to closely estimate the time required to reach the mean stationary RNA level.

3.1.2 Dependence of RNA levels and rise time on kinetic parameters

The level of stationary RNA molecules and time taken to achieve this level depends on several kinetic parameters [47]. The parameters used in our model include k_x, k_2, k_6, k_7, k_8, τ_1 and τ_2 where k_n is the stochastic rate constant and τ_n is the time delay between the consumption of reactants and appearance of products as shown in equation (2.12). Even a small change in the values of these parameters can lead to a considerable variation in RNA levels [48]. However, not all the parameters have the same effects on final RNA levels. To study the effects of the parameters on the mean stationary RNA levels and rise time, I have generated several sets of simulation results.

Effect of varying k_x on mean stationary RNA level and rise time

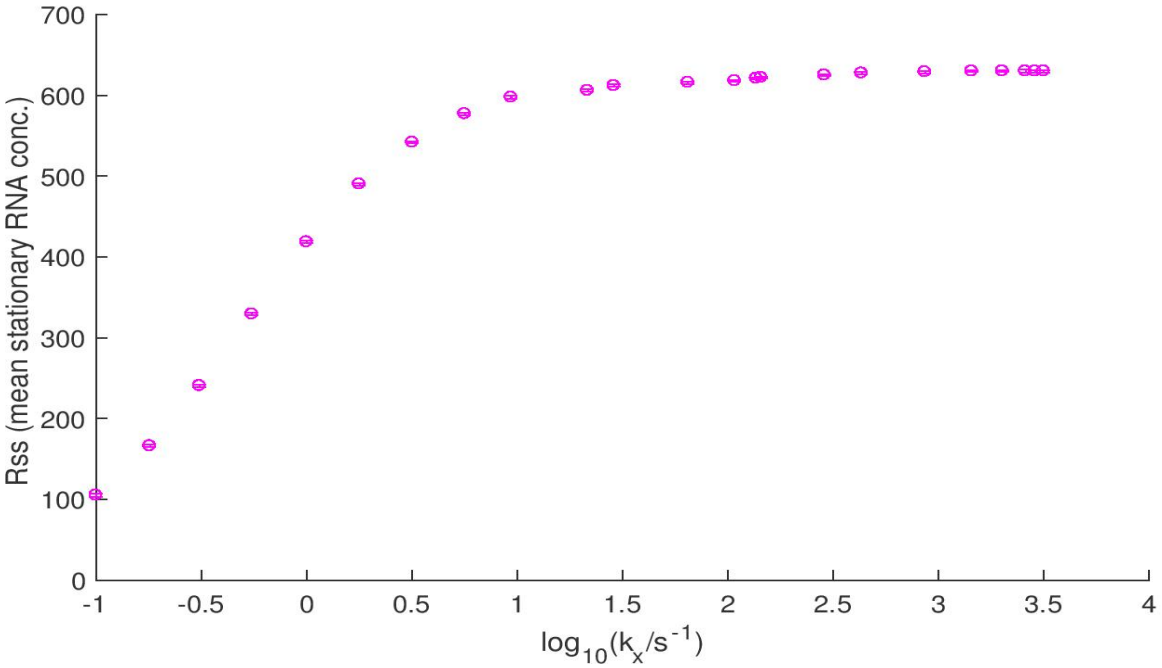

Figure 3.2: R_{ss} versus $\log_{10} k_x$. All parameters, except k_x, are from Table 2.1. The standard errors in the mean values are smaller than the plotting symbols in this figure.

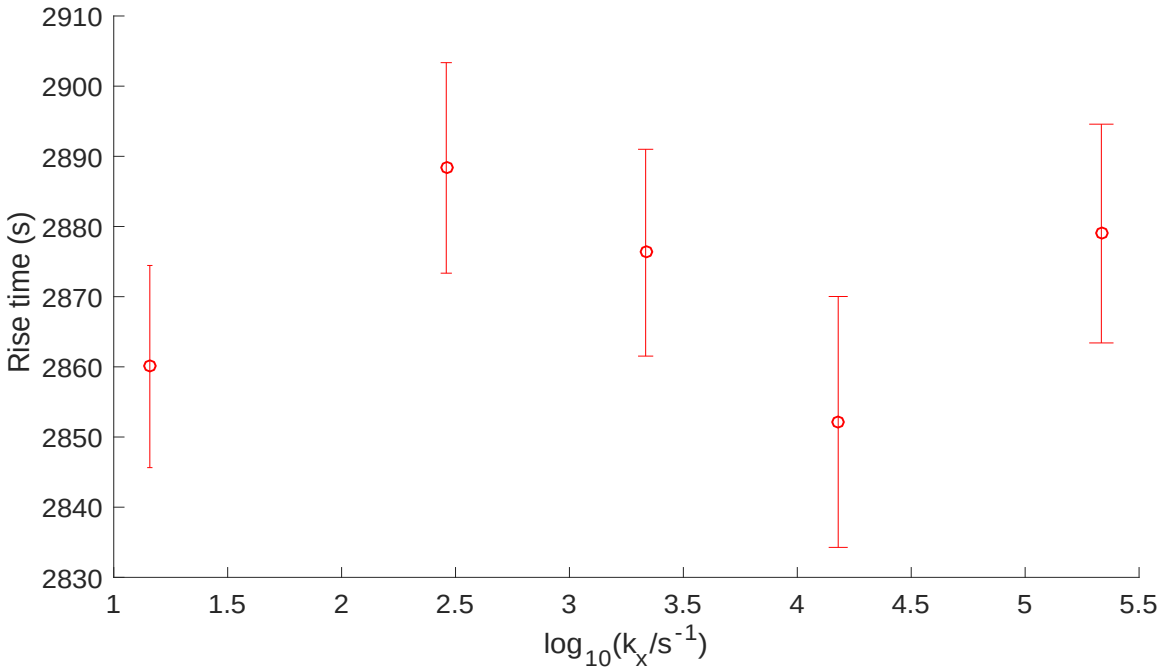

Figure 3.3: Rise time versus $\log_{10} k_x$. All parameters, except k_x, are from Table 2.1. Here, error bars represent the standard error of the mean. The meaning of the error bars will be the same in subsequent figures.

k_x is the stochastic rate constant for the initiation step of the transcription process as shown in equation (2.3). This step can be rate determining in the transcription process [30]. Hence, we can expect that the variation in the value of k_x will have a great impact on R_{ss}.

Figure 3.2 depicts that increasing the value of k_x increases R_{ss}, the mean stationary number of RNA molecules produced during the transcription process, to a certain limit but at the bigger values of k_x, the value of R_{ss} becomes independent of the value of k_x. The steady-state level in these systems is set by a balance of the production and destruction terms. As mentioned above, the transcription initiation is the rate-determining step for transcription (production) [30], so we would expect the steady-state level to depend on k_x, at least until the value of k_x is large enough that it is no longer rate-limiting [49]. If we increase the rate of binding of RNA polymerase to the promoter, we expect the mean level of RNA to increase. Saturation in this case is caused by promoter saturation, i.e. by polymerases having to wait for the previous polymerase to clear the promoter. Note that in figure 3.2, estimate of the uncertainty or standard error is smaller than the size of the plotting symbols in this figure.

In addition to this, from figure 3.3 we can tell that the value of rise time is independent of the value of k_x. I will discuss the reason for the independence of the rise time from k_x in chapter 4.

Effect of varying k_2 on mean stationary RNA level and rise time

Here, k_2 is the stochastic rate constant for the transition from initiation to elongation step of the transcription process as shown in equation (2.12). Figure 3.4 shows that the value of R_{ss} increases on increasing the value of k_2. The values of k_2 considered here are much smaller than the value of k_x from Table 2.1. So, we can say that k_2 being rate-limiting, if we increase this rate constant, the production of RNA is faster. So, in the same amount of time, more RNA molecules would be produced at the end.

In addition to this, from figure 3.5 we can see that initially on increasing the value of

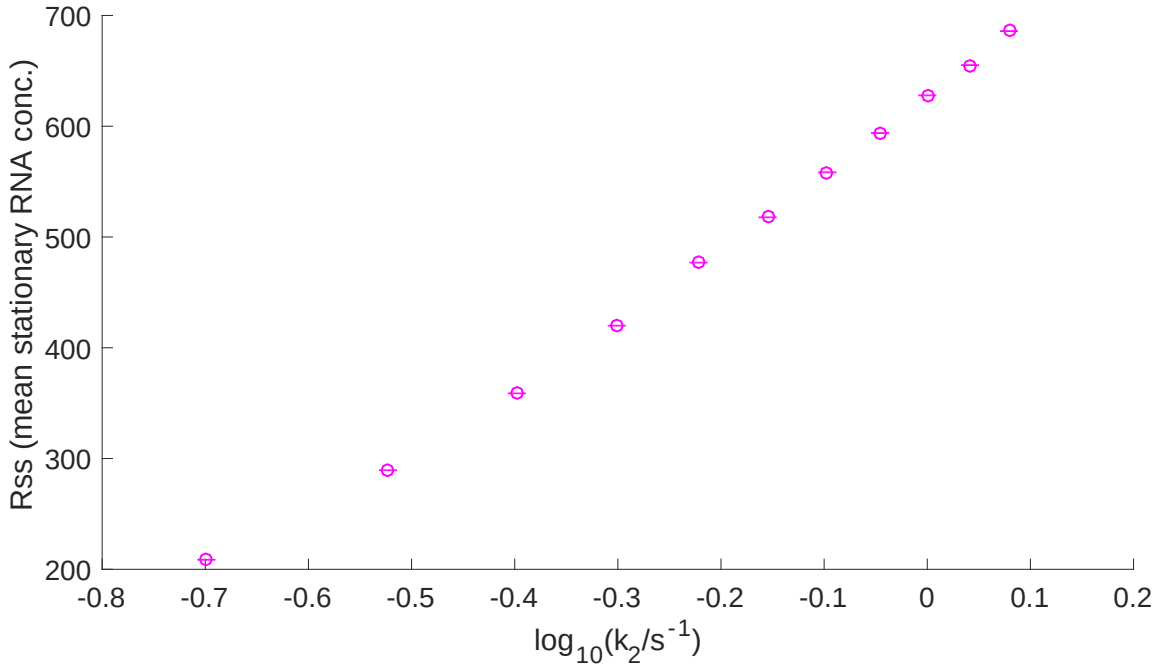

Figure 3.4: R_{ss} versus $\log_{10} k_2$. All parameters, except k_2, are from Table 2.1.

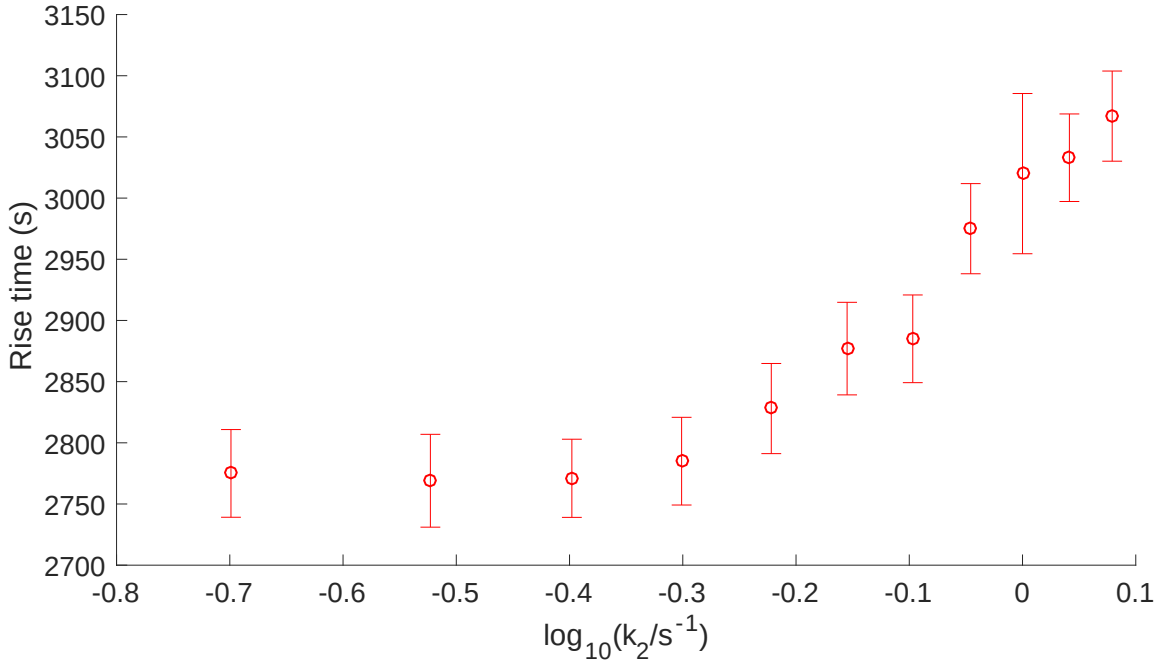

Figure 3.5: Rise time versus $\log_{10} k_2$. All parameters, except k_2, are from Table 2.1.

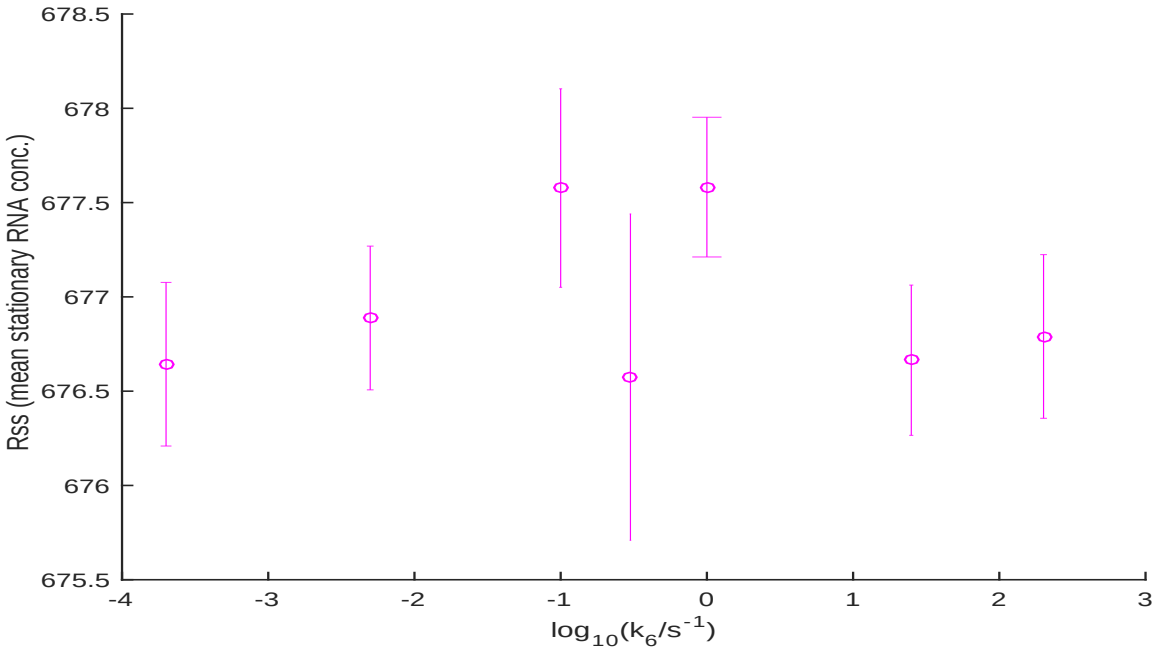

Figure 3.6: R_{ss} versus $\log_{10} k_6$. All parameters, except k_6, are from Table 2.1.

k_2 the value of rise time didn't change much. Specifically, the rise time starts to increase somewhere between $\log_{10} k_2$ equals –0.4 and –0.3. It is interesting that the rise time curve is flat for $\log_{10} k_2 \leq$ –0.4, but the reason for this is not clear. After the value of $\log_{10} k_2$ reaches –0.3, the stochastic rate constant k_2 had a significant impact on the value of rise time. The rise time increases when we increase k_2. To put it another way, making the entry into productive elongation faster causes the rise time to go up.

Effect of varying k_6 on mean stationary RNA level and rise time

k_6 is the stochastic rate constant for the termination step where the termination complex 1 is converting into termination complex 2 as shown in equation (2.13). Figure 3.6 shows that increasing the value of k_6 doesn't change the mean RNA number.

However, figure 3.7 shows that increasing the value of k_6 decreases the value of rise time significantly at smaller values of k_6. In other words, for the reaction process where the termination complex 1 is converting to termination complex 2, if we make this reaction step faster (up to a certain limit) then it takes less time to reach the stationary state. But when k_6

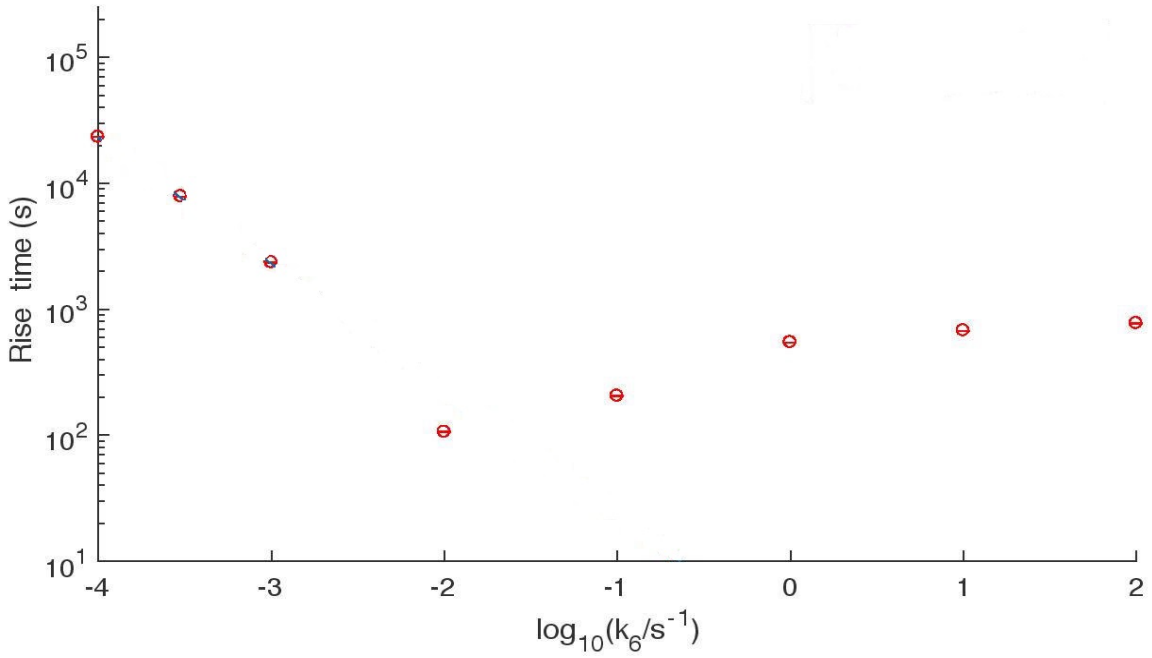

Figure 3.7: Rise time versus $\log_{10} k_6$. All parameters, except k_6, are from Table 2.1.

becomes large enough, it no longer controls the rise time.

Effect of varying k_7 on mean stationary RNA level and rise time

k_7 is the stochastic rate constant for the termination step of the transcription process as shown in equation (2.14). As shown in figure 3.8 for R_{ss} versus k_7, the value of R_{ss} is independent of the value of k_7. I will tackle this issue in chapter 4.

However, as shown in figure 3.9, k_7 has a significant impact on the rise time at very small values. On increasing the value of the stochastic rate constant k_7, the value of the rise time decreases initially and then becomes independent at large enough values of k_7. This means that really small values of k_7 make this step rate limiting, and the rise time will then be largely dependent on this rate constant.

Effect of varying k_8 on mean stationary RNA level and rise time

As shown in equation (2.15), the stochastic rate constant k_8 is responsible for the rate of degradation of the RNA transcript produced. A comparatively big value of k_8 denotes

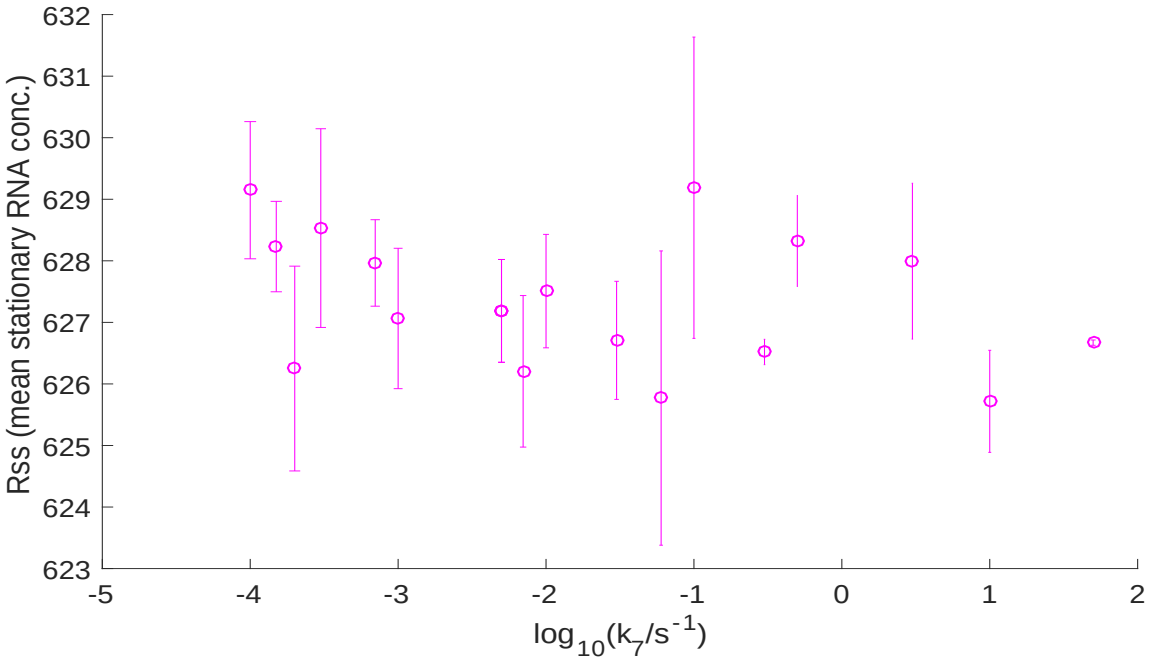

Figure 3.8: R_{ss} versus $\log_{10} k_7$. All parameters, except k_7, are from Table 2.1.

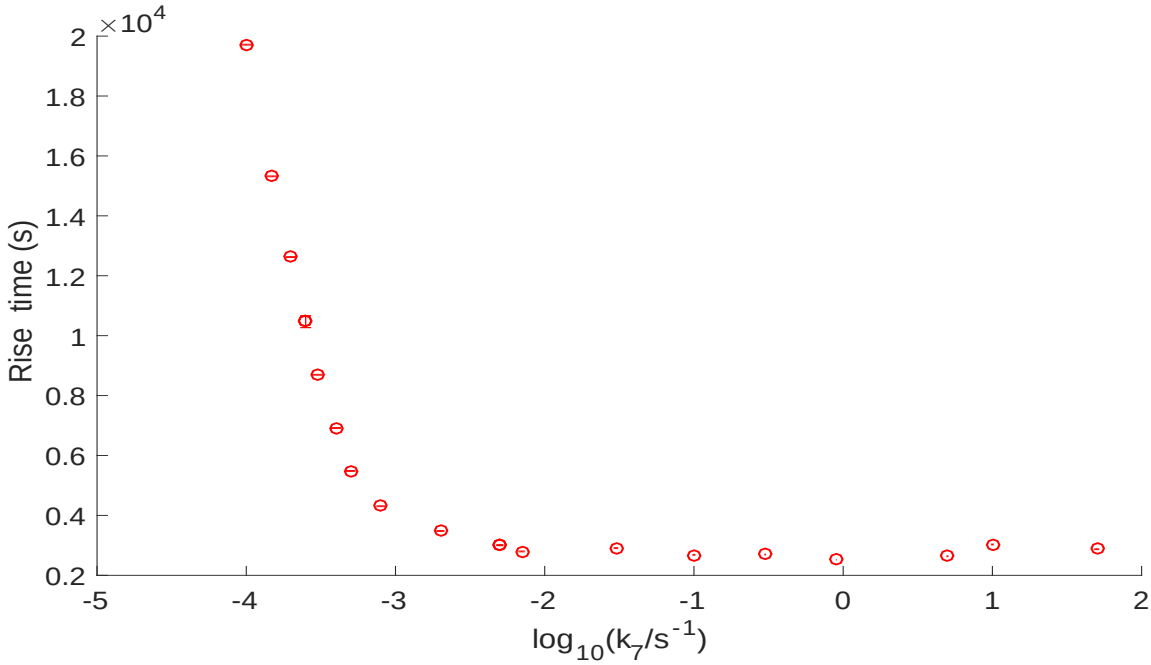

Figure 3.9: Rise time versus $\log_{10} k_7$. All parameters, except k_7, are from Table 2.1.

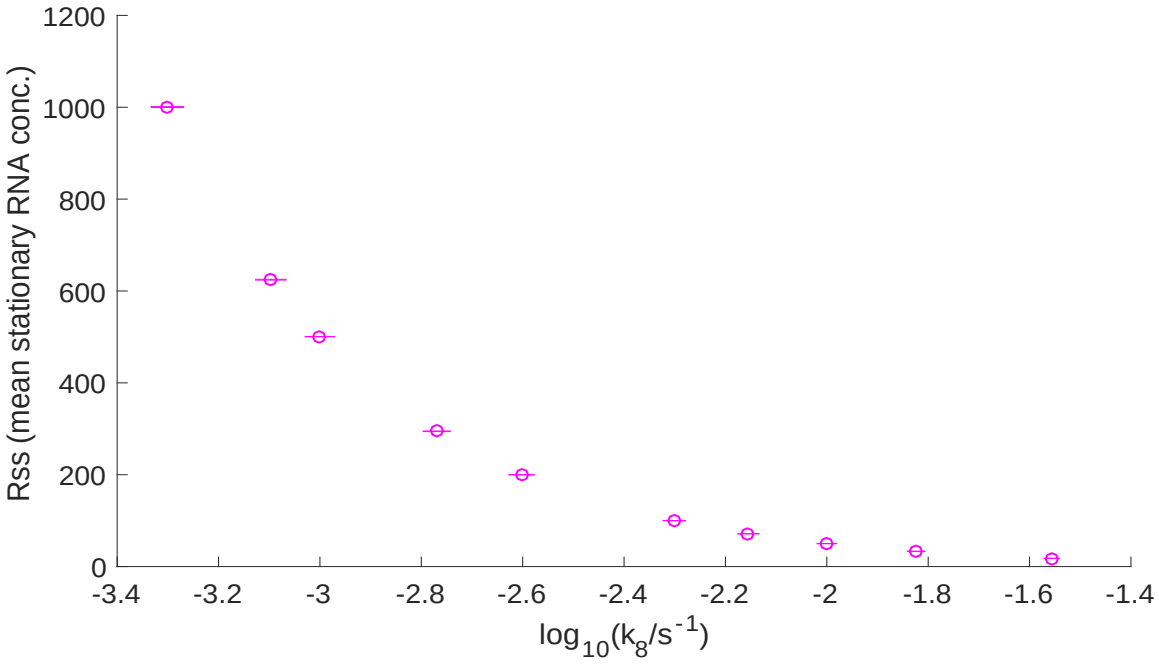

Figure 3.10: R_{ss} versus $\log_{10} k_8$. All parameters, except k_8, are from Table 2.1.

a faster RNA degradation, hence decreasing the mean stationary RNA number or R_{ss} as shown in figure 3.10.

In addition to this, it is simply a case of k_8, the smallest rate constant, limiting the rate at which the steady state is approached. Hence, rise time decreases on decreasing the value of k_8. A wide range of the values of k_8 is plotted here in figures 3.10 and 3.11. The values of k_8 plotted here are based on the range of mRNA lifetimes in *E. coli* (i.e. $1/k_8$) which are found to be between 0.6 and 36 minutes [18].

Effect of varying τ_1 on mean stationary RNA level and rise time

As shown in equation (2.12), τ_1 is the time delay for the clearance of promoter after the initiation of elongation at time t. So, if the promoter gets cleared more quickly or, in other words, if the delay associated with promoter clearance is decreased, then the number of RNA polymerases getting attached to the promoter per unit time increases, hence leading to an increased number of transcription cycles per unit time. So the overall yield of the RNA molecules increases. The range of τ_1 values plotted here is decided on the basis of

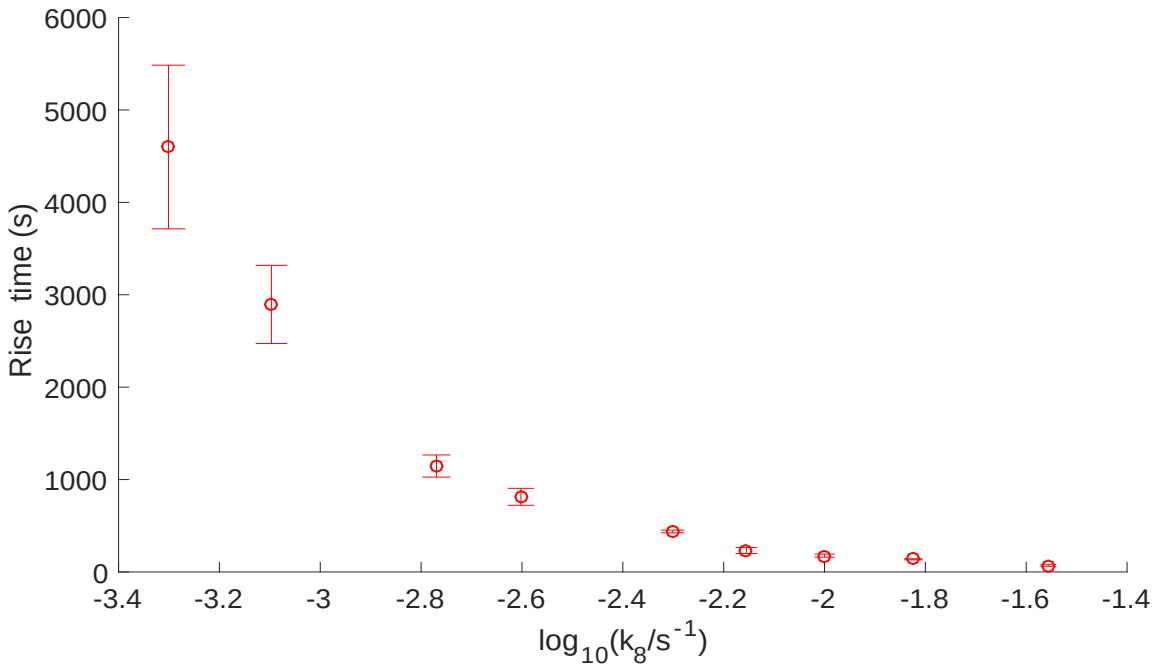

Figure 3.11: Rise time versus $\log_{10} k_8$. All parameters, except k_8, are from Table 2.1.

the physically reasonable range of values of τ_1.

In conclusion, increasing the delay time τ_1 decreases the mean RNA number produced during the transcription process as shown in figure 3.12. On the other hand varying τ_1 has no impact on rise time as shown in figure 3.13. Typically, when R_{ss} decreases, we expect the rise time to also decrease since it takes less time to reach any given level. However, in this case the slower re-initiation frequency balances the decrease in R_{ss}.

Effect of varying τ_2 on mean stationary RNA level and rise time

From equation (2.12), τ_2 is the time delay in the release of termination complex 1 after the formation of elongation complex at time t. So in terms of the pipe metaphor, a longer delay corresponds to a longer pipe. If we have a longer pipe at some given initiation rate, we can have more transcription complexes in the pipe, but in the steady state, they come out at exactly the same rate as they go in, regardless of the length of the pipe. So, it doesn't directly impact the RNA number in the transcription process overall. Hence, changing the value of τ_2 will not affect the value of R_{ss} and hence doesn't affect rise time as well. The

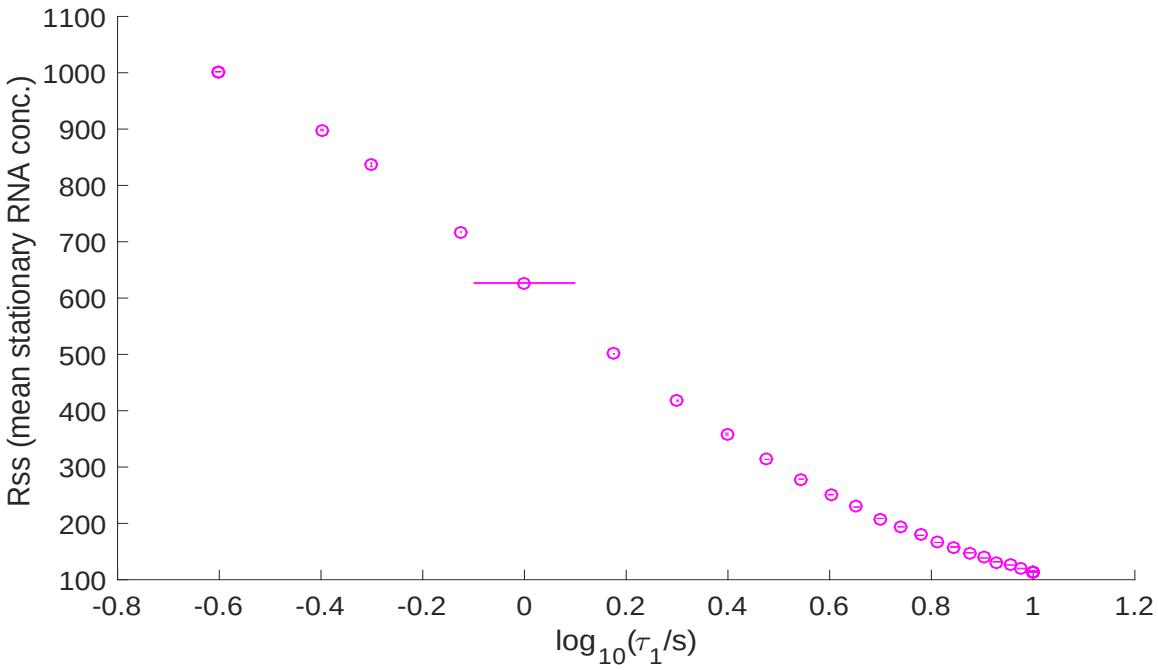

Figure 3.12: R_{ss} versus $\log_{10}\tau_1$. All parameters, except τ_1, are from Table 2.1.

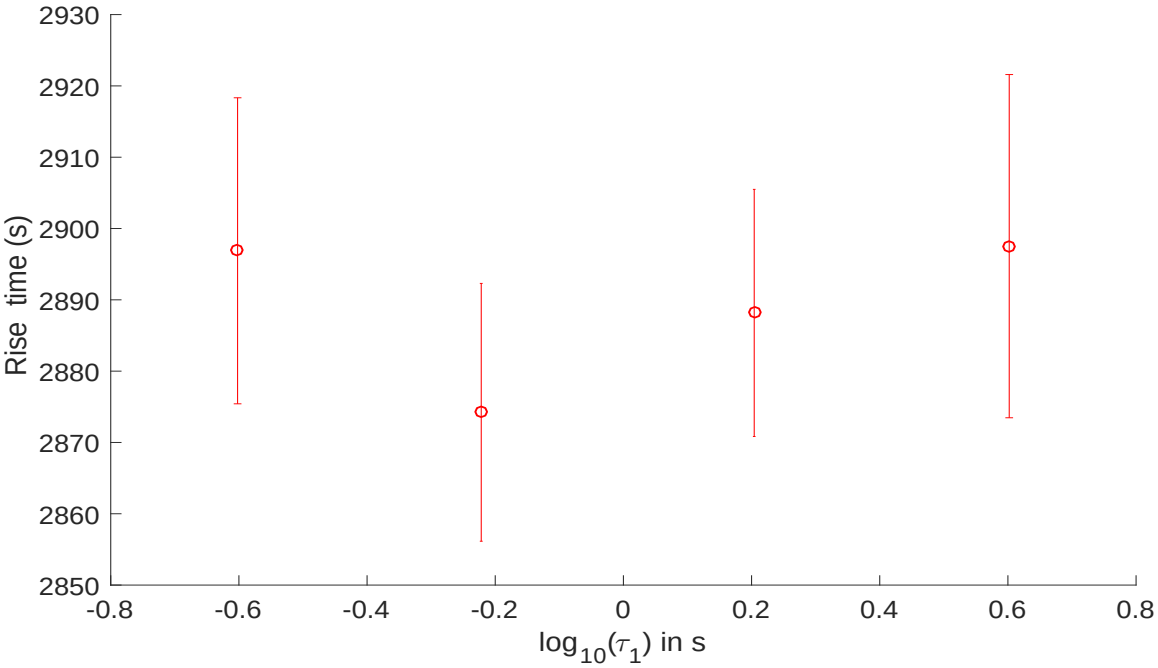

Figure 3.13: Rise time versus $\log_{10}\tau_1$. All parameters, except τ_1, are from Table 2.1.

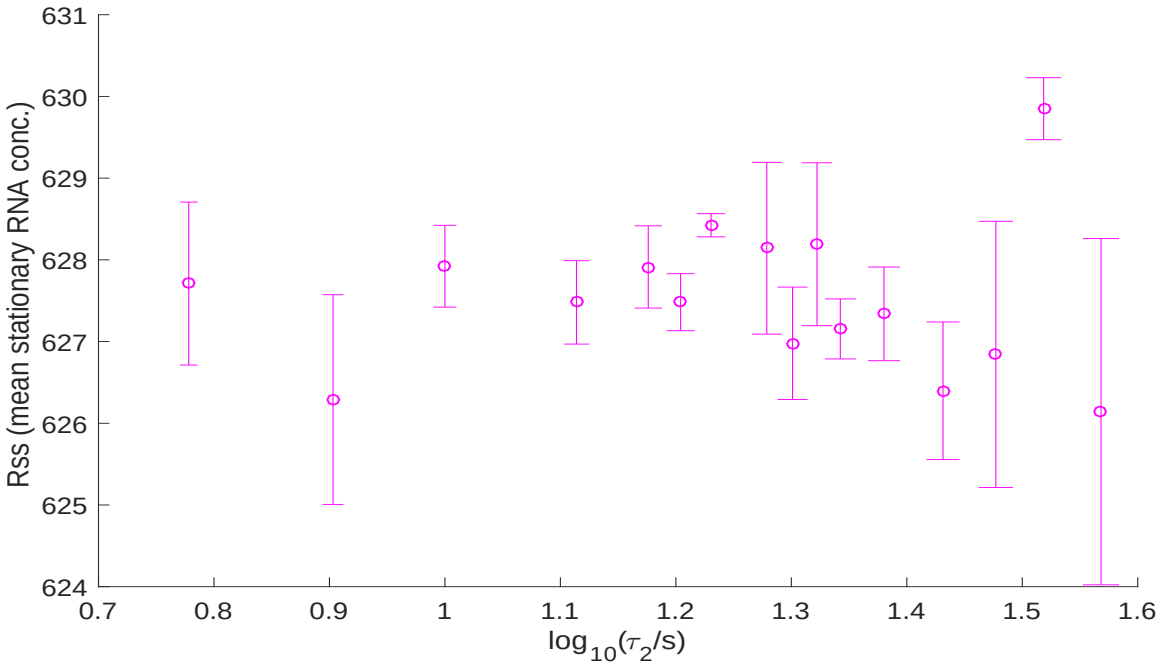

Figure 3.14: R_{ss} versus $\log_{10}\tau_2$. All parameters, except τ_2, are from Table 2.1.

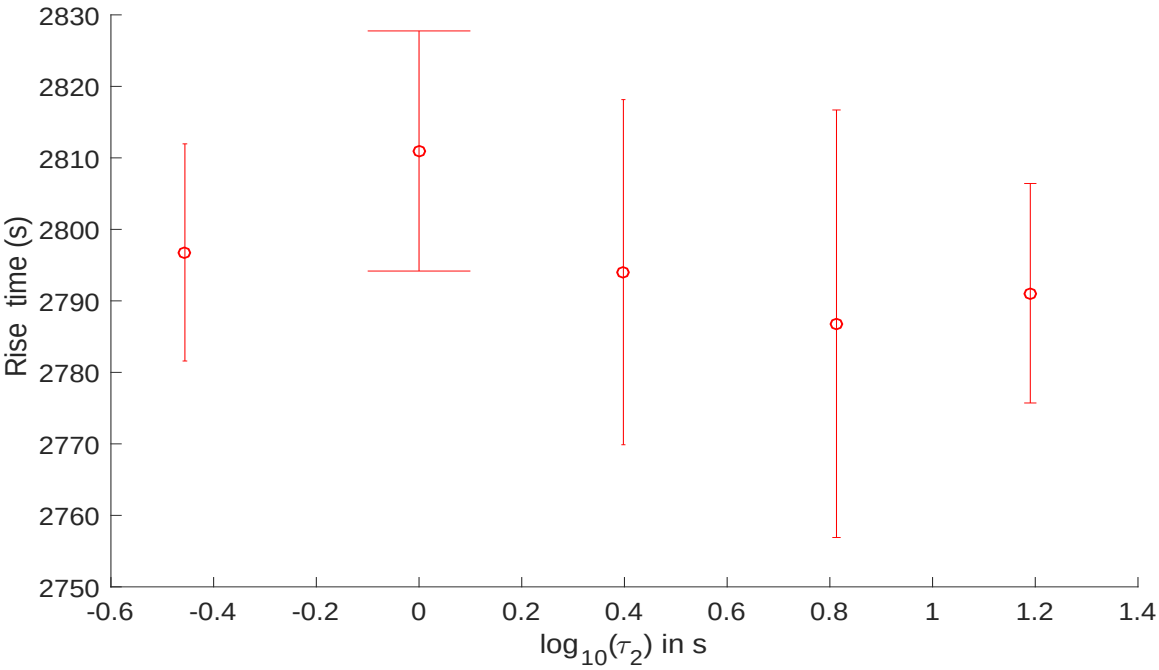

Figure 3.15: Rise time versus $\log_{10}\tau_2$. All parameters, except τ_2, are from Table 2.1.

range of τ_2 values plotted here is again decided on the basis of the physically reasonable range in which τ_2 could exist.

Figure 3.14 shows that, as expected, changing the value of τ_2 doesn't change the RNA number, and figure 3.15 shows that changing the value of τ_2 doesn't affect the value of rise time.

3.2 Delay stochastic simulation model with arrest

Transcription is a complex biological process in which the movement of RNA polymerase is not always continuous [50]. The simple RNA expression model described above is very effective in dealing with the transcription process where RNA polymerase arrest is not included [22]. It gave a clear insight into the dependence of bacterial transcription kinetics on several parameters. But in order to understand and simulate the process of bacterial transcription as it happens in the biological system where RNA polymerase undergoes abortive initiation during the elongation phase of transcription, the delay stochastic simulation model with arrest is studied here in detail. This model will help us to understand the effects of the abortive initiation process by RNA polymerase on the mean stationary RNA concentration (R_{ss}) and rise time.

The delay stochastic simulation model with RNA polymerase arrest that was described in section 2.3 was simulated using the delay stochastic simulation method [22].

3.2.1 Analysis of delay stochastic simulation model with arrest

In this section, I will discuss the simulation results generated for the model with abortive initiation in comparison to the results generated by the simple RNA expression model. In particular, the dependence of rise time and R_{ss} on several parameters used in the delay stochastic simulation model with abortive initiation is also analyzed. The simulation is done using a set of parameters that are given in table 2.3. These simulations were run 10,000 times to generate accurate statistics.

Figure 3.16 shows that the mean stationary number of RNA molecules produced with respect to time by simulating the simple RNA expression model (in red) is larger than that of the simulation of the delay stochastic simulation model with abortive initiation (in blue). This is because of the fact that, if the abortive initiation of an RNA polymerase happens during the early elongation process then the arrested RNA polymerase stays on the arrested site for a certain period of time, and then detaches from the template strand [51]. In other words, it is assumed that an arrested RNA polymerase never resumes its movement again on the DNA template strand. Hence, the mean stationary number of RNA molecules produced with abortive initiation is less in comparison to the number produced in the non-arrest case.

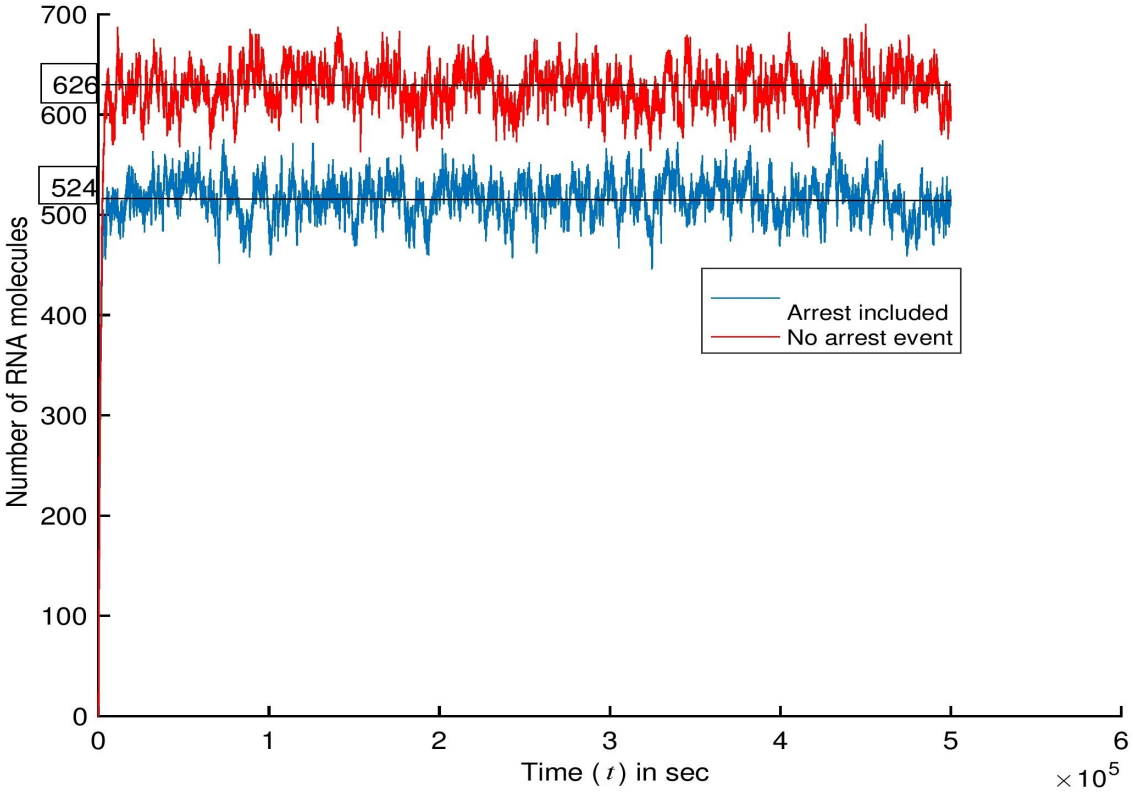

Figure 3.16: Curves showing the number of RNA molecules produced with respect to time by simulating the simple RNA expression model (in red) and by simulating the delay stochastic simulation model with abortive initiation (in blue). For the red curve, all parameters are from Table 2.1 and for the blue curve, all parameters are from Table 2.3.

In addition to the reasoning given above, I have decided to look deeper into the reason

Table 3.1: Tables showing 1) on the left, the values of k_2 and k_{arrest} for non-arrest and arrest cases corresponding to the red and blue curves from figure 3.17 respectively; 2) on the right, the values of k_2 and k_{arrest} for non-arrest and arrest cases corresponding to the yellow and green curves respectively.

	k_2/s^{-1}	$k_{\text{arrest}}/\text{s}^{-1}$
Non-arrest case	1	0
Arrest case	1	0.3

	k_2/s^{-1}	$k_{\text{arrest}}/\text{s}^{-1}$
Non-arrest case	100	0
Arrest case	100	0.3

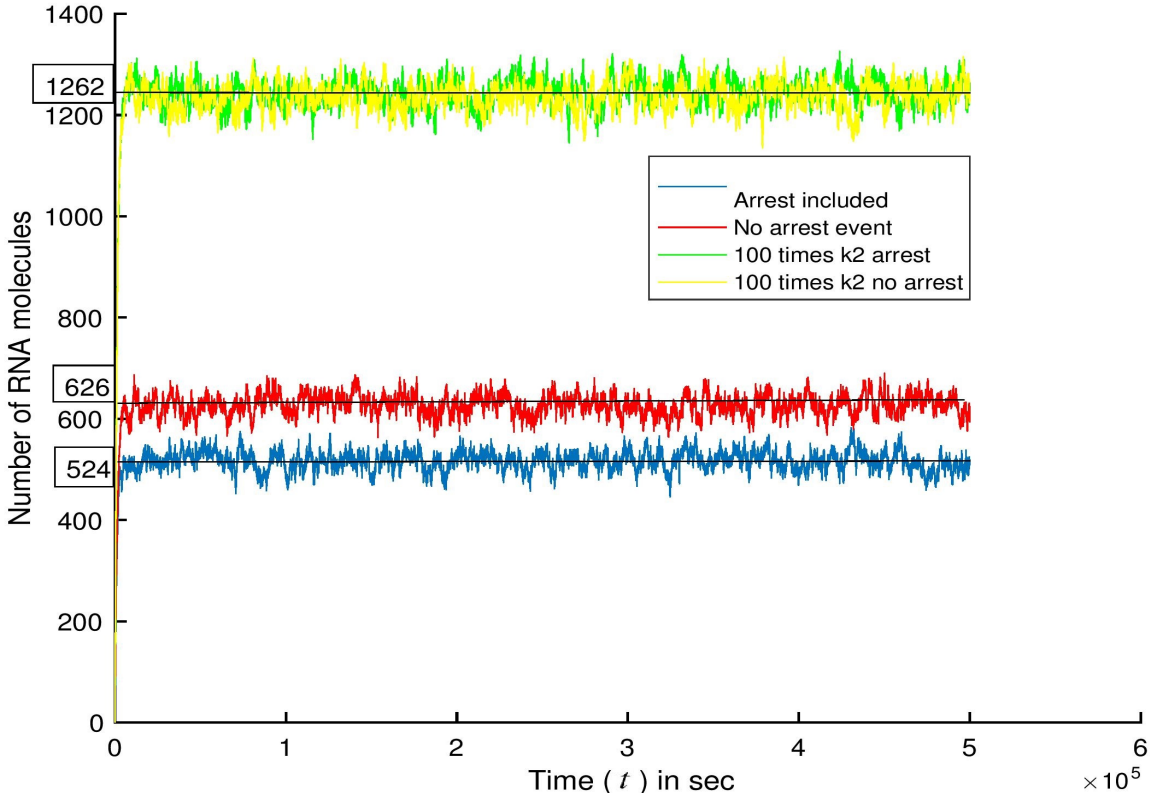

Figure 3.17: Curves showing the number of RNA molecules produced with respect to time by simulating the simple RNA model with the parameters of table 2.1 (in red), the model with abortive initiation with the parameters of table 2.3 (in blue), the simple RNA model with k_2 increased to 100 s^{-1} (in yellow), and the model with abortive initiation and k_2 increased to 100 s^{-1} (in green).

behind the curves depicted in figure 3.16. For that, I have increased the value of k_2 which is the stochastic rate constant for starting the elongation process in the non-arrest case as shown in equation (2.12). I have chosen k_2 specifically because we know that in the delay stochastic simulation model with arrest the reaction steps leading to arrest or productive elongation are competing with each other (as explained in section 2.3). In table 3.1, the

values of k_2 and k_{arrest} are given for arrest and non-arrest cases corresponding to the curves given in figure 3.17. In figure 3.17, the red curve indicates the non-arrest case and the blue curve indicates the RNA polymerase arrest case. These red and blue curves are the same as in figure 3.16. The curves in yellow and green colors indicate the cases when the value of the stochastic rate constant k_2 was multiplied by 100 in order to decrease the effect of RNA polymerase arrest by considerably decreasing the chances of arrest (by decreasing the arrest probability as derived from the formula given in equation (2.18)). The yellow curve corresponding to the simple RNA expression model and the green colored curve corresponding to the model with RNA polymerase arrest are overlapping each other. This overlapping shows that, if we will increase the value of the stochastic rate constant k_2 by 100 times then it will override the effects of RNA polymerase arrest by making the value of k_{arrest} much smaller than k_2, thereby decreasing the probability of abortive initiation.

3.2.2 Dependence of RNA levels and rise time on kinetic parameters

As explained above, the final RNA concentrations and the rise time corresponding to these concentrations depend on several kinetic parameters [47]. The parameters used in our model include k_x, k_2, k_{arrest}, k_6, k_7, k_8, τ_1, τ_2 and τ_3, where k_{arrest} is the stochastic rate constant for the reaction that represents abortive initiation and which is competing with the elongation reaction (having stochastic rate constant k_2) as shown in equation (2.12). τ_3 is the delay in the abortive initiation step from equation (2.17). Since all the parameters in this model are the same as given in the simple RNA expression model (studied in section 3.1) except k_{arrest} and τ_3, I have decided to study the dependency of R_{ss} and rise time on k_{arrest} and τ_3 here in this section.

Effect of varying k_{arrest} on mean stationary RNA level and rise time

In figure 3.18, we can see that increasing the value of k_{arrest} decreases the mean RNA concentration or R_{ss}. The reason behind this kind of curve is because of the fact that the arrest of RNA polymerase on the template strand during the transcription process doesn't

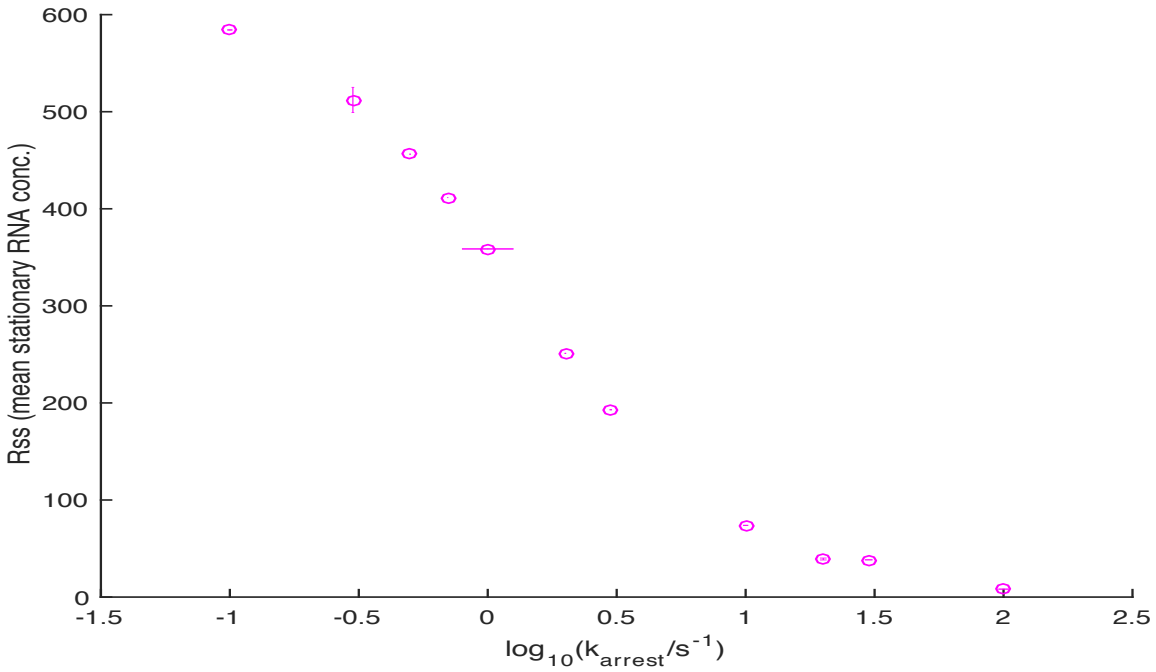

Figure 3.18: R_{ss} versus $\log_{10} k_{\text{arrest}}$. All parameters, except k_{arrest}, are from Table 2.3.

result in a successful RNA product [51]. k_{arrest} has the effect of increasing the time between successful productive initiation events which is analogous to the delay τ_1. Hence, the overall RNA concentration decreases.

I ran several simulations with the value of the arrest probability varying from 10 percent to 99 percent and analyzed its impact on rise time. Table 3.2 represents the value of arrest probability corresponding to the value of k_{arrest} as used in this simulation. The arrest percentage used in this graph is calculated by using the formula given for arrest probability in equation (2.18). Here, the value of k_2 is kept fixed and the value of probability of arrest is varied from 10 percent to 99 percent which gives rise to several different values of k_{arrest}. As shown in figure 3.19, the value of rise time decreases on increasing the value of k_{arrest}. This shape of the rise time curve relates to the R_{ss} curve in figure 3.18 in which the value of R_{ss} is decreasing on increasing the value of k_{arrest}. Therefore, if the stationary RNA level is less then the rise time associated with that stationary RNA level would also be correspondingly less.

Table 3.2: Table showing the value of arrest probability corresponding to the value of k_{arrest}

$k_{\text{arrest}}/\text{s}^{-1}$	Arrest probability
0.1	10%
0.3	25%
0.5	75%
0.7	40%
1	50%
2	65%
3	75%
10	90%
20	95%
30	97%
100	99%

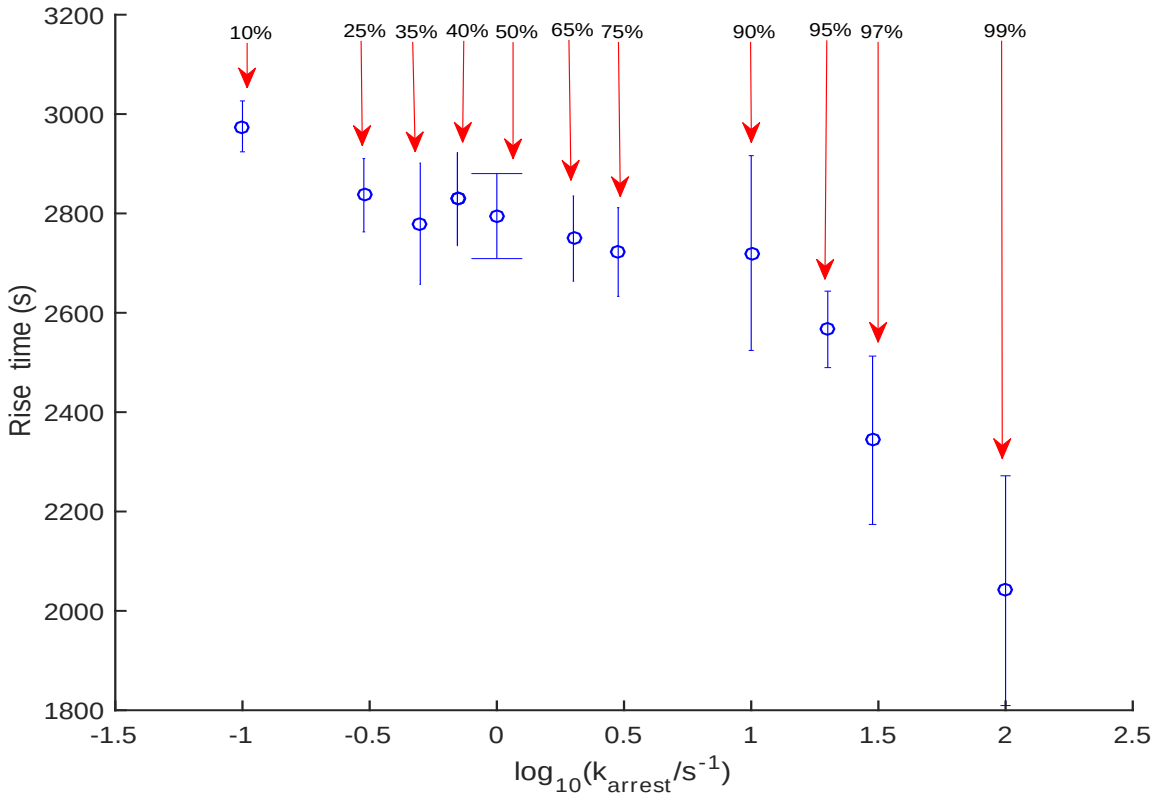

Figure 3.19: Rise time versus $\log_{10} k_{\text{arrest}}$. For every value of k_{arrest}, there is an associated arrest probability which varies from 10 percent to 99 percent. The values of the rise time which are plotted here with respect to k_{arrest} and their corresponding arrest probabilities are shown in table 3.2. All parameters, except k_{arrest}, are from Table 2.3.

As shown in section 3.2.2 above for the analysis of the simple RNA expression model, varying the values of stochastic rate constants k_2, k_6 and k_8 can have a significant impact on the rise time. Specifically, on varying the values of stochastic rate constants k_2 and k_8, the values of both R_{ss} and rise time change, whereas in the case of the stochastic rate constant k_6, the value of R_{ss} doesn't change significantly but the value of rise time shows a dramatic change as shown in figure 3.7. To understand the relationship between k_6 and rise time in the presence of abortive initiation, I have studied them in more detail as described in the subsection below.

Effect of varying k_6 and k_{arrest} on rise time

In our model, k_6 is the stochastic rate constant for the reaction process where the termination complex 1 is converted into the more stable state of termination complex 2 (as shown in equation (2.13)). So, in order to understand the combined effect of k_6 and k_{arrest} on rise time, a combined graph is generated (figure 3.20). In this graph, the value of k_6 is varied as 10, 20 and 100 s^{-1}. Note that the values of all other parameters were kept constant and were taken from table 2.3. The data presented here correspond to the wide range of values of k_{arrest} starting from the small values of the order of 10^{-5} s^{-1} all the way to 10 s^{-1}. The simulations were run for 10,000 times in order to generate these graphs.

On the vertical rise time axis of figure 3.20, we can see that as the value of k_6 increases from 10 s^{-1} (in blue) to 100 s^{-1} (in pink), the value of rise time decreases from about 2890 s to 2800 s for the smallest values of k_{arrest} I looked at, $k_{\text{arrest}} = 5 \times 10^{-6}\, s^{-1}$. Changes of a similar order of magnitude are observed at other values of k_{arrest}. This result is in agreement with the result obtained from figure 3.7 in the case of the simple RNA expression model and the reason behind this shape of the curve was already explained above. Also, we can see here that on increasing the value of k_6 by tenfold, the rise time only changes by about 3 percent, so the rise time is not very sensitive to this rate constant.

In addition to this, figure 3.20 shows here the variation of rise time with k_{arrest} is similar

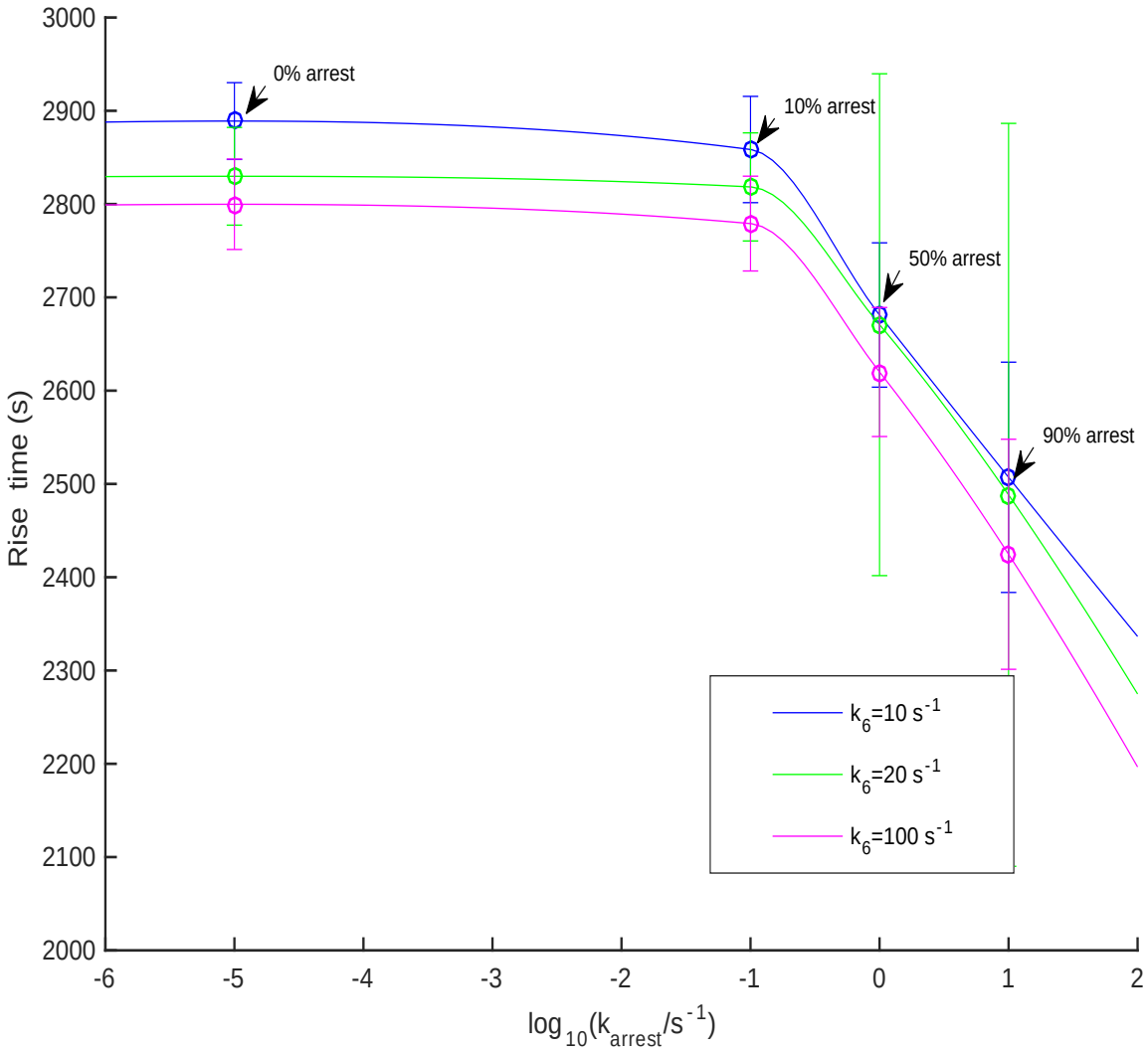

Figure 3.20: Rise time versus $\log_{10} k_{\text{arrest}}$ at several values of k_6. All parameters, except k_6 and k_{arrest}, are from Table 2.3.

regardless of the value of k_6.

Effect of varying τ_3 on mean stationary RNA level and rise time

From equation (2.17), the term τ_3 is assumed to be the sum of τ_{as} and τ_{arrest}. τ_{as} is the time taken by the RNA polymerase to reach the arrest site 'X' and τ_{arrest} is the time for which the RNA polymerase remains arrested on site 'X' before getting detached.

Figure 3.21 shows that increasing the value of τ_3 decreases the mean RNA number. However, figure 3.22 shows that changing the value of τ_3 doesn't affect the value of rise

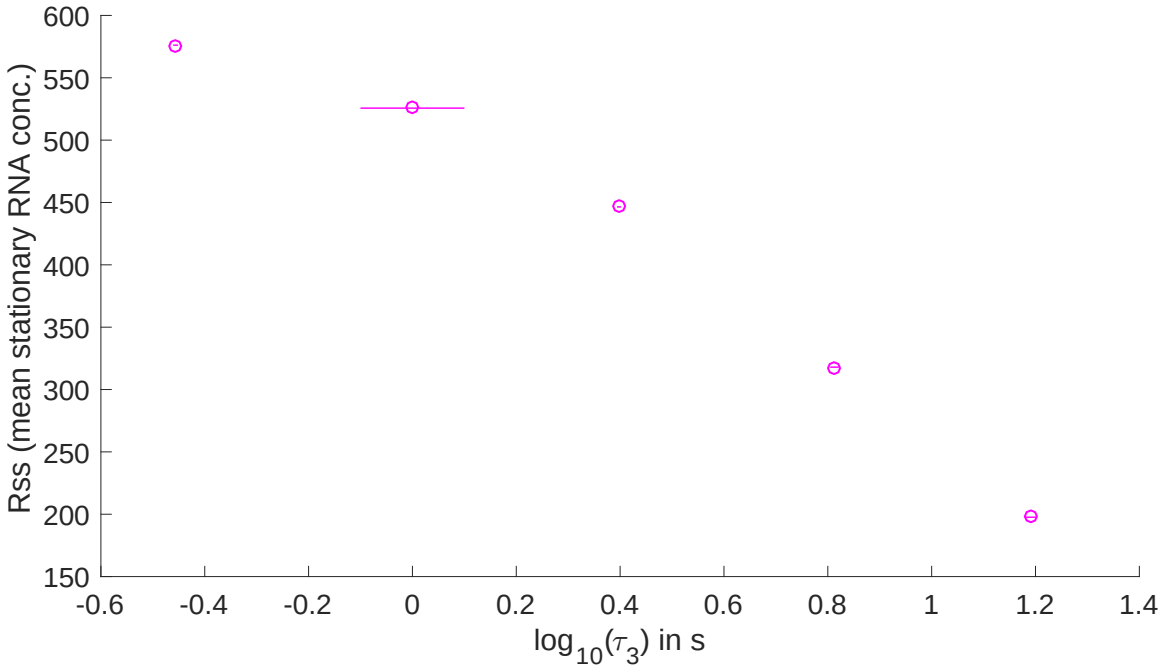

Figure 3.21: R_{ss} versus $\log_{10}\tau_3$. All parameters, except τ_3, are from Table 2.3.

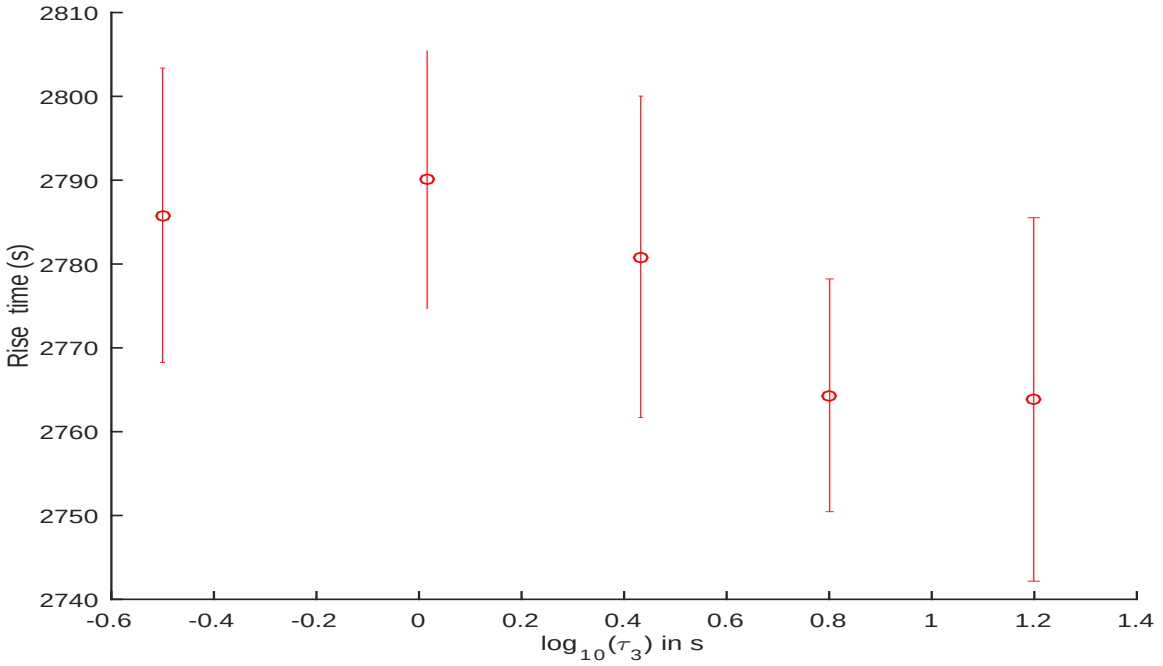

Figure 3.22: Rise time versus $\log_{10}\tau_3$. All parameters, except τ_3, are from Table 2.3.

time. In this case, as shown in the case of τ_1, the slower re-initiation frequency balances the decrease in R_{ss}. The range of the values chosen here is an assumption based on the physically reasonable range.

3.3 Summary

Summarizing this chapter, the results above give rise to two observations: first, in the case of the simple RNA expression model, R_{ss} depends on the stochastic rate constants k_x, k_2, k_8 and τ_1 whereas the rise time depends on k_2, k_6, k_7 and k_8; and second, in the case of the model with abortive initiation, in addition to the stochastic rate constants mentioned for the simple RNA expression model, both R_{ss} and rise time depend on k_{arrest} whereas only R_{ss} depends on the delay τ_3. It is important to note here that the rise time depends on k_7 but only at very small values when k_7 is rate limiting. Also, the reason for the independence of the rise time from k_x is due to the fact that k_x controls the initiation process along with many other factors. So, even if the initiation process has a large stochastic rate constant, it wouldn't impact the rise time because it depends on the re-initiation frequency. The re-initiation frequency depends on how fast the promoter is available for the next initiation process.

I wouldn't have predicted that the value of rise time is independent of the values of the delays τ_1 and τ_3 even when the value of R_{ss} changes as shown in figures 3.13 and 3.22 respectively. The probable reason behind this kind of curve would be due to the slower re-initiation frequency which balances the decrease in R_{ss}. In addition to this, it was also surprising to see that on increasing the value of k_6 by about ten-fold, the value of rise time doesn't change significantly as seen in figure 3.20. This wasn't predicted earlier because as shown in figure 3.7, the value of k_6 has a significant impact on rise time in the absence of RNA polymerase arrest.

Chapter 4

Mathematical analysis

The results obtained from chapter 3 were obtained from the delay stochastic simulation model with and without RNAP arrest. These simulation results are dependent on the values of parameters like stochastic rate constants and delays. To provide a general relationship between the variables and the parameters I have derived mathematical expressions for the mean stationary RNA number and rise time. These mathematical expressions are derived from a set of five differential equations taken from the six reaction steps in the transcription model with RNAP arrest.

The stochastic model can help us to estimate the statistics for different variables in a stochastic system and the delay-differential equations can help us to predict the mean behavior of the stochastic system [52]. In models without delays, there is a well-established theory that connects stochastic treatments with the conventional mass-action kinetics. The theory is based on a moment expansion, the lowest moment being the average which predicts the mass-action kinetics. That can be shown to reduce to the usual mass-action equations under some conditions [53]. The standard stochastic simulation methods sample from the master equation, so the average from stochastic simulations is predicted by the mass-action differential equations. So, by looking at the solution of these equations we can find a direct relationship between these differential equations and the stochastic simulation model. A similar relationship should hold for stochastic simulation models with delays [52, 53]. This chapter is a validation of results obtained through the stochastic simulation method for the model with abortive initiation as explained in chapter 3. In the following sections,

analytic calculations of the RNA number and rise time are presented.

The reactions (2.2) to (2.7) are the reaction steps for the model that includes abortive initiation. As before, the value of RNAP is considered as a fixed number that stays constant with time. Firstly, I have generated a set of five delay differential equations. These equations were analyzed in the steady state to get the mean stationary RNA number (R_{ss}). From the final value of R_{ss}, it is concluded that R_{ss} depends on all parameters except k_6, k_7 and τ_2. This is in agreement with the results we got from the computer simulations.

Secondly, we have to validate the dependence of rise time on the parameters. Here, rise time is defined to be the time taken to reach the ninety percent of the stationary RNA level [44, 45]. In order to do this, I have approximated the initial transient approach to the steady state. It is the same set of equations as used previously for R_{ss}, but simplified. These differential equations were then integrated to get an expression for rise time. From the expression we get, we can predict the dependence of rise time on the parameters. Here, I looked at specific cases where k_6 and k_8 were limiting, and derived equations for the rise time valid in these cases.

4.1 Analytic theory of mean stationary RNA number

Five differential equations were obtained from the six reaction steps above using the delayed mass-action law [54] for each variable including promoter (Pro), elongation complex (ElnC), termination complexes (C_1 and C_2) and RNA transcript (R).

$$\frac{\mathrm{d}\,\mathrm{Pro}}{\mathrm{d}t} = -k_x\mathrm{Pro}(t) + k_2\mathrm{ElnC}(t-\tau_1) + k_{\mathrm{arrest}}\mathrm{ElnC}(t-\tau_3) \tag{4.1}$$

$$\frac{\mathrm{d}\,\mathrm{ElnC}}{\mathrm{d}t} = k_x\mathrm{Pro}(t) - k_2\mathrm{ElnC}(t) - k_{\mathrm{arrest}}\mathrm{ElnC}(t) \tag{4.2}$$

$$\frac{\mathrm{d}\,\mathrm{C}_1}{\mathrm{d}t} = k_2\mathrm{ElnC}(t-\tau_2) - k_6\mathrm{C}_1(t) \tag{4.3}$$

$$\frac{\mathrm{d}\,\mathrm{C}_2}{\mathrm{d}t} = k_6\mathrm{C}_1(t) - k_7\mathrm{C}_2(t) \tag{4.4}$$

$$\frac{\mathrm{d}\mathrm{R}}{\mathrm{d}t} = k_7\mathrm{C}_2(t) - k_8\mathrm{R}(t) \tag{4.5}$$

The equations above represent the average values of the population of molecules present at any time during a transcription process. Under steady state conditions, the rates of change of promoter (Pro), elongation complex (ElnC), termination complexes (C_1 and C_2), and RNA transcript (R) with respect to time are zero. So, we set the right hand sides of the equations from (4.1) to (4.5) equal to zero. In addition, there are terms in these equations with the delays present in them and since the number of molecules at steady state (at time t) will be the same as the molecular population at any time delay (say at time (t - τ)), so we will write the equations from (4.1) to (4.5) without any time delay. The equations (4.1) and (4.2) are redundant in the steady state and this linear dependence between (4.1) and (4.2) implies the existence of a conservation law. An appropriately constructed conservation law will be explained shortly. The following equations give the steady-state conditions derived from equations (4.1) to (4.5):

$$0 = -k_x\mathrm{Pro} + k_2\mathrm{ElnC} + k_{\mathrm{arrest}}\mathrm{ElnC} \tag{4.6}$$

$$0 = k_2\mathrm{ElnC} - k_6\mathrm{C}_1 \tag{4.7}$$

$$0 = k_6\mathrm{C}_1 - k_7\mathrm{C}_2 \tag{4.8}$$

$$0 = k_7\mathrm{C}_2 - k_8\mathrm{R} \tag{4.9}$$

Adding equations (4.7) and (4.8), we get:

$$k_2\mathrm{ElnC} - k_7\mathrm{C}_2 = 0 \tag{4.10}$$

Adding equations (4.9) and (4.10), we get:

$$k_2\mathrm{ElnC} = k_8\mathrm{R} \tag{4.11}$$

The mass of a system remains conserved which means that the mass of the system before and after the reaction remains the same. In order to find the steady state, we have to take mass conservation into account. Consider the pipe model here again where once the elongation complex (ElnC) is in the pipe, there are only two ways by which the promoter could be recovered at the other end of the pipe: firstly, when the elongation happens without the arrest of RNA polymerase then promoter is released τ_1 time units later after the formation of elongation complex at time t; secondly, when the arrest of RNA polymerase happens then it leads to an unsuccessful elongation step where the promoter comes out after τ_3 time units later after the formation of elongation complex at time t. The mass conservation equation for the promoter in the process of transcription can be written as:

$$\mathrm{Pro}_{\mathrm{total}} = \mathrm{Pro}(t) + \mathrm{ElnC}(t) + \int_t^{t+\tau_1} k_2 \mathrm{ElnC}(t'-\tau_1)\mathrm{d}t' + \int_t^{t+\tau_3} k_{\mathrm{arrest}} \mathrm{ElnC}(t'-\tau_3)\mathrm{d}t' \tag{4.12}$$

We can prove that the number of promoters is constant by differentiating the right-hand side and showing that this derivative is zero as given below:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Pro}_{\mathrm{total}} = \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Pro}(t) + \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{ElnC}(t) + \frac{\mathrm{d}}{\mathrm{d}t}\int_t^{t+\tau_1} k_2 \mathrm{ElnC}(t'-\tau_1)\mathrm{d}t' + \frac{\mathrm{d}}{\mathrm{d}t}\int_t^{t+\tau_3} k_{\mathrm{arrest}} \mathrm{ElnC}(t'-\tau_3)\mathrm{d}t' \tag{4.13}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Pro}_{\mathrm{total}} = \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Pro}(t) + \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{ElnC}(t) + \frac{\mathrm{d}}{\mathrm{d}t}\int_t^{a} k_2 \mathrm{ElnC}(t'-\tau_1)\mathrm{d}t' + \frac{\mathrm{d}}{\mathrm{d}t}\int_a^{t+\tau_1} k_2 \mathrm{ElnC}(t'-\tau_1)\mathrm{d}t' + \frac{\mathrm{d}}{\mathrm{d}t}\int_t^{a} k_{\mathrm{arrest}} \mathrm{ElnC}(t'-\tau_3)\mathrm{d}t' + \frac{\mathrm{d}}{\mathrm{d}t}\int_a^{t+\tau_3} k_{\mathrm{arrest}} \mathrm{ElnC}(t'-\tau_3)\mathrm{d}t' \tag{4.14}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Pro}_{\mathrm{total}} = \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Pro}(t) + \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{ElnC}(t) - \frac{\mathrm{d}}{\mathrm{d}t}\int_a^{t} k_2 \mathrm{ElnC}(t'-\tau_1)\mathrm{d}t' + \frac{\mathrm{d}}{\mathrm{d}t}\int_a^{t+\tau_1} k_2 \mathrm{ElnC}(t'-\tau_1)\mathrm{d}t' - \frac{\mathrm{d}}{\mathrm{d}t}\int_a^{t} k_{\mathrm{arrest}} \mathrm{ElnC}(t'-\tau_3)\mathrm{d}t' + \frac{\mathrm{d}}{\mathrm{d}t}\int_a^{t+\tau_3} k_{\mathrm{arrest}} \mathrm{ElnC}(t'-\tau_3)\mathrm{d}t' \tag{4.15}$$

Putting the value of $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Pro}(t)$ and $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{ElnC}(t)$ from equation (4.1) and (4.2), respectively, into equation (4.15), and using the fundamental theorem of calculus, we get

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Pro}_{\mathrm{total}} = &-k_x\mathrm{Pro}(t) + k_2\mathrm{ElnC}(t-\tau_1) + k_{\mathrm{arrest}}\mathrm{ElnC}(t-\tau_3) + k_x\mathrm{Pro}(t) - k_2\mathrm{ElnC}(t) \\ &- k_{\mathrm{arrest}}\mathrm{ElnC}(t) - k_2\mathrm{ElnC}(t-\tau_1) + k_2\mathrm{ElnC}(t) \\ &- k_{\mathrm{arrest}}\mathrm{ElnC}(t-\tau_3) + k_{\mathrm{arrest}}\mathrm{ElnC}(t) \quad (4.16)\end{aligned}$$

Simplifying equation (4.16), we get

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Pro}_{\mathrm{total}} = 0} \quad (4.17)$$

Hence, we can say that the total number of promoters are constant at all values of t. This is to be noted that equation (4.13) is generally true during the time evolution of the system.

Again from equation (4.13), we can say that in the stationary state, the population of elongation complex at time (t' - τ_1) and at (t' - τ_3) will be the same as at the stationary state population. Evaluating the integrals, we get:

$$\mathrm{Pro}_{\mathrm{total}} = \mathrm{Pro} + \mathrm{ElnC} + \tau_1 k_2 \mathrm{ElnC} + \tau_3 k_{\mathrm{arrest}}\mathrm{ElnC} \quad (4.18)$$

Note that equation (4.18) applies only in the stationary state. Also, from equation (4.6) we know that,

$$\mathrm{Pro} = \frac{(k_{\mathrm{arrest}} + k_2)\mathrm{ElnC}}{k_x} \quad (4.19)$$

Putting the value of Pro from equation (4.19) in equation (4.18), we get:

$$\mathrm{Pro}_{\mathrm{total}} = (k_{\mathrm{arrest}} + k_2)\mathrm{ElnC}/k_x + \mathrm{ElnC} + \tau_1 k_2 \mathrm{ElnC} + \tau_3 k_{\mathrm{arrest}}\mathrm{ElnC} \quad (4.20)$$

$$\mathrm{Pro}_{\mathrm{total}} = \frac{\mathrm{ElnC}((k_{\mathrm{arrest}} + k_2) + k_x + \tau_1 k_x k_2 + \tau_3 k_x k_{\mathrm{arrest}})}{k_x}$$

From equation (4.11), k_2 ElnC = k_8 R:

$$\mathrm{ElnC} = \frac{k_8 \mathrm{R}}{k_2}$$

$$\frac{k_8 \mathrm{R}}{k_2} = \frac{k_x \mathrm{Pro}_{\mathrm{total}}}{(k_{\mathrm{arrest}} + k_2) + k_x + \tau_1 k_x k_2 + \tau_3 k_x k_{\mathrm{arrest}}}$$

$$\boxed{\mathrm{R}_{\mathrm{ss}} = \frac{k_x k_2 \mathrm{Pro}_{\mathrm{total}}}{k_8[(k_{\mathrm{arrest}} + k_2) + k_x + \tau_1 k_x k_2 + \tau_3 k_x k_{\mathrm{arrest}}]}} \tag{4.21}$$

So from equation (4.21), we can see that the steady state value of R depends on certain parameters which include k_x, k_2, k_8, k_{arrest}, $\mathrm{Pro}_{\mathrm{total}}$, τ_1 and τ_3. I have added the subscript 'ss' on the left hand side in equation (4.21) to indicate a steady-state concentration. It makes sense that R_{ss} depends on all the delays except τ_2 because if we think in terms of the pipe metaphor, a longer delay corresponds to a longer pipe. If we have a longer pipe at some given initiation rate, we can have more transcription complexes in the pipe, but in the steady state, they come out at exactly the same rate as they go in, regardless of the length of the pipe. So, it doesn't directly impact the RNA number in the transcription process overall. Hence, changing the value of τ_2 will not affect the value of R_{ss}. Also, R_{ss} depends on all the stochastic rate constants except k_6 and k_7, the probable reason behind this is the steady-state level in these systems is set by a balance of the production and destruction terms. All the stochastic rate constants present in equation (4.21) are either involved in the initiation of production or in destruction. The stochastic rate constants k_6 and k_7 are in the middle so the first-order rates with which they are associated just increase as the initiation rate increases.

I have generated analytical results for the parameters on which R_{ss} depends in order to see how well the data fit the theory. The analytical results were generated using the same parametric values as used in the simulation of these models which are given in table 2.3 and then substituted in equation (4.21). The analytical results obtained are in agreement with the simulation results as shown in figures from 4.1 to 4.5. It can be seen that the simulation results are systematically a bit higher than the analytic results in all the figures from 4.1 to

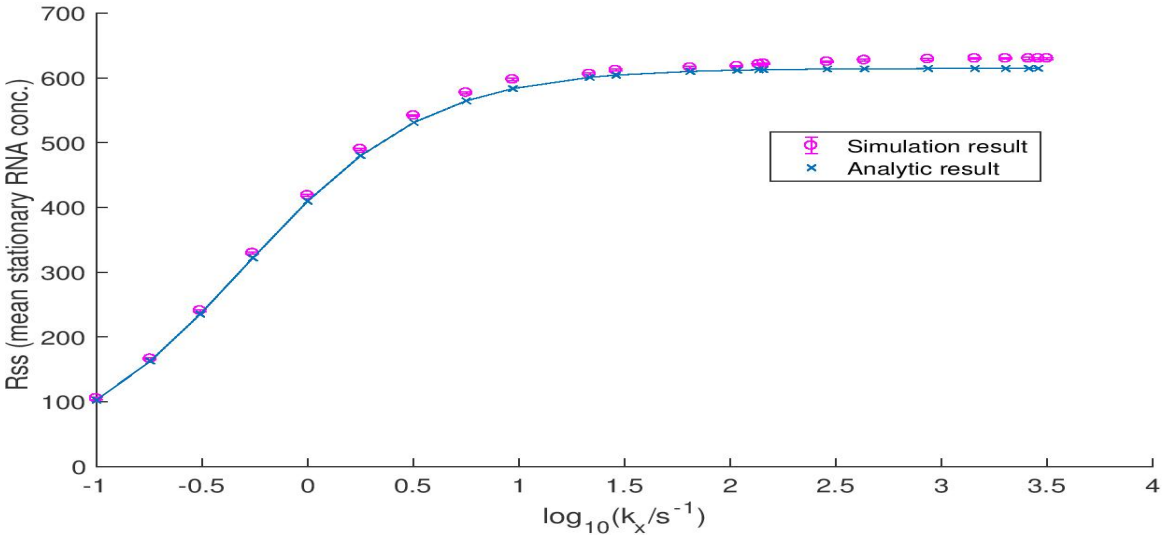

Figure 4.1: Figure showing the analytical curve with the simulation results for R_{ss} versus $\log_{10} k_x$. The simulation results are carried forward from figure 3.2. This makes these analytical results consistent with the simulations, which were subject to the initial conditions given in table 2.2. The above results were obtained for the simple RNA expression model.

4.5. This systematic bias is to be investigated at a later date.

4.2 Analytic theory of rise time

Some variables equilibrate much faster than the timescale for the slow rise in RNA levels. We can therefore treat those variables as being in equilibrium, which then allows us to solve a simple differential equation that we can use further to validate the simulation results for rise time. The analytic treatment of the rise time requires different approaches in different parameter ranges, and I am going to treat a few cases below to illustrate the theory and to validate selected simulation results.

4.2.1 At small values of k_8

In our model, variable C_2 equilibrates much faster than the timescale for the slow rise in RNA levels. This is so with the default parameters from table 2.1. We can therefore treat C_2 as being in equilibrium. So, we have assumed that all of the complexes up to and including C_2 equilibrate quickly. C_2 in the equation below represents the steady-state value. Hence

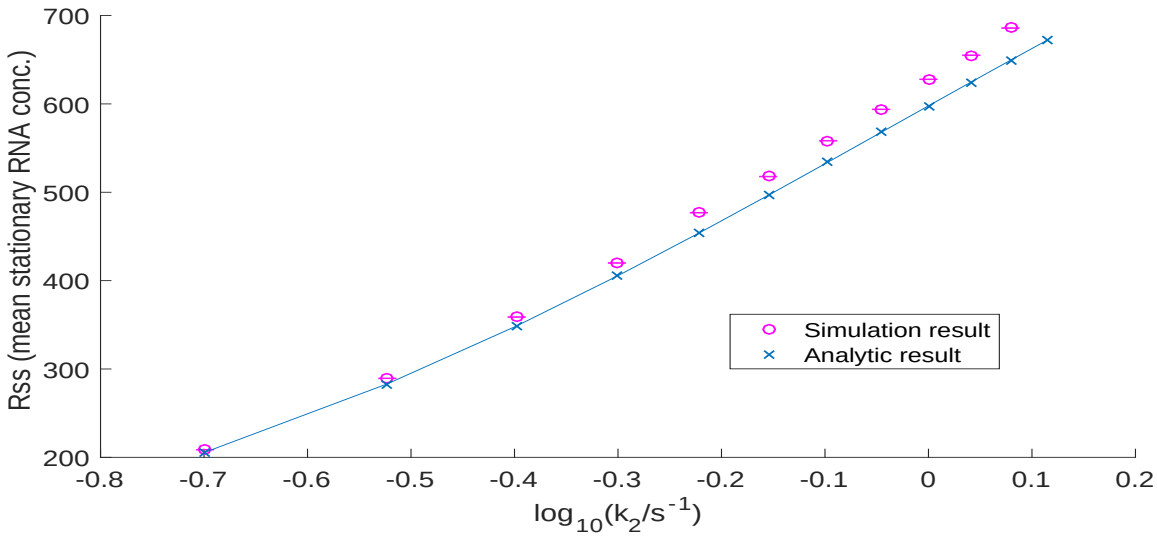

Figure 4.2: Figure showing the analytical curve with the simulation results for R_{ss} versus $\log_{10} k_2$. The simulation results are carried forward from figure 3.4. This makes these analytical results consistent with the simulations, which were subject to the initial conditions given in table 2.2. The above results were obtained for the simple RNA expression model.

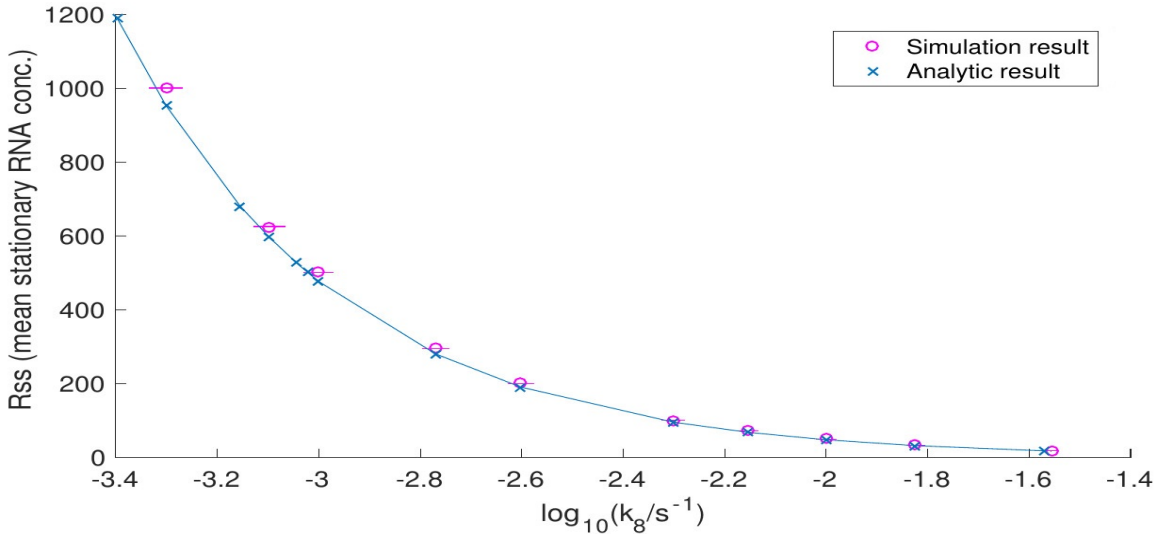

Figure 4.3: Figure showing the analytical curve with the simulation results for R_{ss} versus $\log_{10} k_8$. The simulation results are carried forward from figure 3.10. This makes these analytical results consistent with the simulations, which were subject to the initial conditions given in table 2.2. The above results were obtained for the simple RNA expression model.

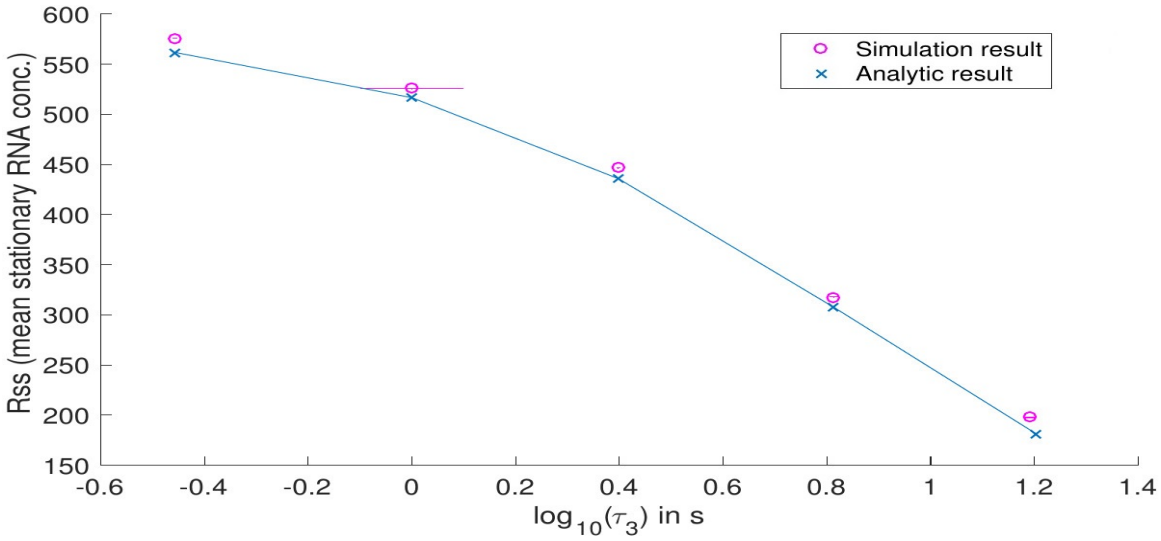

Figure 4.4: Figure showing the analytical curve with the simulation results for R_{ss} versus $\log_{10}\tau_3$. The simulation results are carried forward from figure 3.21. This makes these analytical results consistent with the simulations, which were subject to the initial conditions given in table 2.3. The above results were obtained for the model with abortive initiation.

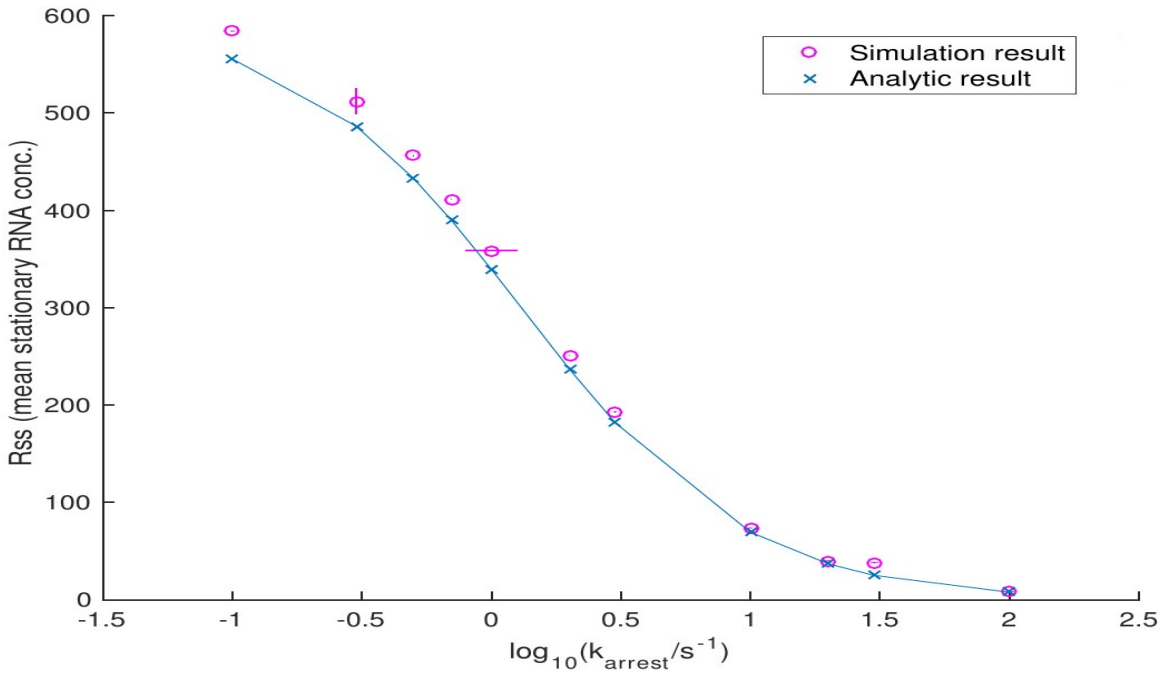

Figure 4.5: Figure showing the analytical curve with the simulation results for R_{ss} versus $\log_{10} k_{\text{arrest}}$. The simulation results are carried forward from figure 3.18. This makes these analytical results consistent with the simulations, which were subject to the initial conditions given in table 2.3. The above results were obtained for the model with abortive initiation.

from equation (4.5),

$$\int_0^{fR_{ss}} \frac{\mathrm{dR}}{k_7\mathrm{C}_2 - k_8\mathrm{R}} = \int_0^{t_{\mathrm{rise}}} \mathrm{d}t \tag{4.22}$$

where

$f = 0.9$, the fraction of the steady state level used to define the rise time [44, 45];

t_{rise} = time taken to reach the 90% level.

By using the technique of integration by substitution, we will solve this further. Let

$$u = k_7\mathrm{C}_2 - k_8\mathrm{R} \tag{4.23}$$

$$\mathrm{d}u = -k_8\ \mathrm{dR}$$

$$\mathrm{dR} = \frac{(-1)\mathrm{d}u}{k_8}$$

Putting the value of dR in equation (4.22), we get:

$$\int_0^{fR_{ss}} \frac{-1}{k_8}\frac{1}{u}du = \int_0^{t_{Rise}} \mathrm{d}t$$

$$\frac{-1}{k_8}\ln u\Big|_0^{fR_{ss}} = t\Big|_0^{t_{Rise}}$$

Replacing the value of u from equation (4.23), we get:

$$\frac{-1}{k_8}\ln\left(\frac{k_7\mathrm{C}_2 - k_8 fR_{ss}}{k_7\mathrm{C}_2}\right) = t_{Rise} \tag{4.24}$$

Combining equation (4.10) and equation (4.11) and then putting the values in equation (4.24), we get:

$$\boxed{\frac{1}{k_8}\ln\left(\frac{1}{1-f}\right) = t_{Rise}} \tag{4.25}$$

A graph of the relationship between rise time and $\log_{10} k_8$ for simulation versus analyti-

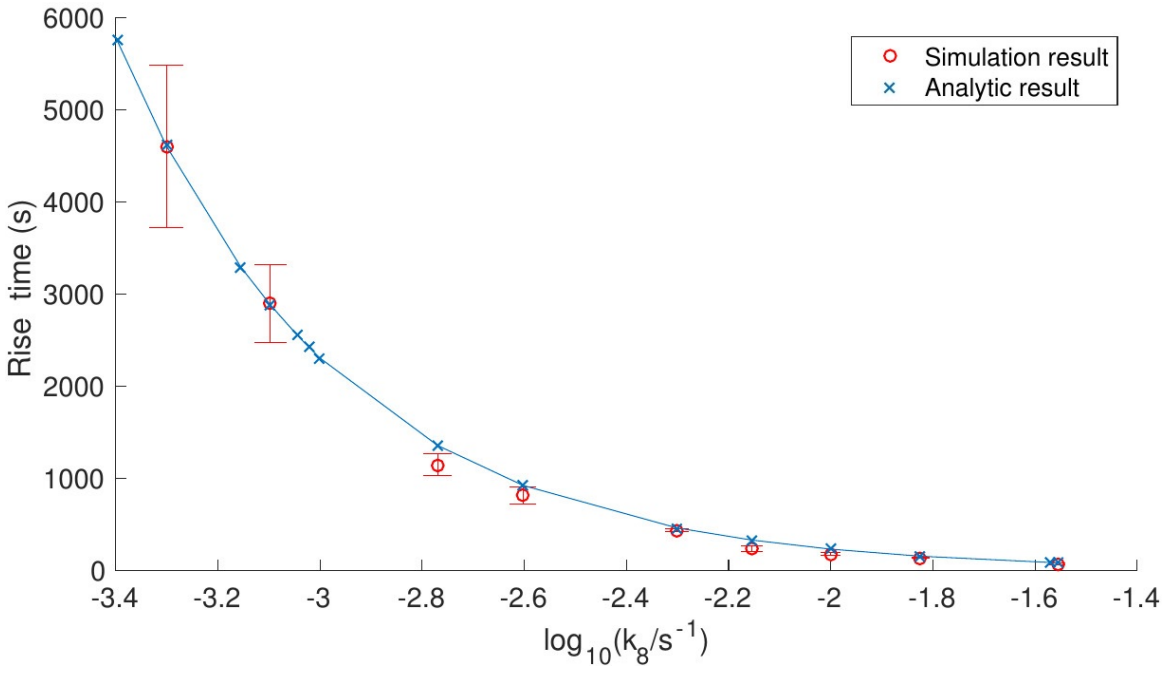

Figure 4.6: Figure showing the analytical curve with the simulation results for the rise time versus $\log_{10} k_8$. The simulation results are carried forward from figure 3.11. The analytical result is from equation (4.25). The above results were obtained for the model with abortive initiation.

cal results is shown in figure 4.6. Since the value of rise time varies so much, the simulated results are plotted with the error bars to show the variability of the data and how widely the rise time data are distributed. The results obtained are in agreement with the simulation results as shown in figure 4.6 except when k_8 is large enough, which validates our results quantitatively. It can be seen that the analytical results deviate from the theoretical curve once k_8 is no longer rate-limiting.

The relationship (4.25) is valid provided all of the complexes equilibrate much faster than the rise time, which is then strictly controlled by k_8. I am able to compare my numerical results, at least in cases in which k_8 is not too large, to this theoretical curve. Therefore, I considered one other parameter to check the dependence of rise time when the value of k_8 is large. For example, when k_6 is small, the complexes before C_1 are equilibrated, and also the complexes after C_1 rapidly adjust to (i.e. equilibrate with) the time-dependent value of C_1.

4.2.2 At small values of k_6

I have assumed that ElnC is in a steady state throughout much of the time required for R to reach the steady state. On solving equation (4.3) for C_1 to validate the relationship of rise time, we get-

$$\frac{dC_1}{dt} = k_2 ElnC - k_6 C_1 \tag{4.26}$$

$$\int_0^{C_{1_{Rise}}} \frac{dC_1}{k_2 ElnC - k_6 C_1} = \int_0^{t_{Rise}} dt \tag{4.27}$$

where,

$C_{1_{Rise}}$ = number of molecules of C_1 until it reaches the value corresponding to 90% of the steady-state value of R. Let

$$u = k_2 ElnC - k_6 C_1 \tag{4.28}$$

$$dC_1 = \frac{-1}{k_6} du$$

Putting the value of dC_1 in equation 4.27

$$\int_0^{C_{1_{Rise}}} \frac{-1}{k_6} \frac{1}{u} du = \int_0^{t_{Rise}} dt$$

$$\frac{-1}{k_6} \ln u \Big|_0^{C_{1_{Rise}}} = t \Big|_0^{t_{Rise}}$$

Replace the value of u from equation (4.28):

$$\frac{-1}{k_6} [\ln(\frac{k_2 ElnC_{ss} - k_6 C_{1_{Rise}}}{k_2 ElnC_{ss}})] = t_{Rise}$$

As we know, from equations (1.8) and (1.9):

$$k_6 C_1 = k_8 R$$

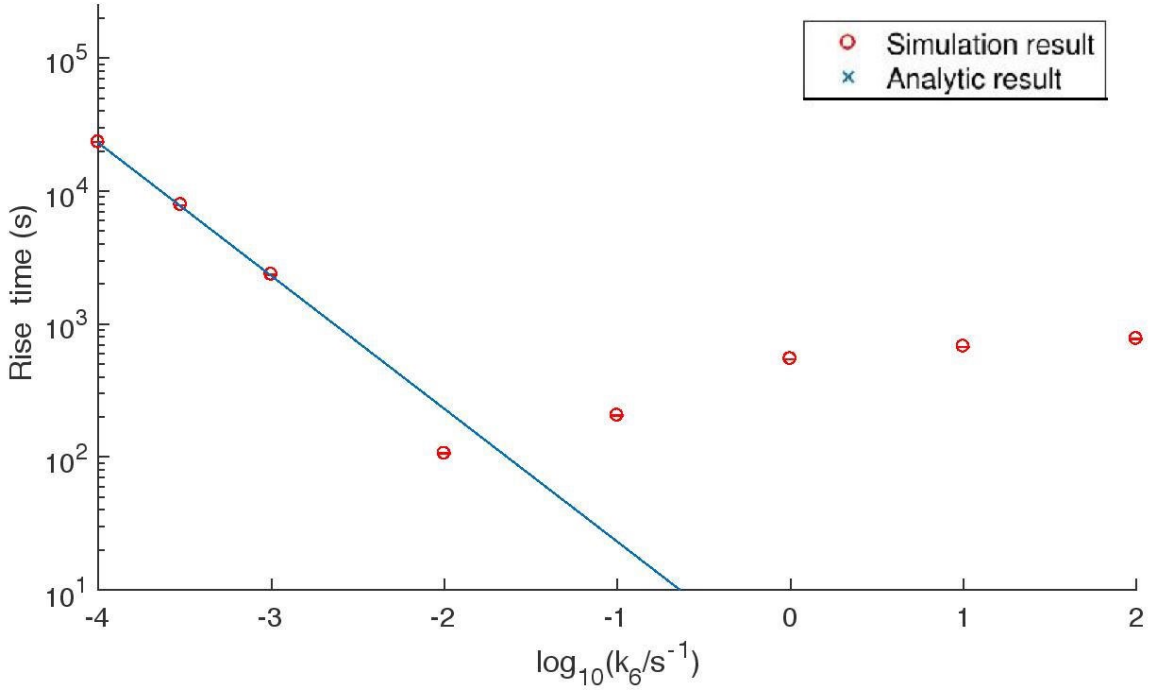

Figure 4.7: Figure showing the analytical curve with the simulation results for the rise time versus $\log_{10} k_6$. The simulation results are carried forward from figure 3.7. The analytical result is from equation (4.30). The above results were obtained for the model with abortive initiation.

As we are considering here the time limits equal to rise time, so the RNA population would be the ninety percent of the stationary RNA level which is fR_{ss}. So,

$$k_6 \mathrm{C}1_{Rise} = k_8 f R_{ss}$$

$$\boxed{\frac{-1}{k_6} \ln\left(\frac{k_2 \mathrm{ElnC}_{ss} - k_8 f R_{ss}}{k_2 \mathrm{ElnC}_{ss}}\right) = t_{Rise}} \tag{4.29}$$

By putting the value of $k_2\mathrm{ElnC}_{ss}$ from equation (4.11), we get:

$$\boxed{\frac{1}{k_6} \ln\left(\frac{1}{1-f}\right) = t_{Rise}} \tag{4.30}$$

So from equation (4.30), we can see that the rise time depends on f, which is the fraction of the steady state level used to define the rise time, and on k_6.

A graph of the relationship between rise time and $\log_{10} k_6$ for simulation versus analyti-

cal results is shown in figure 4.7. The results obtained are in agreement with the simulation results as shown in figure 4.7 except when k_6 is large enough, which validates our results quantitatively. It can be seen that the analytical results deviate from the theoretical curve once k_6 is no longer rate-limiting.

The mathematical expression obtained in this chapter for mean stationary RNA number provides a general formula for calculating this quantity but the equations for rise time are special cases. Other special cases could be derived as needed. These special cases include the consideration of cases when the stochastic rate constants k_x or k_2 are rate limiting.

Chapter 5

Conclusions

5.1 Summary

In this thesis, firstly a detailed study of a simple RNA expression model was done in order to study the effects of the parameters on the mean stationary RNA level and on the rise time. Stochastic simulations were used to do an in-depth study of this model by taking steps like initiation, elongation, termination and RNA degradation into consideration. The simulation results obtained are in full agreement with analytic results from chapter 4.

Furthermore, a delayed mass-action model with abortive initiation has been developed. This is an extension of the simple model. I carried out a simulation study to determine the effects of abortive initiation on the mean stationary RNA level and rise time. This study was done using the delay stochastic simulation algorithm as introduced by Roussel and Zhu [22]. On adding abortive initiation, it was observed that the mean stationary RNA level and rise time decrease as we increase the arrest probability of RNA polymerase.

I have developed an analytic theory of rise time in gene expression based on the delayed mass-action formalism and, to my knowledge, this is a first of its kind. There is a well-developed theory of mean first-passage times in stochastic kinetics [55, 56], but the theory that I have developed is much simpler.

In this chapter, I will discuss the conclusions made from the research work that I have presented in my thesis in relation to the results obtained from the stochastic simulations done for both the models, and the future perspectives to take this research work to the next level.

5.2 Conclusions

In the models studied, the number of RNA molecules initially increases sharply and then attains a stationary state. This stationary state happens when the rates of RNA formation and degradation become equal on average. The timescale for the rise is found to be in the range of tens of minutes to a couple of hours. Furthermore, when the dependence of RNA levels and rise time on several kinetic parameters was determined, it was seen that the mean stationary RNA level depends on the stochastic rate constants k_1, k_2, k_8 and delays τ_1 whereas the rise time depends on the stochastic rate constants k_2, k_6 and k_8 only.

In the process of transcription, it is assumed that the movement of RNA polymerase is not always continuous and it gets arrested while traversing the template DNA strand [50]. I have extended the model given by Roussel and Zhu [22] by adding an extra step of abortive initiation. I have assumed the arrest of RNA polymerase at some of the early sites on the template strand even before the promoter is cleared. The arrest process is competing with the elongation process in this model. Accordingly, increasing any one of the stochastic rate constants either for the arrest process or the elongation process could override the effects of the other. Furthermore, the dependence of RNA levels and rise time on k_{arrest} and delay τ_3 was determined. It was observed that on increasing the value of k_{arrest}, the value of both R_{ss} and rise time decreases. In the model, the abortive initiation of an RNA polymerase happens during the early elongation process. The arrested RNA polymerase stays on the arrested site for a certain period of time, and then detaches from the template strand [51]. In other words, it is assumed that an arrested RNA polymerase never resumes its movement again on the DNA template strand. Hence, the mean stationary number of RNA molecules produced with abortive initiation is reduced. Therefore, if the mean stationary number of RNA molecules produced is less, then the time taken to achieve that level would be less. Hence, the rise time decreases as we increase the value of k_{arrest}. However, it was observed that the mean stationary RNA level decreases on increasing the value of τ_3 whereas the rise time is independent of the value of τ_3. The reason behind this kind of the curve is that the

slower re-initiation frequency balances the decrease in R_{ss}.

Finally, the results generated from the simulation of these models were then analyzed using differential equations in chapter 4. I have illustrated how to determine rise times using a couple of particular cases, and other cases could be worked out as needed. The results generated by the simulations of these two models were validated using the mathematical analysis. I didn't attempt to analyze what happens if initiation events are rate-limiting. For example, the graphs showing the dependence of rise time on k_{arrest} and τ_3 could not be explained using the mathematical analysis. It's harder because I have to solve a delay-differential equation to get the rise time. According to the mathematical analysis done in chapter 4, it was concluded that the rise time depends on the stochastic rate constants k_2, k_6 and k_8 only, whereas the simulation results shows the dependence of rise time on k_{arrest} and τ_3 as well. This dependence of rise time on k_{arrest} and τ_3 makes sense because they control the time between two consecutive initiations.

5.3 Future perspectives

This study of abortive initiation using a delay stochastic simulation model is just the beginning. There is still a lot more to work on over here. The immediate step after this work would be to study the reason behind why the simulation results are systematically a bit higher than the analytic results in the analytic theory of mean stationary RNA number given in chapter 4. This systematic bias needs to be investigated in detail by doing a study of the parameters which can give rise to this difference in the simulation and analytical result. Also, in the analytical theory of rise time, I have not looked at cases where the rate-limiting step occurs before the delays; this is something that we can do in the future. In addition to this, I would like to consider the traffic issues on the template DNA strand in which, if a polymerase pauses or arrests somewhere in the middle of a gene, it can hold up the polymerases following it. We would have to look at different consequences of having several polymerases on the template strand and how they will react to an arrested

polymerase in front of them. After that, we can also think about adding the reactions like backtracking and pausing to a delay-stochastic model and analyze how easy or not it would be to modify the model accordingly. Once we have an efficient simulation method developed including these reactions, we can characterize the response of a model with arrest of RNA polymerase to the changes in parameters. Finally, since we have been working on a prokaryotic model organism, in future we can think of modifying the model in such a way that the transcription and translation will be simulated simultaneously. So, we can study the effects of varying several kinetic parameters for the transcription process and analyzing their impacts on the translation process. This is a very wide field and the scope of advancement is unlimited.

Bibliography

[1] Geoffrey M Cooper and Robert E Hausman. *The Cell*. Sinauer Associates Sunderland, 2000.

[2] J. Berger. A comment on information theory and central dogma of molecular biology. *Journal of Theoretical Biology*, 65(2):393–5, 1977.

[3] E. Nudler, A. Mustaev, E. Lukhtanov, and A. Goldfarb. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell*, 89(1):33–41, 1997.

[4] Suzanne Clancy. RNA transcription by RNA polymerase: prokaryotes vs eukaryotes. *Nature Education*, 1(1):125, 2008.

[5] Emmanuel Margeat, Achillefs N. Kapanidis, Philip Tinnefeld, You Wang, Jayanta Mukhopadhyay, Richard H. Ebright, and Shimon Weiss. Direct observation of abortive initiation and promoter escape within single immobilized transcription complexes. *Biophysical Journal*, 90(4):1419–1431, 2006.

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

[7] Tomoko Kubori and Nobuo Shimamoto. A branched pathway in the early stage of transcription by *Escherichia coli* RNA Polymerase. *Journal of Molecular Biology*, 256(3):449–457, 1996.

[8] Achillefs N Kapanidis, Emmanuel Margeat, Sam On Ho, Ekaterine Kortkhonjia, Shimon Weiss, and Richard H Ebright. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*, 314(5802):1144–1147, 2006.

[9] Hubert P Yockey. An application of information theory to the central dogma and the sequence hypothesis. *Journal of Theoretical Biology*, 46(2):369–406, 1974.

[10] M Sofia Ciampi. Rho-dependent terminators and transcription termination. *Microbiology*, 152(9):2515–2528, 2006.

[11] Wendy A Breyer and Brian W Matthews. A structural basis for processivity. *Protein Science*, 10(9):1699–1711, 2001.

[12] Ellen R Edenberg, Michael Downey, and David Toczyski. Polymerase stalling during replication, transcription and translation. *Current Biology*, 24(10):R445–R452, 2014.

[13] Daniel D Kephart, Nicholas F Marshall, and David H Price. Stability of drosophila RNA polymerase II elongation complexes in vitro. *Molecular and cellular biology*, 12(5):2067–2077, 1992.

[14] E. Nudler, M. Kashlev, V. Nikiforov, and A. Goldfarb. Coupling between transcription termination and RNA polymerase inchworming. *Cell*, 81(3):351–7, 1995.

[15] X. Gu and W. F. Marzluff. 3' processing and termination of mouse histone transcripts synthesized in vitro by RNA polymerase ii. *Nucleic Acids Research*, 24(19):3797–3805, 1996. 8871561[pmid] Nucleic Acids Res.

[16] Natalia Komissarova and Mikhail Kashlev. Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proceedings of the National Academy of Sciences*, 94(5):1755–1760, 1997.

[17] Neal Crampton, William A. Bonass, Jennifer Kirkham, Claudio Rivetti, and Neil H. Thomson. Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. *Nucleic Acids Research*, 34(19):5416–5425, 2006.

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

[19] Gregory F Lawler. *Introduction to stochastic processes*. CRC Press, 2006.

[20] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[21] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry A*, 104(9):1876–1889, 2000.

[22] Marc R Roussel and Rui Zhu. Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression. *Physical Biology*, 3(4):274, 2006.

[23] Yang Cao, Daniel T Gillespie, and Linda R Petzold. The slow-scale stochastic simulation algorithm. *The Journal of Chemical Physics*, 122(1):014116, 2005.

[24] Dmitri Bratsun, Dmitri Volfson, Lev S. Tsimring, and Jeff Hasty. Delay-induced stochastic oscillations in gene regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 102(41):14593–14598, 2005.

[25] Thorleif Møller, Thomas Franch, Christina Udesen, Kenn Gerdes, and Poul Valentin-Hansen. Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. *Genes & Development*, 16(13):1696–1706, 2002.

[26] Jonathan A Bernstein, Arkady B Khodursky, Pei-Hsun Lin, Sue Lin-Chao, and Stanley N Cohen. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proceedings of the National Academy of Sciences*, 99(15):9697–9702, 2002.

[27] Marc R Roussel and Rui Zhu. Stochastic kinetics description of a simple transcription model. *Bulletin of Mathematical Biology*, 68(7):1681–1713, 2006.

[28] Lin Xu, Hong Chen, Xiaohua Hu, Rongmei Zhang, Ze Zhang, and ZW Luo. Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms. *Molecular biology and evolution*, 23(6):1107–1108, 2006.

[29] M Thomas Record, William S Reznikoff, Maria L Craig, Kristi L McQuade, and Paula J Schlax. Escherichia coli RNA polymerase (eσ70), promoters, and the kinetics of the steps of transcription initiation. *Escherichia coli and Salmonella*, 1:792–820, 1996.

[30] Joseph T Wade and Kevin Struhl. The transition from transcriptional initiation to elongation. *Current Opinion in Genetics & Development*, 18(2):130–136, 2008.

[31] Paul Singer and Cheng-Wen Wu. Promoter search by *Escherichia coli* RNA polymerase on a circular DNA template. *Journal of Biological Chemistry*, 262(29):14178–14189, 1987.

[32] Shasha Chong, Chongyi Chen, Hao Ge, and X Sunney Xie. Mechanism of transcriptional bursting in bacteria. *Cell*, 158(2):314–326, 2014.

[33] Andre S Ribeiro, Olli-Pekka Smolander, Tiina Rajala, Antti Hkkinen, and Olli Yli-Harja. Delayed stochastic model of transcription at the single nucleotide level. *Journal of Computational Biology*, 16(4):539–553, 2009.

[34] H. Bremer, P. Dennis, and M. Ehrenberg. Free RNA polymerase and modeling global transcription in *Escherichia coli*. *Biochimie*, 85(6):597–609, 2003.

[35] Somenath Bakshi, Albert Siryaporn, Mark Goulian, and James C Weisshaar. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Molecular Microbiology*, 85(1):21–38, 2012.

[36] Marko Djordjevic and Ralf Bundschuh. Formation of the open complex by bacterial RNA polymerasea quantitative model. *Biophysical Journal*, 94(11):4233–4248, 2008.

[37] Andrey Revyakin, Richard H Ebright, and Terence R Strick. Promoter unwinding and promoter clearance by RNA polymerase: detection by single-molecule DNA nanomanipulation. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14):4776–4780, 2004.

[38] J. Zeitlinger, A. Stark, M. Kellis, J. W. Hong, S. Nechaev, K. Adelman, M. Levine, and R. A. Young. RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nature Genetics*, 39(12):1512–6, 2007.

[39] David A Hendrix, Joung-Woo Hong, Julia Zeitlinger, Daniel S Rokhsar, and Michael S Levine. Promoter elements associated with RNA Pol II stalling in the *Drosophila* embryo. *Proceedings of the National Academy of Sciences*, 105(22):7762–7767, 2008.

[40] William R McClure, Carol L Cech, and David E Johnston. A steady state assay for the RNA polymerase initiation reaction. *Journal of Biological Chemistry*, 253(24):8941–8948, 1978.

[41] Hans Bremer and Patrick P Dennis. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, 3(1), 2008.

[42] Tamas Szekely and Kevin Burrage. Stochastic simulation in systems biology. *Computational and Structural Biotechnology Journal*, 12(20):14–25, 2014.

[43] Stephen Ramsey, David Orrell, and Hamid Bolouri. Dizzy: stochastic simulation of large-scale genetic regulatory networks. *Journal of Bioinformatics and Computational Biology*, 3(02):415–436, 2005.

[44] William S Levine. *Control System Fundamentals*. CRC press, 1999.

[45] Dwarkadas Pralhaddas Kothari and IJ Nagrath. *Basic Electrical Engineering*. Tata McGraw Hill, 2010.

[46] A. Pannenborg. *Electrical Instruments / Elektrische Instrumente*. Springer Berlin Heidelberg, 2012.

[47] Lars Dölken, Zsolt Ruzsics, Bernd Rädle, Caroline C Friedel, Ralf Zimmer, Jörg Mages, Reinhard Hoffmann, Paul Dickinson, Thorsten Forster, Peter Ghazal, et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA*, 14(9):1959–1972, 2008.

[48] Yuriy Mileyko, Richard I Joh, and Joshua S Weitz. Small-scale copy number variation and large-scale changes in gene expression. *Proceedings of the National Academy of Sciences*, 105(43):16659–16664, 2008.

[49] Uwe an der Heiden and Michael C Mackey. The dynamics of production and destruction: analytic insight into complex behavior. *Journal of Mathematical Biology*, 16(1):75–101, 1982.

[50] Daguang Wang, Timothy I Meier, Cathleen L Chan, Guohua Feng, Donna N Lee, and Robert Landick. Discontinuous movements of DNA and RNA in RNA polymerase accompany formation of a paused transcription complex. *Cell*, 81(3):341–350, 1995.

[51] M Schaechter. *Desk Encyclopedia of Microbiology*. Elsevier Science, 2003.

[52] Tianhai Tian, Kevin Burrage, Pamela M Burrage, and Margherita Carletti. Stochastic delay differential equations for genetic regulatory networks. *Journal of Computational and Applied Mathematics*, 205(2):696–707, 2007.

[53] Thomas G Kurtz. The relationship between stochastic and deterministic models for chemical reactions. *The Journal of Chemical Physics*, 57(7):2976–2978, 1972.

[54] Marc R Roussel. The use of delay differential equations in chemical kinetics. *The Journal of Physical Chemistry*, 100(20):8323–8330, 1996.

[55] Crispin W Gardiner et al. *Handbook of Stochastic Methods*, volume 3. Springer Berlin, 1985.

[56] P Reimann, GJ Schmid, and P Hänggi. Universal equivalence of mean first-passage time and kramers rate. *Physical Review E*, 60(1), 1999.

Appendix A

The MatLab code for the delay stochastic simulation modelling of arrested transcription as used in section 3.2 is given below.

```
% RNAP+PRO-> ELNC  [Firstreaction]  Rate constant:k1
% ELNC-> PRO(t+tau_1)+C1(t+tau_2)  [Second reaction]  Rate constant:k2
% ELNC-> PRO(t+tau_as+tau_arrest)+ELNC_as(t+tau_as)+RNAP(t+tau_as+tau_arrest)  [Third reaction]  Rate constant:k_arrest
% C1-> C2  [Fourth reaction]  Rate constant:k6
% C2-> RNAP+R  [Fifth reaction]  Rate constant:k7
% R-> Waste_product  [Sixth reaction]  Rate constant:k8
w=1; p=[]; q=[];
% Null array for entering Rss & rise time after every loop
t_half=14.5*60;
% Average half life of mRNA in E. coli in seconds
k1=10;  k2=1;  k_arrest=0.5;  k6=10;  k7=1;  k8=(0.693/t_half);  rnap=144;  kx=(k1*rnap);
% Pr(arrest)=(#arrest/total initiations)& Pr(arrest)=(k_arrest/(k_arrest+k2))
%Pr(arrest)=0.3 so #arrest=15000 when max_time=500000 and k_arrest=0.428
max_time=500000;  rng('shuffle');  numrepeats=2;  tau1=1;  tau2=20;  tau_as=0.5;  tau_arrest=1;
%Arrest happens usually at 20 nt from the start site and RNAP transcribes 40nt/sec so tau_as=20/40)
%% Step 1: Initialization
elnc=0;  c1=0;  c2=0;  r=0;  t=0;  rnap=50;  n=0;  pro=1;  elnc_as=0;
L=[];  arrest_array=[];  event_no=0;
% Empty list L for delayed generating events and corresponding time
% Constant labels: Initialize the value of generatepro, generatec1 & no. of events in L
generatepro=1;  generatec1=2;  generatepro_as=3;  num_events_in_L=0;
% Calculate initial propensities
```

```
23 A=zeros(6,1);
24 A(1)=kx*pro;
25 A(2)=k2*elnc;
26 A(3)=k_arrest*elnc;
27 A(4)=k6*c1;
28 A(5)=k7*c2;
29 A(6)=k8*r;
30
31 %% Step 2: Stochastic Simulation Algorithm step for reacting
       events to get the next reacting events R1 &
     corresponding occurence time t1
32 % Step 2(a): Calculate and store values of reaction
     propensities a1, a2, a3, a4, a5 & store the sum as Asums
33 for w=1:numrepeats
34     while t<max_time
35         Asums=cumsum(A);
36         % Step 2(b): Generates two random numbers r1 & r2
37         random_num=rand(2,1);
38         % Step 2(c): Calculate & store the values of time(t1
              ) & mu
39         t1=(1/Asums(6))*log(1/random_num(1));
40         mu=find(random_num(2)*Asums(6)<Asums,1);
41         % Displays the very first time condition is true &
              save value as mu, to represent the reaction that
              has     occurred
42         %% Step 3: Compare t1 with least time tau(min) in L
43         % Step 3(a): Get the generating events corresponding
               to the reacting event  R1 & generate delays tau(
              j) for each generating event G(j)
44         % Adds the generating events & corresponding delays
              in L by checking the condition, if t1<tau(min)
45         if(size(L,1)==0)||(t1<min(L(:,2)))
46             if (mu==1)  % Pro -> Elnc
47              % Value of mu equal to one, denotes that First
                    reaction occurred
48                 pro=pro-1;
49                 % Reduces the number of reactants and
                       increases no.of products molecules
50                 elnc=elnc+1;
51                 A(1)=kx*pro;
52                 % Recalculates new propensities
53                 A(2)=k2*elnc;
54                 A(3)=k_arrest*elnc;
55
56             elseif (mu==2)  % Elnc -> Pro(t+tau1)+C1(t+tau2)
```

```
57              % Value of mu equal to two, denotes that second
                   reaction occurred
58                  elnc=elnc-1;
59                  A(2)=k2*elnc;
60                  A(3)=k_arrest*elnc;
61
62                  num_events_in_L=num_events_in_L+1;
63                  % Increases the count of the no. of events
                       in L as delayed events are added
64                  L(num_events_in_L,1)=generatepro;
65                  % Adds the generatepro in list of generating
                        events in L
66                  L(num_events_in_L,2)=tau1+t1;
67                  % Adds the tau2(corresponding delay of
                       generatec1) in list of delays in L
68                  % Adds the tau1(corresponding delay of
                       generatepro) in list of  delays in L
69                  % We add t1 so that the subsequent
                       subtraction of t1 puts this event tau1
                       events into the future from t
70                  num_events_in_L=num_events_in_L+1;
71                  L(num_events_in_L,1)=generatec1;
72                  % Adds the generatec1 in list of generating
                       events in L
73                  L(num_events_in_L,2)=tau2+t1;
74                  % Adds the tau2(corresponding delay of
                       generatec1) in list of delays in L
75
76              elseif (mu==3)  % Elnc -> Pro(t+tau_as+
                   tau_arrest)+ElnC_as(t+tau_as)+RnaP(t+tau_as+
                   tau_arrest)
77              % Value of mu equal to three, denotes that third
                    reaction occurred
78                  elnc=elnc-1;
79                  elnc_as=elnc_as+1;
80                  A(3)=k_arrest*elnc;
81                  A(2)=k2*elnc;
82
83                  % Increases the count of the no. of events
                       in L as delayed events are added where L
                       is the list for the arrested entities
84                  num_events_in_L=num_events_in_L + 1;
85                  L(num_events_in_L,1)=generatepro_as;
86                  % Adds the generatepro in list of generating
                        events in L
```

```
87                  L(num_events_in_L,2)=tau_as+tau_arrest+t1;
88                  % Adds the tau1(corresponding delay of
                        generatepro) in list of delays in L
89
90              elseif (mu==4)  % C1 -> C2
91              % Value of mu equal to four, denotes that fourth
                    reaction occurred
92                  c1=c1-1;
93                  c2=c2+1;
94                  A(4)=k6*c1;
95                  A(5)=k7*c2;
96
97              elseif (mu==5)  % C2 -> R
98              % Value of mu equal to five, denotes that fifth
                    reaction occurred
99                  c2=c2-1;
100                 r=r+1;
101                 A(5)=k7*c2;
102                 A(6)=k8*r;
103
104             else  % R -> waste_product
105                 r=r-1;
106                 % Sixth reaction occurred, leading to the
                        degradation of one rna
107                 A(6)=k8*r;
108             end
109
110             % Step 3(b): Deletes the generating events &
                    their corresponding time from L, if t1>tau(
                    min)
111         else
112             [t1,nextevent]=min(L(:,2));
113             % Identify which delay in L is smallest & its
                    position to get the corresponding generating
                    event
114             % If generating event corresponding to the min.
                    delay is generatepro or generatec1 then one
                    promoter or c1 is released resp.
115
116             if L(nextevent,1)==generatepro
117                 pro=pro+1;
118                 A(1)=pro*kx;
119                 num_events_in_L=num_events_in_L-1;
120                 % Total number of reacting events in L are
                        reduced by one as delayed events are
```

```
                deleted
121             L=[L(1:nextevent-1,:);L(nextevent+1:end,:)];
122             % Delete generating event & corresponding
                    delay in L
123
124         elseif L(nextevent,1)==generatec1
125             c1=c1+1;
126             A(4)=k6*c1;
127             num_events_in_L=num_events_in_L-1;
128             L = [L(1:nextevent-1,:);L(nextevent+1:end,:)
                    ];
129
130         elseif L(nextevent,1)==generatepro_as
131             pro=pro+1;
132             A(1)=pro*kx;
133             num_events_in_L = num_events_in_L-1;
134             % Total number of reacting events in L are
                    reduced by one as delayed events are
                    deleted
135             L = [L(1:nextevent-1,:);L(nextevent+1:end,:)
                    ];
136
137         else
138             error('error\n');
139         end
140     end
141
142     t=t+t1;
143     % Increases the value of time variable by t1
144
145     % Decrement all delays in L by t1
146     if size(L)~=0
147         L(:,2)=L(:,2)-t1;
148     end
149
150     %% Step 4: Stores the value of variables at n
151     n=n+1;
152     T(1,n)=t;
153     %PRO(1,n)=pro;
154     ELNC(1,n)=elnc;
155     ELNC_as(1,n)=elnc_as;
156     R(1,n)=r;
157     %RNAP(1,n)=rnap;
158     %C1(1,n)=c1;
159     %C2(1,n)=c2;
```

```
160     end
161
162     %% tells us about the first steady state in RNA mol
          concentration
163     lower_limit=round(n/2);
164     rise_time_index=0;
165     last_rise_time_index=n;
166     while abs(rise_time_index-last_rise_time_index) > 1
167         meanR=mean(R(lower_limit:n));
168         last_rise_time_index = rise_time_index;
169         rise_time_index=find(R>0.9*meanR,1);
170         lower_limit=rise_time_index*10;
171         if lower_limit > n
172             error('Rise time calculation failed.');
173         end
174     end
175     rise_time = T(rise_time_index);
176
177     %% stationary rna concentration
178     mean_1=mean(R(lower_limit:n));
179     std_1=std(R(lower_limit:n));
180
181     %% to record the value of rss & rise time after every
          loop
182     p(w)=mean_1;
183     q(w)=rise_time;
184 end
185
186 mean_rss=mean(p);
187 mean_rise_time=mean(q);
188
189 % standard error(SE=SD/sqrt(N))
190 SD_rss = std(p(1:numrepeats));
191 SE_rss = SD_rss/sqrt(numrepeats);
192 SD_rise_time= std(q(1:numrepeats));
193 SE_rise_time= SD_rise_time/sqrt(numrepeats);
```