Modeling of nucleation-based stochastic processes in cellular systems

Xiaohua Xu

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Physics

Rahul Kulkarni, Chair

Michel Pleimling

Kyungwha Park

Hans Robinson

August 11th, 2010

Blacksburg, Virginia

Keywords: Stochastic modeling, nucleation, first passage time

Modeling of nucleation-based stochastic processes in cellular systems

Xiaohua Xu

(ABSTRACT)

Molecular cell biology has been an intensively studied interdisciplinary field with the rapid development of experimental techniques and fast upgrade of computational hardware and numerical tools. Recent technological developments have led to single-cell experiments which allow us to probe the role of stochasticity in cellular processes. Stochastic modeling of the corresponding processes is thus an essential ingredient for the understanding and interpretation of cellular systems of interest. In this thesis, we explore several nucleation-based stochastic cellular processes, i.e. Min protein oscillation in *Escherichia coli*, pausing phenomena in DNA transcription, and single-molecule enzyme kinetics. We focus on the key experimental results and build up stochastic models accordingly to provide quantitative insights to the underlying physical mechanisms for the corresponding biological processes. We utilize specific mathematical methods and computational algorithms to gain a better understanding and make predictions for further experimental explorations in the relevant fields.

Acknowledgments

First of all, I would like to thank my adviser Dr. Rahul Kulkarni for his tremendous help, professional guidance and support throughout my PhD study at Virginia Tech. He has not only taught me knowledge of statistical mechanics, cellular biology, stochastic modeling, etc., but also guided me on how to make a successful presentation, how to pursue a suitable career, how to take priority and have perspectives in life, and many other things that will benefit me in the long run. I have always been greatly inspired by his passionate way of discussion, stimulated by his creative interpretation in research, and I am extremely grateful for his encouragement and emotional support during my PhD study and job searching. I feel very lucky to have him as my adviser. There is one Chinese saying that "mentor for once, parent for life". I will carry on my respect for him and all that I have learned from him in my future life long after graduation.

I would also like to thank all my committee members: Dr. Michel Pleimling, Dr. Kyungwha Park, and Dr. Hans Robinson. I really appreciate the time and effort they put in reading and commenting on this manuscript. A special thank you goes to Dr. Michel Pleimling, who read my thesis with utmost care and patience. His suggestions were extremely helpful.

I would like to acknowledge my friends Jason Ridley, Tao Jia and Eugene Halpin for helpful and insightful discussions when they helped me with my

defense presentation.

My special thanks to my parents and my dearest friends Krisha Chachra and Derek Klinedinst, Michael Kevic and Bing Feng, David Lillemon, Vitor Leite Nunes, Xiaoyi Zhong and Ying Xiong. My life in Blacksburg would have turned out totally different if they weren't there for me. I will love them for always. I especially thank Krisha. Her love and care has accompanied me ever since my arrival in this foreign country. She has been someone way beyond a friend to me. Krisha and Derek treat me as a family member. I'm so glad that we have developed this trusting relationship along the way I'm pursuing my PhD.

Contents

List of Figures

Chapter 1

Introduction

Over the past several decades, progress in experimental techniques has allowed quantitative observations of intercellular or intracellular processes and precise measurements of the reaction rates. With the vast amount of quantitative information uncovered, our understanding of the dynamic phenomena in cells has gained a revolutionary improvement. On the other hand, an increased demand on mathematical analysis and computational simulations has been raised to confirm previous hypothesis and to develop models and predictions for further experimental explorations.

1.1 Role of computational modeling

Computational modeling of biological processes has drawn an increasing attention in cell biology with the emerging quantitative experimental cellular-level data. Most of the kinetic processes in biological cells such as growth, cell

division, intercellular communication or cells' response to the environment, are complicated nonlinear processes, so that approximations are required to evaluate their evolution in time. When an exact solution is generally difficult or impossible to obtain analytically using traditional mathematical methods, we can derive quite good estimates using numerical methods implemented on computers. To build a successful computational model we usually make some simplifications that not only make sense in terms of the physical process of interest but also are valid from a mathematical standpoint, apart from which we also need to develop algorithms that are well balanced between accuracy and efficiency. For many years computer models have proved to be an important tool in helping to understand the experimental measurements and to guide further explorations.

1.2 Research background

Chemical reactions are stochastic events, meaning that it is not possible to know when and where the next reaction will occur. However, the probabilities for the reaction events can be modeled and the time evolution of the system can therefore be described probabilistically. At any particular time, the state of the biological system is modeled as a random variable. The system evolves by undergoing transitions to other states based on the reaction rates characterizing system interactions. In some cases the stochastic process is irreversible, i.e. the backward process is not permitted, once the system

reaches a specific state which we broadly term as "nucleation". Originally, nucleation refers to the onset of the accumulation of a certain number of molecules when the cluster becomes stable and the dissociation can be negligible compared to the dramatically increased association. In molecular cell biology, this term is borrowed to describe a system with several transient states before reaching a final absorbing state. This reaction scheme is seen to occur frequently when developing models for important biological processes. Modeling of such stochastic processes occurring in different contexts can potentially provide insight into the role of "nucleation " in cellular processes. In this dissertation we analyze three nucleation-based systems: the Min protein oscillation in *E. coli*, pausing phenomena in DNA transcription and single-molecule enzyme kinetics.

1.2.1 Min protein oscillation in *Escherichia coli*

Spatial variation is often an important feature in cellular mechanisms and the specific subcellular localization of the relevant proteins sometimes determines the bacterial cell fate. One prominent example is the Min proteins regulating accurate cell division in *Escherichia coli* (*E. coli*) that we will discuss in detail in Chapter 3. Research has revealed that the rod-shaped bacteria *E. coli* symmetrically divide at mid-cell to guarantee one copy of the chromosomes in each daughter cell during cell division. This self-organized behavior has been discovered to be regulated by a group of proteins called "Min proteins". The

diffusion and interactions between the Min proteins cause some interesting oscillatory spatiotemporal pattern formation, which is believed to be crucial for an accurate cell division in *E. coli*. The mechanism of Min protein oscillation is complicated and not analytically tractable, but we present a simplified computational model which focused on some key processes to explain experimental observations. We provide a novel understanding of the dynamics behind precise subcellular localization of Min protein pattern formation and *E. coli* cell division.

1.2.2 DNA transcriptional pausing

In Chapter 4 we will look into the transcriptional pausing phenomena that has been studied for decades but is not yet fully understood. DNA transcription, also called mRNA synthesis, plays an important role in gene expression. Transcription is roughly divided into three phases: initiation, elongation and termination. The transcriptional elongation phase has been detected to pause frequently both in prokaryotes and eukaryotes. Most of the pauses are transient and the elongation is resumed after the interruption. We focus on the distribution of times for the system to get back onto the elongation pathway from a pausing state, and we propose a quantitative model to understand how the pausing duration changes with different kinetic rates characterizing pause state interactions.

1.2.3 Single-molecule enzyme kinetics

Enzymes are needed in almost all cellular processes and enzyme kinetics has been well interpreted by the Michaelis-Menten (MM) model. According to the MM equation as well as experimental results, the reaction rate of enzymes increases with the substrate concentration but saturates when all the free enzymes are converted into the substrate-bound form. In Chapter 5, we study the issue on a single-molecule level. We particularly focus on the role that substrate fluctuations plays on the turnovers of the enzymatic reactions.

1.3 Summary

The stochastic dynamics of different molecular biochemical reactions can be analyzed and simulated by computational models. Analytical solutions or efficient numerical simulations for a quantitative understanding of complex reaction systems can provide fresh insights. In this research, we model and analyze the underlying mechanisms of several nucleation-based cellular processes. With reasonable assumptions and approximations, we derive analytical relations which are validated by stochastic simulations. With continuing technological innovations leading to more wide-spread single-cell and single-molecule assays, we believe that the results derived will be of considerable interest to future explorations, both theoretical and experimental, of biological systems.

Chapter 2

Modeling methods

In cellular biology, many phenomena can be interpreted as probabilistic occurrences and the problems in the biological systems can therefore be approached statistically. The biochemical processes such as the interaction between relevant proteins, their diffusion in the cytoplasm or binding to the cell membrane are often modeled as Markov processes. In this chapter we introduce several theoretical approaches and tools, i.e. the chemical master equation, generating function and first passage time calculations, which we will use to analyze stochastic cellular processes and to model the corresponding dynamics in the later chapters. We also discuss a specific computational method, the Gillespie algorithm, that is used for the stochastic simulations.

2.1 Chemical master equation

Biological cells are complex systems in which the biochemical reactions between the protein molecules are considered as discrete and random encounters, therefore it is important to consider stochasticity in the cell dynamics. We use the chemical master equation (CME) to describe the time evolution of the probability of the system occupying one of the discrete states.

$$\frac{dp(x,t)}{dt} = \sum_{x \neq x'} [T_{x' \to x} p(x',t) - T_{x \to x'} p(x,t)] \tag{2.1}$$

where $p(x,t)$ is the probability for the biological system to be in state x at time t, and $T_{x \to x'}$ is the rate constant that corresponds to the transition from state x to state x'. The R.H.S. thus indicates the gain or loss in probability for the system to be in state x. The CME exhibits detailed balance if the system is in equilibrium.

$$T_{x' \to x} p(x',t) = T_{x \to x'} p(x,t)$$

In the analysis, we typically consider various paths the system may take starting from the given initial state. In contrast to this, we can also list all possible paths the system may undergo to reach a given final state. The latter corresponds to the "backward" master equations [4]. For instance, consider the reaction scheme corresponding to a 1D random walk in Eq.(2.2),

$$\cdots \; m \underset{\mu_{m+1}}{\overset{\lambda_m}{\rightleftharpoons}} m+1 \underset{\mu_{m+2}}{\overset{\lambda_{m+1}}{\rightleftharpoons}} \cdots \underset{\mu_n}{\overset{\lambda_{n-1}}{\rightleftharpoons}} n \underset{\mu_{n+1}}{\overset{\lambda_n}{\rightleftharpoons}} n+1 \; \cdots \tag{2.2}$$

if we want to describe the probability that the system starts from state m and ends up at state n at time t, we analyze the forward master equation which is

$$\frac{dP_{n,m}(t)}{dt} = \lambda_{n-1}P_{n-1,m}(t) - (\mu_n + \lambda_n)P_{n,m}(t) + \mu_{n+1}P_{n+1,m}(t) \tag{2.3}$$

where μ_n is the rate for a transition to the left, to state $n-1$, and λ_n is the rate for a transition to the right, to state $n+1$. Alternatively, sometimes we are interested in the case that the system is in state n at time t, i.e. instead of specifying the initial condition at time $t = 0$, we have fixed the final condition at time t. We want to know, for every state m, what is the probability of the system ending up in state n at time t. In such cases, we have the backward master equation as follows.

$$\frac{dP_{n,m}(t)}{dt} = \lambda_m P_{n,m+1}(t) - (\mu_m + \lambda_m)P_{n,m}(t) + \mu_m P_{n,m-1}(t) \tag{2.4}$$

The calculations based on the master equation for a particular model can be compared to experimental results, which will then tell us whether or not the model is an accurate one for a particular situation. In some cases we can solve the master equations directly, while in others we have to resort to numerical

solutions such as using the Monte Carlo method which will be discussed later on in this chapter.

2.2 Generating function

A generating function is also called a probability-generating function when it represents a power series of the probability mass function, which is a function that gives the probability that a discrete random variable is exactly equal to some value. The one indeterminate parameter z usually has a sequence of non-negative coefficients p_n that encode the probabilistic information of the system. For a single random variable with probability distribution $p_n(t)$ at time t, the generating function is defined as

$$G(z,t) = \sum_{n=-\infty}^{\infty} p_n(t) z^n \tag{2.5}$$

Then the system can be analyzed through the manipulation of the generating function in case the calculations of the probabilities are difficult or underivable. For example the k^{th} moment of the random variable x can be found by

$$< x^k > (t) \equiv \sum_{n=-\infty}^{\infty} x^k p_n(t) = \left. \frac{\partial^k G(z,t)}{\partial (\ln z)^k} \right|_{z=1} \tag{2.6}$$

Sometimes the generating function is extremely helpful to solve the master equations. In that case the appropriate initial condition $G(z,0) = \sum_n p_n(0) z^n$ and the boundary condition $G(1,t) = \sum_n p_n(t) = 1$ are utilized.

2.3 First passage time

In many cases, the process of interest can be modeled as a 1D random walk process in which the lower bound is a reflecting boundary and the upper bound is an absorbing boundary. Then given infinitely long time, the random walker will eventually hit the absorbing boundary thus exit the system. Now we wish to find out characteristics of the distribution of times when the random walker leaves the system.

2.3.1 Mean first passage time

First, we intend to obtain the mean value of the distribution of first passage times for the system with one reflecting boundary and one absorbing boundary [5]. Particularly, when a random walker starts from position $x = x_0$ at time $t = t_0$, after a time interval $t - t_0$, the survival probability for the walker to remain in the system is

$$G(x_0, t - t_0) \equiv \int_a^b p(x, t|x_0, t_0)\, dx \tag{2.7}$$

where a and b are the boundaries of the system and $p(x,t|x_0,t_0)$ is the probability that the walker is at position x at time t after starting from position x_0 at time t_0. Eq. (2.7) also represents the probability that first passage time is equal to or greater than the time t, i.e. the cumulative probability for the first passage time T_{fp},

$$\int_a^b p(x,t|x_0,t_0)\,dx = \text{Prob}(T_{fp} \geq t) = G(x_0, t-t_0)$$

Thus the mean first passage time is given by

$$< T_{fp} > = -\int_{t_0}^{\infty} t\;\partial_t G(x_0, t-t_0)dt = \int_{t_0}^{\infty} G(x_0, t-t_0)dt \qquad (2.8)$$

Let us denote $T_{fp}(x_0) \equiv < T_{fp} >$ as the mean first passage time starting from position x_0. Then we use the backward master equation

$$\begin{aligned} \partial_t p(x,t|x_0,t_0) &= \lambda(x_0)\,[p(x,t|x_0+1,t_0) - p(x,t|x_0,t_0)] \\ &+ \mu(x_0)\,[p(x,t|x_0-1,t_0) - p(x,t|x_0,t_0)] \end{aligned}$$

where λ is the forward rate and μ is the backward rate as mentioned before. If we integrate the above equation on both sides over x and t, we have

$$\begin{aligned} \text{R.H.S.} &= \int_{t_0}^{\infty} dt\;\{\;\lambda(x_0)\,[G(x_0+1,t-t_0) - G(x_0,t-t_0)] \\ &\qquad + \mu(x_0)\,[G(x_0-1,t-t_0) - G(x_0,t-t_0)]\;\} \\ &= \lambda(x_0)\,T_{fp}(x_0+1) + \mu(x_0)\,T_{fp}(x_0-1) - [\lambda(x_0)+\mu(x_0)]\,T_{fp}(x_0) \end{aligned}$$

As for the L.H.S.

$$\int_a^b dx \int_{t_0}^{\infty} \partial_t p(x,t|x_0,t_0)dt = G(x_0,\infty) - G(x_0,0) = -1$$

Therefore

$$-1 = \lambda(x_0)\, T_{fp}(x_0+1) + \mu(x_0)\, T_{fp}(x_0-1) - [\lambda(x_0)+\mu(x_0)]\, T_{fp}(x_0) \quad (2.9)$$

Besides, we have reflecting boundary at $x = a$,

$$p(x,t|a-1,t_0) = p(x,t|a,t_0) \quad \text{or} \quad T_{fp}(a-1) = T_{fp}(a)$$

and we have absorbing boundary at $x = b$,

$$p(x,t|b+1,t_0) = 0 \quad \text{or} \quad T_{fp}(b+1) = 0$$

Now we can solve Eq. (2.9) with the above boundary conditions

$$T_{fp}(x_0) = \sum_{y=x_0}^{b} \phi(y) \sum_{z=a}^{y} \frac{1}{\lambda(z)\phi(z)} \quad (2.10)$$

where

$$\phi(x) = \prod_{z=a+1}^{x} \frac{\mu(z)}{\lambda(z)}$$

2.3.2 First passage time probability distribution

In the preceding section we derived the mean value of the first passage time probability distribution. In this section we will focus on deriving the entire probability distribution of the first passage time, $F(t)$. For a random walker that starts from state m and hits the absorbing state n after time t, We use $F_{n,m}(t)$ to denote the probability density function of this first passage time. $F_{n,m}(t)$ is known to follow the backward master equation [6].

$$\frac{dF_{n,m}(t)}{dt} = \lambda_m F_{n,m+1}(t) - (\lambda_m + \mu_m)F_{n,m}(t) + \mu_m F_{n,m-1}(t)$$

Where λ_m is the transition rate to the right and μ_m is the transition rate to the left. Then we can apply the boundary conditions to solve the equations for the specific system of interest, which in our case is a four-state absorbing Markov process (detailed in Appendix B).

2.3.3 Phase-type distribution

There is an alternative way to obtain the first passage time probability distribution – that is to use the phase-type distribution [7]. A phase-type distribution is a probability distribution of time for a system of one or more inter-related Poisson processes occurring in sequence to reach the final absorbing state. The sequence in which each of the phases occur may itself be a stochastic process, and the distribution can be represented by a random vari-

able describing the time until the absorption of a Markov process, in which each of the states represents one of the phases.

If we consider a continuous-time Markov chain with $(n+1)$ states, among which the first n states are transient states and the last one state $n+1$ is an absorbing state, we can write the equation for the Markov chain as

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{x}(\mathbf{t})\, Q \\ \mathbf{x}(0) = \vec{\alpha} \end{cases} \tag{2.11}$$

where $\mathbf{x}(t)$ is a row vector whose i^{th} component stands for the probability of being in state i at time t, and $\vec{\alpha}$ is the initial condition of vector $\mathbf{x}(t)$. Also, Q is an $(n+1) \times (n+1)$ matrix such as

$$Q = \begin{bmatrix} T & \nu \\ 0 & 0 \end{bmatrix}$$

Here T is an $n \times n$ matrix that gives the transition rates of all the non-absorbing states, particularly the T_{ij} component $(i \neq j)$ of which represents the transition rate from state i to state j. While ν is a vector whose i^{th} component denotes the transition from state i to the absorbing state, so $\nu = -T\ \vec{1}$, where $\vec{1}$ is the unit vector. Note that the time derivative of the absorbing state $\dot{\mathbf{x}}_{\mathbf{n+1}}(t)$ is actually the probability density function of the absorption time, $F(t)$, or the probability density function of the first passage time. Let us rewrite the equations in (2.11) in a different way as follows.

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{x}(t)\ T \\ F(t) = \dot{\mathbf{x}}_{\mathbf{n+1}}(t) = \mathbf{x}(t)\ \nu \\ \mathbf{x}(0) = \vec{\alpha} \end{cases} \tag{2.12}$$

The solution for $F(t)$ is

$$F(t) = \vec{\alpha}\ e^{t\ T}\nu$$

We can see that $F(t)$ is a vector whose i^{th} component denotes the probability of first passage time starting from state i, which is equivalent to $F_{n+1,i}(t)$ in the previous section. We find it less complicated to deal with the exponential term in the result if we can diagonalize the transition rate matrix T. Once we can find a diagonalized similar matrix of T such that $M = P^{-1}TP$, the n^{th} power of T can be rewritten as

$$T^n = (PMP^{-1})^n = (PMP^{-1}) \cdot (PMP^{-1}) \cdots (PMP^{-1}) = PM^nP^{-1}$$

Consequently

$$F(t) = \vec{\alpha} \sum_n \frac{(t\ T)^n}{n!}\ \nu = \vec{\alpha} \sum_n P\ \frac{(t\ M)^n}{n!}\ P^{-1}\nu$$

So

$$F(t) = \vec{\alpha}\ e^{t\ T}\nu = \vec{\alpha}\ Pe^{t\ M}P^{-1}\nu \tag{2.13}$$

where M is a diagonalized similar matrix of the transition rate matrix T.

The method of phase-type distribution is equivalent to the backward master equations method. We have demonstrated one detailed practice calculation in Appendix C as an example.

2.4 Numerical methods: the Gillespie algorithm

Molecules in a biological cell exhibit some degree of randomness in their dynamical behavior, therefore researchers who study the time evolution of some well-stirred chemically reacting systems should take into account their stochastic characteristic during the analysis of these cellular systems in biology. One successful method for the numerical simulations of the stochastic chemical kinetics in cells is the Gillespie algorithm [8].

Given a Master equation describing a reaction scheme, the Gillespie algorithm is an accurate approach to simulate the stochastic process. It is essentially a real-time Monte Carlo algorithm which involves averaging over different realizations of the stochastic dynamics. Each realization is generated by iteratively updating the state of the system. The updates involve drawing random variables from appropriate probability distributions to determine: (i) when the next reaction takes place and (ii) which of the possible reactions occurs. One advantage to utilize the Gillespie algorithm is that it is more time efficient than the regular stochastic simulation algorithm (SSA) especially when the reaction rates are small and the rejection rates are high. In other

words, there are many time intervals to be skipped before one action is taken place in the regular SSA, but for the Gillespie algorithm we can find out the time τ between the occurrence of two events and its probability distribution $f(\tau)$.

2.4.1 Survival probability

The probability density for the occurrence of the next event during $[t, t+\Delta\tau)$ can be derived from the probability of no events happening until time t (the survival probability $S(t)$). For the simplest case of a event occurring with constant probability per unit time k, we have

$$\begin{aligned} S(t+\Delta\tau) &= S(t)[1-k\Delta\tau] \\ \frac{S(t+\Delta\tau)-S(t)}{\Delta\tau} &= -kS(t) \\ S(t) &\sim e^{-kt} \end{aligned}$$

The waiting time distribution for the next event $f(t) = -dS(t)/dt$ then follows an exponential distribution. Note that the value of the corresponding cumulative distribution function of $f(t)$ varies from 0 to 1. If we pick a random number from a uniformly distributed interval $[0, 1]$, we can obtain a corresponding random number τ from the consistent interval $[0, \infty]$ distributed according to $f(t)$ using

$$\tau = \frac{1}{k} \ln \left[\frac{1}{r} \right] \tag{2.14}$$

2.4.2 Simulation steps

The simulation steps for the Gillespie algorithm are as follows

1. Generate two random numbers r_1, r_2 from a uniformly distributed interval $[0, 1]$;

2. Compute the propensity functions $\alpha_i(t)$ for each reaction. A propensity function for a given reaction gives the rate of occurrence for a single outcome based on averaging over large ensembles. The total propensity function for n reactions is thus $\alpha_0 = \sum_{i=1}^{n} \alpha_i(t)$;

3. Compute the time at which the next chemical reaction takes place as $t + \tau$ where τ is derived from Eq.(2.14);

$$\tau = \frac{1}{\alpha_0} \ln \left[\frac{1}{r_1} \right];$$

4. Compute which reaction occurs at time $t + \tau$, find j such that $r_2 \geq \frac{1}{\alpha_0} \sum_{i=1}^{j-1} \alpha_i$ and $r_2 < \frac{1}{\alpha_0} \sum_{i=1}^{j} \alpha_i$;

5. Update the biological system with the corresponding reaction and continue with $t = t + \tau$.

Chapter 3

Min protein oscillation in *Escherichia coli*

Recent research has highlighted several examples wherein bacterial cell fate is determined by precise subcellular localization of proteins. A prominent example is the polar localization and oscillation of the Min proteins which is necessary for accurate cell division in *E. coli*. Several computational models have been proposed which reproduce the oscillatory behavior and observed phenotypes. However, these models use varying assumptions to do so leading to different mechanisms for precise polar localization of MinD zones. In this chapter, we present a simplified computational model which focused on some key processes to explain the observed length scale for MinD zone formation. Using analytical approaches and numerical simulations, we explore cellular MinD distributions produced by these processes and we propose a mechanism for precise polar localization of MinD.

3.1 Introduction

With the development of genetics and cell biology techniques, increasing attention has been drawn to mechanisms of precise subcellular localization of proteins in bacterial cells. For instance, spatial localization of proteins governs asymmetric cell division in *Caulobacter cresentus* [9], the *Spo0J-Soj* system plays a critical role during sporulation in *Bacillus subtilis* [10] and the Min proteins regulate accurate cell division in *E. coli* [2, 11, 12, 13, 14, 15, 16, 17]. It has been shown that cell fate is largely dependent on the spatial localization of the relevant proteins in the above critical bacterial processes. Understanding the mechanisms governing cellular protein localization is thus an important challenge in microbiology. A prominent example is to study the mechanism of Min oscillations for accurate cell division in *E. coli*.

3.1.1 The cytokinesis machinery

E. coli is a rod-shaped bacterium that grows and reproduces by elongating along its long axis, duplicating its DNA into two nucleoids and dividing symmetrically into two daughter cells. Each daughter cell receives one of the two replicated and segregated chromosomes that are collected in the region of nucleoids on either side of the mother cell. If the division septum forms in the polar region, an asymmetric division occurs and results in minicelling, a phenotype such that one daughter cell gets two chromosomes and the other

one none. Studies however revealed that the position of the septum in *E. coli* cell division is precisely placed at mid-cell and the process is governed by FtsZ ring formation [13]. The protein FtsZ can be considered as the bacterial homologue of tubulin, which assembles into a polymeric ring on the inner side of the cytoplasmic membrane called the Z-ring. The Z-ring serves as the scaffold that recruits other proteins to form a complete septal ring to carry out cytokinesis. Therefore, the Z-ring determines the division site and its location is critical.

The placement of the Z-ring is regulated both by nucleoid occlusion [18] and by the Min protein system. Nucleoid occlusion suppresses the Z-ring formation in the regions surrounding the nucleoids to exclude cell division over chromosomes, which restricts the possible division site to the mid-cell region and the two cell poles. Min proteins undergo rapid pole-to-pole oscillations to inhibit polymerization of the FtsZ protein in the polar areas [2, 11, 12, 13, 14, 15, 16, 17]. Thus the combined effect of the two directs the location of the Z-ring as well as the site of *E. coli* cell division to mid-cell.

3.1.2 The Min protein system of *E. coli*

It has been discovered that the precise mid-cell localization of Z-ring in *E. coli* cell division predominantly depends on the regulation of the Min protein system, which is so called for its function of preventing minicell phenotype. The Min system consists of three Min proteins, MinC, MinD and MinE.

The formation of the Z-ring is suppressed by MinC [19, 20], whose minimum time-averaged concentration at mid-cell enables FtsZ polymerization in the cell center. MinC however does not contribute to the oscillatory behavior of the system and follows the dynamics of MinD [1, 21]. MinD is an ATPase and the ATP-bound form (MinD:ATP) can independently attach to the cell membrane [22] and recruit MinC and MinE from the cytoplasm [16, 23, 24, 25]. When MinC is expressed in the absence of MinD, it is homogeneously distributed in the cytoplasm (See Figure 3.1A) and there is no specifically preferred site for the Z-ring formation. On the other hand the oscillation of MinD (together with MinC) is largely dependent on MinE, which induces ATP hydrolysis and releases MinD:ADP, MinC and itself from the membrane [15, 17, 25, 26, 27, 28]. If MinE is absent, MinD and MinC will be distributed evenly over the cell membrane and block the Z-ring formation along the entire membrane (See Figure 3.1B). However, the cytoplasmic MinD:ADP cannot rebind to the cell membrane until it undergoes diffusion and nucleotide exchange into MinD:ATP. In short, the observed oscillatory pattern of MinD:ATP associating with the membrane to form transient polar zones is believed to be resulted from the self-organization of MinD and MinE.

3.1.3 Experimental observations

Researchers have shown intensive interest in the remarkable dynamic behavior of the Min protein oscillation that provides positional information for the

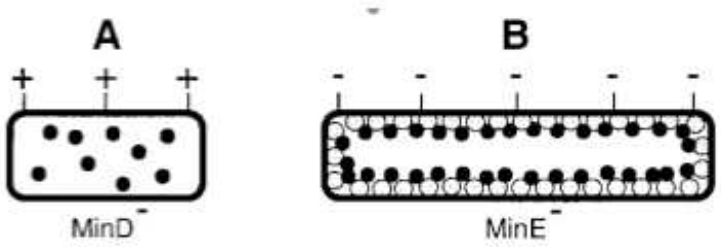


Figure 3.1: Schematic illustration of Min protein interactions. The circles represent MinD molecules and the solid dots are MinC molecules. The "+" sign indicates a possible division site and "-" sign refers to a site that Z-ring formation is blocked. (A) In the absence of MinD, MinC localizes randomly in the cytoplasm and has no effect on Z-ring formation. (B) In the absence of MinE, MinC associates with MinD along the cell membrane and prevents Z-ring formation at all sites. (This figure is modified from D. Raskin's paper [1].)

cell division without stationary positional markers. The experiments show that MinD undergoes a pole-to-pole oscillation with a period of $40 \sim 60$ s [16, 17]. At the beginning of an oscillatory cycle most MinD is associated in a membrane bound polar zone (old polar zone), and most MinE forms a ring structure (called E-ring) at the edge of the MinD polar zone. As the cycle progresses, MinD is released as E-ring moves toward the pole, observed as the shrinkage of the old polar zone. As this polar zone disassembles, a new zone is established at the *opposite* pole of the cell, and the cycle is repeated as shown in Figure 3.2. In contrast to the scenario in wild-type (wt) *E. coli* that the new MinD zone always starts at the opposite cell end, MinD forms an oscillating striped pattern with a characteristic length scale across the filamentous *E. coli* cells [14]. A filamentous cell occurs when FtsZ protein is extracted and the cell keeps growing without dividing, but the Min protein oscillatory manner remains. The new MinD zones are observed to be located

in the middle of the two adjacent old MinD zones.

3.2 The model

3.2.1 Overview of existing models

Several numerical models have been developed which successfully reproduce the oscillations of Min proteins and, to varying extents, the observed phenotypes of various mutants [29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43]. All of these models take into account the interaction between MinD, MinE and the cytoplasmic membrane, but they use varying assumptions leading to different mechanisms for precise polar localization of MinD zones. The model provided by Meinhardt *et al* [29] includes synthesis and degradation of the proteins and it states that at high concentrations MinD no longer recruits MinE. Howard *et al* propose some models that require cytoplasmic MinD to recruit cytoplasmic MinE and form MinDE complexes, and they consider no delay for nucleotide exchange [30, 31, 32]. In some model the membrane-bound MinD molecules are treated with diffusion rates [36]. However, the key questions relating to the mechanism for precise polar localization of MinD zones remain. For example, one of the striking features of the experimental results noted above is that for wt *E. coli* cells, the new MinD zone always begins from the opposite (bare) pole and then extends towards mid cell. While it is possible that this is due to the poles being more

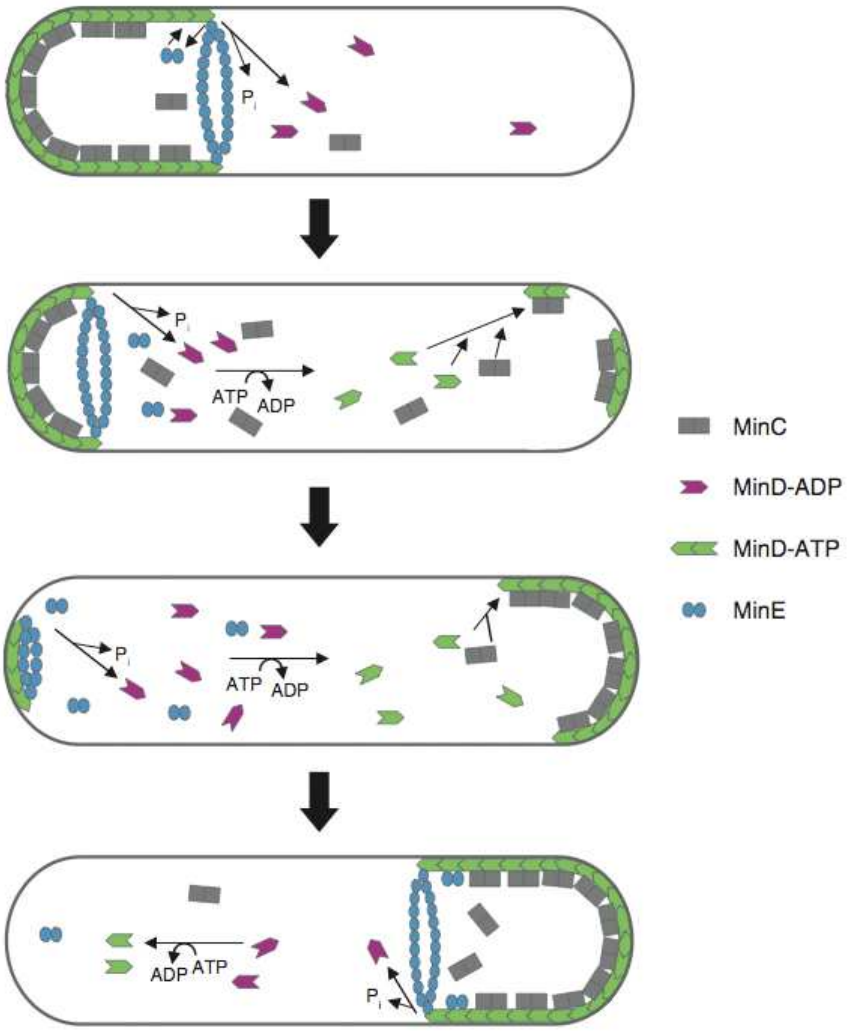


Figure 3.2: Demonstration of Min protein oscillation in *E.coli.*. In each oscillation cycle MinE forms an E-ring at the rim of the old MinD polar zone, which dissociates MinC and MinD:ADP from the membrane. MinD:ADP undergoes nucleotide exchange into MinD:ATP in the cytoplasm before it rebinds to the membrane on the opposite cell pole, and meanwhile it recruits MinC and MinE to the new MinD polar zone. (This figure is modified from J. Lutkenhaus's paper [2].)

favorable for MinD:ATP binding (as proposed in some models), it would be of interest to see if this feature can be explained based solely on dynamics and interactions between Min proteins.

The model proposed by Huang *et al* [34] introduces MinE:MinD:ATP complexes formed by MinD and MinE, as well as a delay for nucleotide exchange and self-association of MinD on the membrane. It successfully reproduces Min oscillations and polar localization of MinD in wt, filamentous and spherical *E. coli* cells ($rodA^-$) by considering known interactions of MinD and MinE, without specific targeting for MinD at the cell poles. Based on this model Elf's research group [41] simulated both the stochastic and mean-field kinetics of the Min system with a software called MesoRD utilizing the reaction-diffusion master equations. Moreover, Kulkarni *et al* proposed a simplified 1D model [44] that was derived from Huang's 3D model and focused on some key processes, such as diffusion, a delay for nucleotide exchange and different rates of attachment to the bare membrane and occupied membrane, to analyze the observed length scale for MinD zone formation in filamentous cells. The significance of the follow-up work in this thesis is that it extends the previous work to the application of finite cells, which introduces simplification that makes it possible to obtain analytical results for the key processes involved, and it provides a novel approach for a better understanding of the physics insight behind the biological process.

3.2.2 Our model

In previous work [44], a simple model was presented to explain the length scale for formation of MinD zones in filamentous cells. The key elements were (i) differential attachment rates for MinD:ATP to the bare membrane and to the old polar zone and (ii) a delay associated with nucleotide exchange (i.e. conversion from MinD:ADP to MinD:ATP) for cytoplasmic MinD. Since MinD has a much higher rate of binding to the existing MinD zone, the boundary of the old polar zone was considered to be an absorbing boundary for cytoplasmic MinD:ATP. Following ATP hydrolysis by MinE, MinD:ADP is released from the old polar zone boundary, however in order to rebind to the membrane it has to undergo nucleotide exchange (assumed to occur with constant probability per unit time). In filamentous cells, where the location of the poles does not affect the dynamics, this leads to an exponentially decaying profile for the spatial location where MinD:ATP is *first* formed relative to the old polar zone boundary. However, the MinD:ATP molecules formed close to the old polar zone boundary are much more likely to reattach to the old polar zone than to bind to the bare membrane, thus the probability of attachment of MinD:ATP to the bare membrane increases with distance from the old polar zone boundary. These two effects combine to produce a finite length scale (from the old polar zone boundary) where the probability density attachment to the bare membrane by MinD:ATP is maximized. It was suggested that this location corresponds to the most probable location

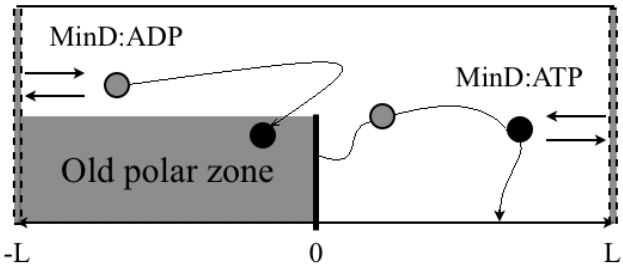


Figure 3.3: Schematic illustration of boundary conditions for the model: reflecting boundary conditions for MinD:ADP and one absorbing one reflecting boundary conditions for MinD:ATP. If a MinD:ATP molecule is converted from MinD:ADP in the region $-L \leq x \leq 0$, or if it diffuses across the boundary x=0, it attaches to the old polar zone with unit probability. The boundary of the old polar zone is thereby considered as an absorbing boundary for MinD:ATP in the region $0 < x \leq L$.

for formation of a new MinD zone thereby accounting for the length scale separating MinD zones in filamentous cells. To gain further insight, we extend this 1D semi-infinite model in the present work to wt cells for which it is critical to take into account the boundary conditions at the cellular poles. This is schematically illustrated in Figure 3.3 which shows the different boundary conditions for MinD:ADP and MinD:ATP. In the following, we use both analytical approaches and numerical simulations to explore cellular MinD distributions produced by the key processes and propose a mechanism for precise polar localization of MinD.

3.2.3 Simulation setup

We set up a 1D model considering discrete sites both in the cytoplasm and on the cell membrane. For a wt *E. coli* cell, we take the length to be 2 μm

[16] and use 100 lattice sites such that the spatial division is $\Delta x = 0.02\ \mu m$. We choose the cytoplasmic diffusion constant to be $D = 16\ \mu m^2 s^{-1}$ [45] for comparison with analytical results. The standard Gillespie algorithm is applied to simulate the processes of reaction and diffusion [8]. We consider $1.6 \times 10^{-5}\ s$ as the unit of time and one spatial division (i.e $0.02\ \mu m$) as the unit of length, correspondingly the diffusion rate between two adjacent sites for the stochastic simulations, d, is 0.64 (since $d = \frac{D}{(\Delta x)^2} = 4 \times 10^4\ s^{-1}$). For the above value of D, the time-scale for diffusion across half of the cell is $< t >= \frac{50^2}{2d} \sim 1950$ in these units. For the simulations, we set the mean nucleotide exchange time as $\tau_n = 6.25 \times 10^4$ which corresponds to $\tau_n = 1\ s$.

For the diffusion specifically we treat it as a multi-reaction process. Instead of tracking the location of each molecule, we track the number of molecules, the concentration of each compartment. Consider the above system to be divided into k compartments with reflecting boundaries. The reaction rate in this case is $d = D/h^2$, with D being the diffusion constant, h being the compartment length and A_i being the concentration in compartment i.

$$A_1 \underset{d}{\overset{d}{\rightleftarrows}} A_2 \underset{d}{\overset{d}{\rightleftarrows}} A_3 \underset{d}{\overset{d}{\rightleftarrows}} \cdots \underset{d}{\overset{d}{\rightleftarrows}} A_k$$

The simulation steps for the system with the Gillespie algorithm are

1. Generate random numbers r_1, r_2 from a uniform distribution in $[0, 1]$.

2. Compute the propensity functions

$$\alpha_i = A_i(t)\, d \qquad \text{and} \qquad \alpha_0 = \sum_{i=1}^{k-1} \alpha_i + \sum_{i=2}^{k} \alpha_i.$$

3. Calculate the time skip τ for the next action to be taken place as

$$\tau = \frac{1}{\alpha_0} \ln \left[\frac{1}{r_1} \right].$$

4. Use r_2 to find the specific reaction channel. We separate the events of molecule moving to the compartment on the right from those on the left.

   If $r_2 < \sum_{i=1}^{k-1} \alpha_i/\alpha_0$, find j such that

$$\frac{1}{\alpha_0} \sum_{i=1}^{j-1} \alpha_i \leq r_2 < \frac{1}{\alpha_0} \sum_{i=1}^{j} \alpha_i \qquad (1 \leq j \leq k-1).$$

   Move the molecule to the right

$$A_j(t+\tau) = A_j(t) - 1 \quad \text{and} \quad A_{j+1}(t+\tau) = A_{j+1}(t) + 1,$$

   else if $r_2 \geq \sum_{i=1}^{k-1} \alpha_i/\alpha_0$, find j such that

$$\frac{1}{\alpha_0} \left(\sum_{i=1}^{k-1} \alpha_i + \sum_{i=2}^{j-1} \alpha_i \right) \leq r_2 < \frac{1}{\alpha_0} \left(\sum_{i=1}^{k-1} \alpha_i + \sum_{i=2}^{j} \alpha_i \right) \qquad (2 \leq j \leq k),$$

   Move the molecule to the left

$$A_j(t+\tau) = A_j(t) - 1 \quad \text{and} \quad A_{j-1}(t+\tau) = A_{j-1}(t) + 1.$$

5. Continue with step 1 for $t + \tau$.

3.3 Results

3.3.1 MinD membrane attachment

We begin by considering the following problem: given a MinD molecule released from the old polar zone boundary ($x = 0$), what is the most probable location for reattachment to the bare membrane? As discussed in the previous section, this was analyzed for the case of a long filamentous cell (i.e. neglecting boundary effects) in earlier work [44] and we now extend the analysis to the case of a finite cell. In other words, we take boundary effects into account. We set up a 1D model with the initial condition that membrane-bound MinD:ADP has an existing polar zone in one half of the cell in the domain $-L \leq x \leq 0$ and the domain $0 < x \leq L$ corresponds to the bare membrane and the opposite pole. We take the end of the old MinD zone as an absorbing boundary for MinD:ATP as before and impose reflecting boundary conditions for the opposite pole (see Figure 3.3). Recent results have measured the cytoplasmic diffusion constant of MinD to be $D \sim 16\ \mu m^2 s^{-1}$ [45]. Correspondingly the time-scale for diffusion to cover half a cell length, i.e in the range $0 < x \leq L$, is $< t >= \frac{L^2}{2D} = \frac{1}{32}\ s$ which is significantly shorter than the mean nucleotide exchange time $\tau_n \sim 1\ s$. Consequently, the probability distribution for the cytoplasmic location of MinD:ATP when

it is *first* formed from MinD:ADP can be approximated as uniform. Note that this is in contrast to the previous analysis [44] for long filamentous cell which corresponded to the limit $\sqrt{2D\tau_n} << L$, whereas in the present case for wt cells, we are working in the limit $\sqrt{2D\tau_n} >> L$. The membrane diffusion coefficient of MinD is significantly lower (about two orders of magnitude smaller [45]) than that of the cytoplasmic MinD and we set it to zero.

Correspondingly, the equations governing the time-dependent probability densities for cytoplasmic $\rho_c(x,t)$ and membrane-bound $\rho_m(x,t)$ MinD:ATP are:

$$\frac{\partial \rho_c(x,t)}{\partial t} = D\frac{\partial^2 \rho_c(x,t)}{\partial x^2} - \frac{1}{\tau}\rho_c(x) \tag{3.1}$$

$$\frac{\partial \rho_m(x,t)}{\partial t} = \frac{1}{\tau}\rho_c(x,t) \tag{3.2}$$

where $1/\tau$ is the attachment rate of a MinD:ATP molecule to the bare membrane. After a sufficiently long time, MinD:ATP will bind either to the old zone or the bare membrane. Correspondingly, the MinD:ATP membrane attachment probability density is given by

$$\rho_m(x) = \frac{1}{\tau}\int_0^\infty \rho_c(x,t)dt \equiv \frac{1}{\tau}f(x)$$

In terms of $f(x)$, the equation to be solved is

$$\rho_c(x,0) = -D\frac{d^2f(x)}{dx^2} + \frac{1}{\tau}f(x) \tag{3.3}$$

For the given initial condition as uniform distribution for cytoplasmic location where MinD:ATP is first converted from MinD:ADP, and with reflecting boundary conditions at $x = L$ and absorbing boundary conditions at $x = 0$, we can achieve the solution using the standard methods as follows [46].

First we consider the scenario with the initial condition as one single MinD:ATP molecule at $x = x_0$: $\delta(x - x_0) = -D(d^2f/dx^2) + f(x)/\tau$, to solve for which we use the following four procedures.

1. For region $x < x_0$, we solve for $0 = -Df''_< + f_</\tau$ with absorbing boundary conditions $f_<|_{x=0} = 0$.

2. For region $x > x_0$, we solve for $0 = -Df''_> + f_>/\tau$ with reflecting boundary conditions $f'_>|_{x=L} = 0$.

3. We take into account the continuity of $f(x)$ at $x = x_0$, $f_<|_{x=x_0} = f_>|_{x=x_0}$.

4. We consider the discontinuity of $f'(x)$ at $x = x_0$, $1 = -Df'_>|_{x=x_0} + Df'_<|_{x=x_0}$

Once we derive the solution from the above we can integrate it over the entire membrane to obtain the final result of membrane attachment probability density distribution in the scenario of uniform initial distribution of MinD:ATP.

$$\begin{aligned}\rho_m(x) &= \frac{1}{\tau}\int_0^x f_> dx_0 + \int_x^L f_< dx_0 \\ &= \frac{1-\cosh\left(\frac{L-x}{\sqrt{D\tau}}\right)/\cosh\left(\frac{L}{\sqrt{D\tau}}\right)}{L-\sqrt{D\tau}\tanh\left(\frac{L}{\sqrt{D\tau}}\right)}\end{aligned} \tag{3.4}$$

where the denominator is the normalization factor. The corresponding profile is shown in Figure 3.4 which indicates that the maximum of the above probability distribution occurs at the opposite pole. If we assume that the maximum indicates the most probable site for formation of the new polar zone, then the above equation suggests that the new MinD zone is most likely to form at the opposite pole in agreement with experimental observations. Note that the above analysis holds in the limit $\sqrt{2D\tau_n} >> L$; as the cell length increases and transitions to the opposite limit $\sqrt{2D\tau_n} << L$ the maximum attachment probability will correspond to the length-scale derived in previous work [44] instead of the cell pole. Moreover, this is consistent with the simulation results from Huang *et al* [34], which show

when $L = 4\ \mu m, \sqrt{2D\tau_n} = 4\sqrt{2} > L,$ cells display pole to pole oscillation;
when $L = 10\ \mu m, \sqrt{2D\tau_n} = 4\sqrt{2} < L,$ cells form a new zone at mid-cell.

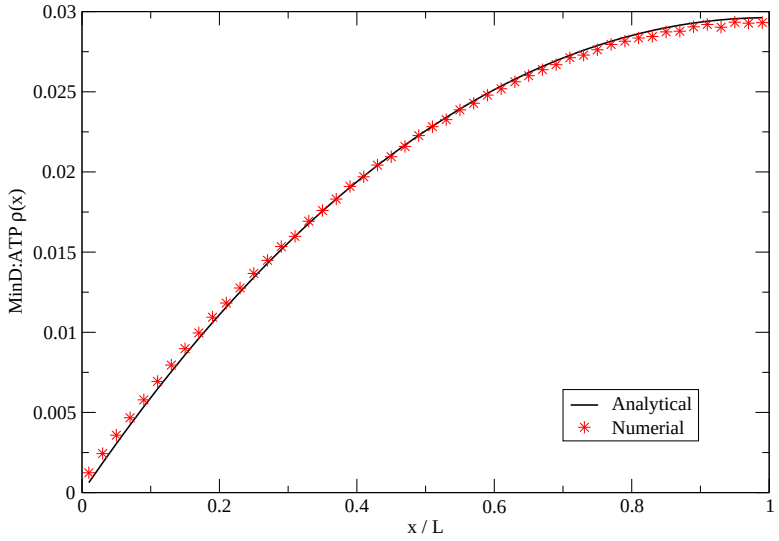


Figure 3.4: Membrane-attachment probability density for MinD:ATP as a function of position. Results from the analytical expression and simulations are shown(corresponding to mean attachment time $\tau = 5000$ units for the simulation). The cell length is taken to be 2 μm.

3.3.2 Filamentous cell attachment profile

When FtsZ is depleted from the cytoplasm, the cell stops dividing but keeps growing into a filamentous cell [14]. Experimental observations reveal that the *E. coli* filamentous cells also display Min protein oscillations. Stripe-patterned MinD zones are observed to switch alternatively between occupied and bare membrane sites. An important observation is that the new MinD attachment zones always relocate between the two existing old zones [16]. We considered the 1D model in the previous section and apply the same equations (3.1) and (3.2) with different boundary conditions: specifically two absorbing boundaries corresponding to the two MinD zone edges at $x = 0$ and $x = 2L$. As before, the initial condition corresponds to uniform MinD:ATP

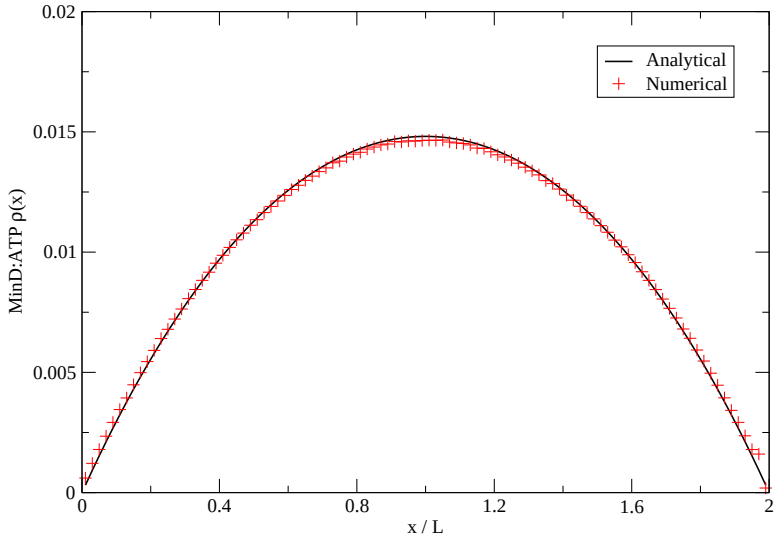


Figure 3.5: Same as Figure 3.4, but for the filamentous cell with two absorbing boundaries (see text). The separation between the two MinD zones is taken to be 4 μm.

cytoplasmic distribution. Following the above four procedures, we solve for the case with the initial condition as a single MinD:ATP resource with two absorbing boundary conditions $f_<|_{x=0} = f_>|_{x=2L} = 0$, and then we integrate the result over the whole membrane region $0 \leq x \leq 2L$ for the uniform initial distribution. The corresponding solution is

$$\rho_m(x) = \frac{1 - \cosh\left(\frac{L-x}{\sqrt{D\tau}}\right) / \cosh\left(\frac{L}{\sqrt{D\tau}}\right)}{2L - 2\sqrt{D\tau}\tanh\left(\frac{L}{\sqrt{D\tau}}\right)} \tag{3.5}$$

From Figure 3.5, the maximum of the MinD:ATP membrane-attachment probability distribution is found to be exactly in the middle of the two boundaries as expected. Thus, the formation of a new zone is most probable at the midpoint of the two old zones which again is consistent with the observed switching of MinD zones in filamentous cells. However, note that as the sep-

aration between the two MinD zones is reduced, the probability P_b that the MinD molecule is reabsorbed at either of the two old zones increases. Within our model, this probability can be readily evaluated as

$$P_b = \frac{\sqrt{D\tau}}{2L} \tanh \frac{L}{\sqrt{D\tau}} \tag{3.6}$$

Thus, although the probability of reattachment to the bare membrane remains maximal at the midpoint, the overall probability that MinD attaches to the bare membrane decreases with decreasing separation between the old MinD zones. If, as discussed in the next section, a minimal cellular concentration of MinD:ATP is required for formation of a new zone, then for sufficiently small separations between the two old MinD zones, formation of a new zone will not take place at the midpoint between the zones.

3.3.3 Time-dependent cytoplasmic density distribution

The previous section considered attachment of MinD monomers to the bare membrane. However, recent experiments show that membrane-associated MinD is polymerized into a helix structure around the cell membrane instead of randomly distributed [47, 48], which indicates that a nucleation event is involved in initiation of a new MinD zone. In particular, we expect the detachment rate for a monomer to be higher than the corresponding concentration-dependent attachment rate until a critical nucleus of monomers is formed. As

discussed in the next section, the mean nucleation time depends strongly on the local cytoplasmic MinD:ATP concentration, therefore it will be helpful to explore the time evolution of the cytoplasmic MinD density distribution. As MinD molecules detach from the polar zone, the cytoplasmic concentration of MinD:ATP increases with time; however at low enough concentrations a nucleation event is very unlikely. To analyze how cellular MinD concentrations evolve in this limit, we ignore membrane attachment. We denote the cytoplasmic probability density of MinD:ADP by $\rho_D(x)$ and that of MinD:ATP by $\rho_T(x)$. Since MinD:ADP cannot bind to the old polar zone the boundary conditions are such that

- MinD:ADP has reflecting boundaries at $x = -L$ and $x = L$, $\frac{\partial \rho_D}{\partial x} \mid_{x=-L} = \frac{\partial \rho_D}{\partial x} \mid_{x=L} = 0$,
- MinD:ATP has a absorbing boundary at $x = 0$, $\rho_T \mid_{x=0} = 0$, and a reflecting boundary at $x = L$, $\frac{\partial \rho_T}{\partial x} \mid_{x=L} = 0$.

The corresponding equations are

$$\frac{\partial \rho_T}{\partial t} = D\frac{\partial^2 \rho_T}{\partial x^2} + \frac{1}{\tau_n}\rho_D \tag{3.7}$$

$$\frac{\partial \rho_D}{\partial t} = D\frac{\partial^2 \rho_D}{\partial x^2} - \frac{1}{\tau_n}\rho_D \tag{3.8}$$

where $1/\tau_n$ is the nucleotide exchange rate in the cytoplasm. The initial condition is that a single MinD:ADP molecule is released at the end of the

old polar zone (i.e. at $x = 0$). To solve the above set of equations we use Laplace transforms $u(x,s) = \int_0^\infty e^{-st}\,\rho(x,t)dt$

$$\begin{aligned}
\rho_T(x,0) &= -D\frac{\partial^2 u_T(x,s)}{\partial x^2} + (s + \frac{1}{\tau_n})u_D(x,s) = 0 \\
\rho_D(x,0) &= -D\frac{\partial^2 u_D(x,s)}{\partial x^2} + (s - \frac{1}{\tau_n})u_D(x,s) = \delta(x - x_0)
\end{aligned}$$

We take the standard approach in the previous section to solve the Laplace transform u_D and u_T.

For $x < 0$,

$$u_D = \frac{1}{2D\alpha}\frac{\cosh[\alpha(L+x)]}{\sinh(\alpha L)} \qquad \text{and} \qquad u_T = 0$$

For $x > 0$,

$$\begin{aligned}
u_D &= \frac{1}{2D\alpha}\frac{\cosh[\alpha(L-x)]}{\sinh(\alpha L)} \\
u_T &= \frac{1}{2D\alpha}\frac{\cosh[\beta(L-x)]\cosh(\alpha L) - \cosh[\alpha(L-x)]\cosh(\beta L)}{\sinh(\alpha L)\cosh(\beta L)}
\end{aligned}$$

where $\alpha = \sqrt{\frac{s}{D} + \frac{1}{D\tau_n}}$, $\beta = \sqrt{\frac{s}{D}}L$. When the Laplace transform is inverted, the corresponding solutions for the time-dependent MinD:ATP and MinD:ADP cytoplasmic probability distribution are

$$\rho_D(x,t) = \frac{1}{2L}e^{-\frac{t}{\tau_n}} + \sum_{n\neq 0}\frac{1}{L}\cos(\lambda_n x)e^{-D\lambda_n^2 t-\frac{t}{\tau_n}} \tag{3.9}$$

$$\begin{aligned}\rho_T(x,t) &= \frac{1}{2L}e^{-\frac{t}{\tau_n}}\left(\frac{\cos\frac{L-x}{\sqrt{D\tau_n}}}{\cos\frac{L}{\sqrt{D\tau_n}}}-1\right) \\ &+ \sum_{n\neq 0}\frac{1}{L}e^{-D\lambda_n^2 t-\frac{t}{\tau_n}}\left(\frac{\cos\left[\sqrt{\lambda_n^2+\frac{1}{D\tau_n}}(L-x)\right]}{\cos\left[\sqrt{\lambda_n^2+\frac{1}{D\tau_n}}L\right]}-\cos(\lambda_n x)\right) \\ &+ \sum\nolimits_{\mu_n^2<\frac{1}{D\tau_n}}\frac{1}{L}\mu_n\sin(\mu_n x)\coth\left(\sqrt{\frac{1}{D\tau_n}-\mu_n^2}L\right)\frac{e^{-D\mu_n^2 t}}{\sqrt{\frac{1}{D\tau_n}-\mu_n^2}} \\ &- \sum\nolimits_{\mu_n^2>\frac{1}{D\tau_n}}\frac{1}{L}\mu_n\sin(\mu_n x)\cot\left(\sqrt{\mu_n^2-\frac{1}{D\tau_n}}L\right)\frac{e^{-D\mu_n^2 t}}{\sqrt{\mu_n^2-\frac{1}{D\tau_n}}}\end{aligned} \tag{3.10}$$

where $\lambda_n = \frac{n\pi}{L}$, $\mu_n = \frac{n\pi+\pi/2}{L}$ and n is an integer.

We plot the cytoplasmic probability density for both MinD:ADP and MinD:ATP at different times in Figure 3.6 and Figure 3.7, with both the analytical results (solid lines) and simulations. We notice that MinD:ADPs reach a uniform cytoplasmic density distribution after diffusion across the cell. Interestingly, after an initial transient, the cytoplasmic probability density of MinD:ATP has a maximum at the bare pole. As discussed in the next section, this is consistent with a nucleation based mechanism for polar localization.

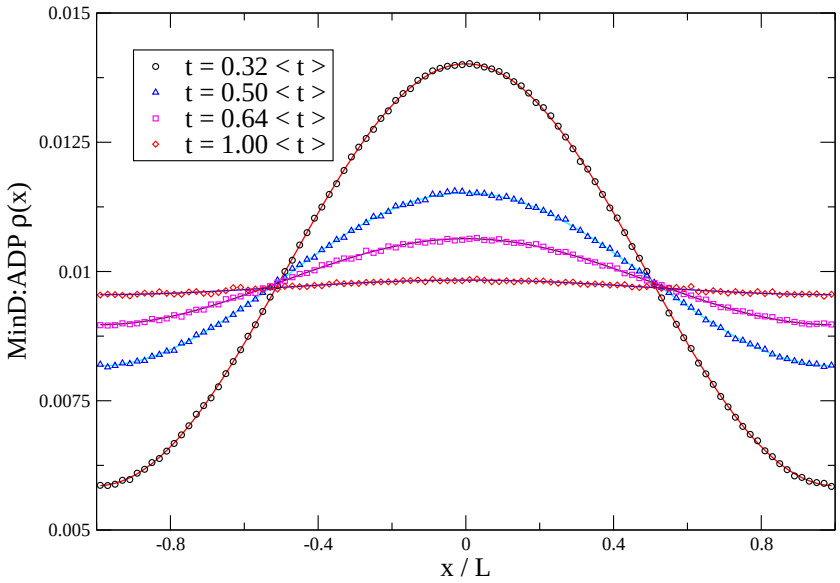


Figure 3.6: MinD:ADP cytoplasmic density distribution from analytical results and simulations. The mean nucleotide exchange time is $\tau_n = 6.25 \times 10^4$ units. The time-scale $< t >= \frac{50^2}{2d} \sim 1950$ units (see text), which roughly corresponds to the time to diffuse one half of the cell for $D = 16\ \mu m^2 s^{-1}$. The cell length is taken to be 2 μm.

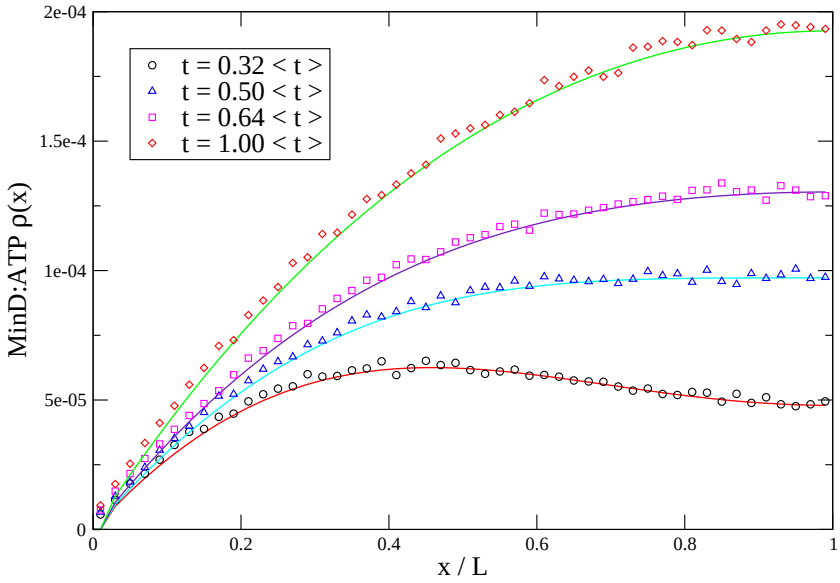


Figure 3.7: Same as Figure 3.6, but for MinD:ATP cytoplasmic density distribution (for the region $0 < x \leq L$).

3.3.4 Mean nucleation time

Nucleation is generally defined as the onset of a phase transition in a small region. In molecular biology, it is used to term the critical stage in the assembly of a polymeric structure. We model the nucleation process by considering only one site on the membrane for MinD attachment and detachment. The local cytoplasmic density of MinD (ρ_c) is taken to be constant, thus the attachment rate is $\rho_c k_a$. We denote the MinD monomer dissociation rate by k_d. Now we have a stochastic process for growth of the MinD cluster size on the membrane varying from zero monomers to the critical nucleus.

$$0 \underset{k_d}{\overset{\rho_c k_a}{\rightleftharpoons}} 1 \underset{k_d}{\overset{\rho_c k_a}{\rightleftharpoons}} 2 \underset{k_d}{\overset{\rho_c k_a}{\rightleftharpoons}} \cdots \underset{k_d}{\overset{\rho_c k_a}{\rightleftharpoons}} N-1 \xrightarrow{\rho_c k_a} N \tag{3.11}$$

We schematically illustrate this process in Eq. (3.11) with the lower bound as a reflecting boundary at $n = 0$ and the upper bound as an absorbing boundary at $n = N$. The mean nucleation time can then be calculated by evaluating the mean first passage time T_{fp} to reach the boundary $n = N$. The mean first passage time for the corresponding process with one reflecting boundary and one absorbing boundary has been discussed in Chapter 2. If the process starts from the reflecting boundary $x = a$ at $t = t_0$, the mean time it takes to reach the absorbing boundary $x = b$ and exit the system is

$$T_{fp} = \sum_{y=0}^{b} \phi(y) \sum_{z=a}^{y} \frac{1}{[\lambda(z)\phi(z)]}, \quad \phi(x) = \prod_{z=a+1}^{x} \frac{\mu(z)}{\lambda(z)}$$

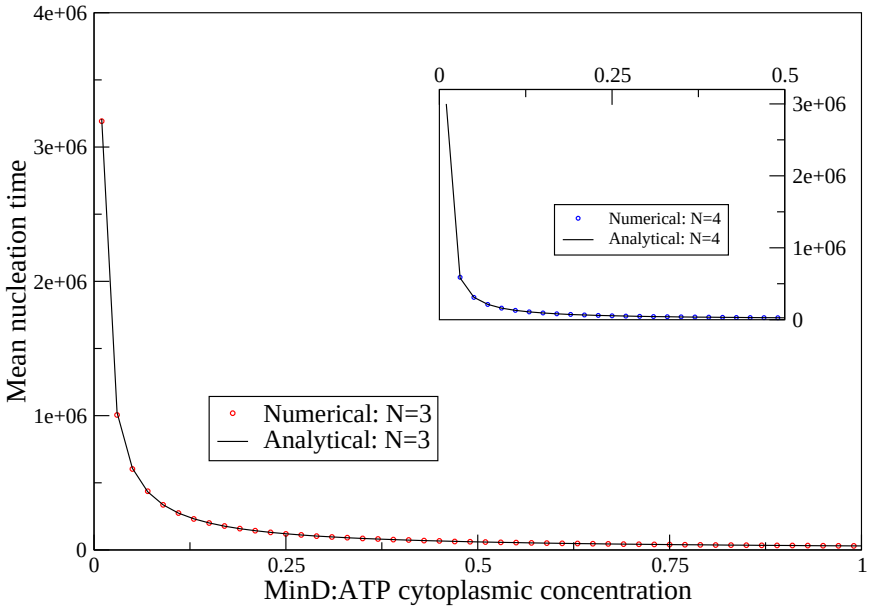


Figure 3.8: Mean nucleation time (T_{fp}) vs local MinD:ATP density (ρ_c) for critical nucleus size $N = 3$ and $N = 4$ (inset). Different ratios of attachment rate to detachment rate k_a/k_d were used for the two cases: $k_a = 10^{-4}\ s^{-1}$ and $k_d = 10^{-7}\ s^{-1}$ for $N = 3$ case, $k_a = 3 \times 10^{-4}\ s^{-1}$ and $k_d = 3 \times 10^{-6}\ s^{-1}$ for $N = 4$ case (ρ_c=0.01). T_{fp} is in units as previously mentioned ($1.6 \times 10^{-5}\ s$ per unit). The solid line represents the analytical results in Eq. (3.12) and circles are the results from the simulations.

We take all the forward transition rates the same as $\lambda(x_0) = \rho_c k_a$, and all the backward transition rates the same as $\mu(x_0) = k_d$, and the reflecting and absorbing boundaries as $a = 0$ and $b = N$ respectively. In such case the mean first passage time or the mean nucleation time can be calculated explicitly (Figure 3.8).

$$T_{fp} = \frac{1}{\rho_c k_a} \sum_{k=0}^{N-1} (N-k)\alpha^k, \qquad \alpha = \frac{k_d}{(\rho_c k_a)} \tag{3.12}$$

3.3.5 Nucleation time probability distribution

We have introduced in the previous section that we model the nucleation process as an 1D random-walk-like stochastic process with a reflecting lower boundary at $n = 0$ and an absorbing upper boundary at $n = N$. Therefore the probability distribution of the nucleation time can be derived by the first passage time of the random walker. Now we consider the case in which the critical nucleus size is set to be $N = 3$, with the same scenario as before that forward rate is the attachment rate $\rho_c k_a$ and the backward rate is the detachment rate k_d. The scheme is shown as follows

$$0 \underset{k_d}{\overset{\rho_c k_a}{\rightleftharpoons}} 1 \underset{k_d}{\overset{\rho_c k_a}{\rightleftharpoons}} 2 \xrightarrow{\rho_c k_a} 3 \tag{3.13}$$

According to the methods in Chapter 2, the first passage time probability distribution density function, or the nucleation time probability distribution $F_{3,m}(t)$ follows the backward master equations, which have been solved in the generalized four-state case in Appendix B or a more specific case Appendix C. For the scenario that the site stars with bare membrane $m = 0$ we have

$$F_{3,0}(t) = \frac{(\rho_c k_a)^3 e^{s_1 t}}{(s_2 - s_1)(s_3 - s_1)} + \frac{(\rho_c k_a)^3 e^{s_2 t}}{(s_1 - s_2)(s_3 - s_2)} + \frac{(\rho_c k_a)^3 e^{s_3 t}}{(s_2 - s_3)(s_1 - s_3)}$$

in which s_1, s_2 and s_3 are real numbers solved as

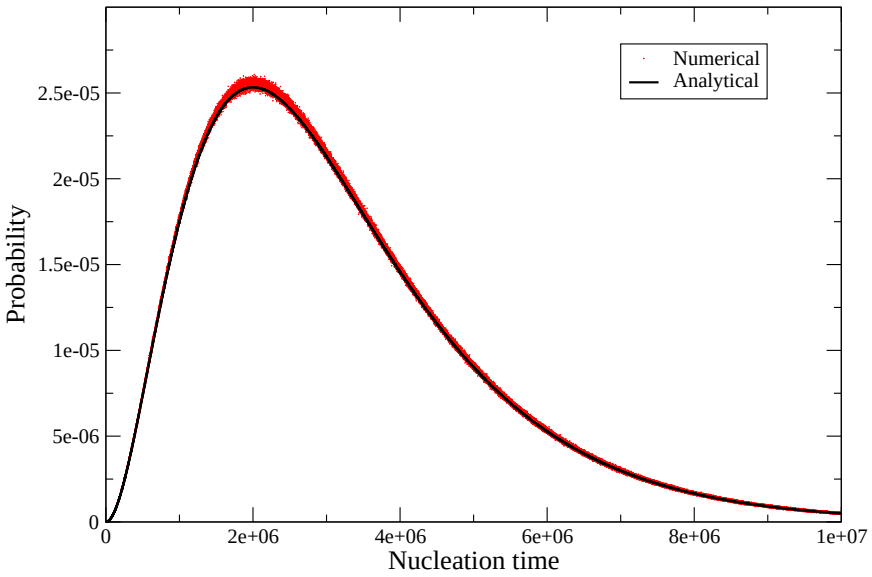


Figure 3.9: Nucleation time probability distribution for a bare membrane site to reach a nucleus sized of 3 monomers. The attachment rate $\rho_c k_a = 10^{-6}\ s^{-1}$ and the dissociation rate $k_d = 10^{-7}\ s^{-1}$. Nucleation time is in units as $1.6 \times 10^{-5}\ s$ per unit. The solid line represents the analytical results and crosses are the results from the simulations.

$$\begin{cases} s_1 = 2\sqrt[3]{r}\cos(\frac{1}{3}\theta) - (\rho k_a + \frac{2}{3}k_d) \\ s_2 = 2\sqrt[3]{r}\cos(\frac{\theta}{3} + \frac{2\pi}{3}) - (\rho k_a + \frac{2}{3}k_d) \\ s_3 = 2\sqrt[3]{r}\cos(\frac{\theta}{3} - \frac{2\pi}{3}) - (\rho k_a + \frac{2}{3}k_d) \end{cases}$$

where

$$r = \sqrt{(\frac{2}{3}\rho k_a k_d + \frac{1}{9}k_d^2)^3} \quad \text{and} \quad \theta = \arccos\left(\frac{-\frac{1}{6}\rho k_a k_d^2 + \frac{1}{27}k_d^3}{r}\right)$$

3.4 Discussion

In summary, our initial analysis focused on identifying the most likely location for MinD:ATP membrane binding assuming that this corresponds to the initiation point for the new MinD zone. This is plausible given the observation that MinD polymerizes on the membrane indicating that initiation of a new MinD zone involves a nucleation event. It should be noted though that experimental data indicates that there is a preexisting backbone of MinD throughout the oscillation cycle [48], which could potentially negate the need for nucleation. Nevertheless, we feel it is of interest to explore the potential role of nucleation in precise polar targeting of MinD. For nucleation to occur multiple MinD monomers (up to a critical nucleus size N) must attach at the same location before the monomers detach. This is more likely to happen at the site where MinD monomers have the highest probability of attachment to the bare membrane. Furthermore the requirement for nucleation also suggests an explanation for how cells reproducibly start the new zone at the same location (the bare pole) despite the inherent stochasticity in the process as argued below.

The basic observation is that the mean first-passage time for a nucleation event has a power-law dependence on the local concentration of MinD:ATP. For cellular regions with low enough concentration of MinD:ATP, the mean nucleation time far exceeds the cell division time; thus only those regions of the cell which have a high enough concentration of MinD:ATP will give rise to

a nucleation event. At the beginning of an oscillation period, the cytoplasmic concentration of MinD:ATP is negligible. As MinD is released from the old polar zone and undergoes nucleotide exchange the cytoplasmic concentration of MinD:ATP starts to increase. As shown in the previous section, the basic processes give rise to a concentration profile which (after initial transients) has a maximum at the opposite/bare pole. Thus the dynamics ensures that as the cytoplasmic MinD:ATP concentration increases, it is always maximal at the bare pole whereas the sharp concentration dependence (due to the requirement for nucleation) ensures that the time-scale for nucleation events to occur elsewhere in the cell is much higher than that at the pole. Correspondingly, initiation of a new MinD zone is most probable at the bare pole and unlikely elsewhere. Thus the combination of the basic processes explored using simplified models in this study can potentially explain reproducible polar localization of MinD in *E. coli*.

Finally, we note that the essential insights from our analysis have potential applications beyond explaining MinD zone formation in *E. coli* cell division. We propose that the combination of processes analyzed can potentially explain several other cases of precise protein localization in cellular processes. For example, in the developing *Drosophila* embryo the hunchback (hb) gene is expressed within a specific domain, which is believed to be related to the earlier concentration profile of the protein bicoid (bcd) across the embryo [49]. The general analysis carried out in our modeling shows how, with two interacting proteins, it is possible to generate cytoplasmic concentration gra-

dients of one of the components, with the maximum value at the subcellular localization of interest such as the poles. When such concentration gradients are coupled with a requirement for nucleation, precise protein localization can be achieved despite the intrinsic stochasticity of the underlying biochemical processes. It will be of interest to consider the application of these ideas to model cellular processes for which protein localization plays a critical role.

Chapter 4

DNA transcriptional pausing

Transcriptional pausing has been intensively studied for three decades and yet the underlying physical processes are still unclear. However, the recent advances in single-molecule techniques provide us a boost of experimental data and possibilities of new perspectives and interpretations upon the issue. It has been discovered that transcriptional pauses are rate-limiting stochastic events. Based on recent data, we propose a quantitative model to understand how the pausing duration changes with different entry and escape rates of the pauses. We obtain analytical solutions for our simplified stochastic model which can provide useful insights and inputs to current and future experiments focusing on time-resolved single-molecule studies of transcriptional pausing.

4.1 Introduction

DNA transcription, also called mRNA synthesis, is a highly regulated process in gene expression. During transcription, a DNA sequence is read by RNA polymerase (RNAP) to produce a copy of complementary antiparallel RNA strand. A DNA transcription unit contains not only the coding sequence, the sequence that will be translated into the protein, but also regulatory sequences that direct and regulate the synthesis of that protein. The regulatory sequence upstream from the coding sequence is called the five prime (5') untranslated region and the sequence downstream from the coding sequence is called the three prime (3') untranslated region. Individual nucleotides are added to the 3' end of the growing RNA chain during elongation, one of the three main stages of transcription (initiation, elongation and termination), which involves RNAP translocation, nucleotide triphosphate (NTP) binding, nucleotide condensation, and the release of inorganic pyrophosphate (PPi).

4.1.1 What is transcriptional pausing

The transcriptional elongation phase has been detected to be frequently interrupted by pausing both in bacteria [3, 50, 51, 52] and in eukaryotes [53, 54, 55]. A pause is characterized by two empirical parameters: the efficiency and the duration. It has been identified that a majority of pauses have a duration of approximately $1 \sim 6$ s on average [3, 50, 56], while others have a longer

duration of 25 s and larger. The "ubiquitous" short-lived pauses have been proposed to be sequence-dependent elemental pauses acting as a common intermediate state preceding the long-lived stabilized pauses [52, 56, 57, 58]. Most of these lengthier stabilized pauses are associated with the formation of hairpin structures in the nascent RNA transcript [57, 59, 60], or that of the RNA:DNA hybrids which are thought to induce enzyme backtracking [3, 61, 62].

Hairpin-stabilized pausing

Hairpin-stabilized pauses are one class of defined long-lived pauses that occur when self-complementary hairpin-like structures form at the exit channel of RNA transcript and allosterically inactivate RNAP. One representative example is the *his* pause in *E. coli*.

Backtrack pausing

A second class of defined stabilized pauses is backtracking pauses whose term comes from the rearward motion of RNAP along the DNA and RNA chains during the pausing. When RNAP encounters a weak RNA:DNA hybrid, it moves upstream on the DNA template to a more energetically stable position, which extrudes the nascent RNA from the NTP entry channel. A backtracked transcription elongation complex (TEC) is catalytically incompetent because the 3' end of the RNA is displaced from the active site. Backtracking pauses

occur both in prokaryotes and eukaryotes. One example is the *ops* pause in *E. coli*.

4.1.2 Role in gene regulation

Transcriptional pausing has been shown to have diverse biological roles in gene regulation. It can not only repress mRNA production and modify subsequent transcription by recruiting regulatory factors to the TEC [63, 64, 65], but also act as a precursor to transcriptional arrest and termination [66]. Besides, pausing is thought to synchronize transcription with translation to coordinate gene expression in prokaryotes, as well as leading to messenger splicing or polyadenylation in eukaryotes. In *Drosophila* and mammalian cells, RNA polymerase II (Pol II) stalls in the promoter proximal region, which limits the overall rate of transcription, thereby coupling transcription with translation and preventing premature rho-dependent termination. This behavior is believed to facilitate a faster response to environmental stimuli than regulation at initiation alone, allowing rapid changes in developmental gene expression during embryogenesis [53, 54, 55, 67, 68, 69]. Other studies suggest that pausing may allow time for the mRNA processing machinery to correctly assemble and cap nascent RNAs [53]. Therefore, pausing is an important step in the regulation of elongation and a checkpoint before committing to productive elongation.

4.1.3 Recent discovery of Pol II stalling

Studies indicate that Pol II is concentrated in the promoter regions of thousands of genes in human and *Drosophila* cells which is believed to diminish cell-to-cell variation in the onset of transcription, thus allowing synchronous rather than stochastic expression of genes induced during development. Such transcriptional synchrony in higher eukaryotes ensures that morphogenetic changes in an organism such as cell division and migration occur in a coordinated fashion [70]. Intense research has been focused on the regulatory role of Pol II pausing upon this rapid and specific activation of gene expression in response to environmental stimuli, but the mechanism that leads to the release of paused Pol II had remained a mystery for years [53, 71, 72]. However, a very recent study in mouse embryonic stem cells demonstrates that the cellular proto-oncogene c-Myc regulates a large subset of its target genes by relieving polymerase pausing, a finding alluding to the importance of this mechanism and its relevance in the onset of cancer [73]. In this research, c-Myc was determined to bind to nearby DNA and recruit pause-release factors such as P-TEFb (positive transcription elongation factor b) to genes where Pol II pausing was detected.

These latest discoveries have provided further motivation for investigating the mechanisms governing pause releasing as well as pause duration. Analysis of the corresponding coarse-grained stochastic models can guide future single-molecule experiments aimed at determining the statistics of pause durations.

In the following, we develop such a stochastic model and derive exact results for it which are validated by stochastic simulations.

4.2 Model

Recent developments in high-resolution measurements, i.e. single-molecule fluorescence and optical-trapping methods, have advanced our knowledge of DNA transcriptional pausing phenomena with direct observations [58, 74, 75]. Based on these, we propose a stochastic model that captures the essence of the pausing phenomena and brings a new quantitative perspective of the issue.

4.2.1 Other models

Several stochastic models have been proposed to study the mechanism of the pausing kinetics at a level of single nucleotide [76, 77, 78, 79] as well as the entire TEC [80]. However, the dependence of the pause duration upon different kinetic rates and elongation velocity has not yet been determined analytically. The distribution of pausing duration has been previously fit by a sum of double-exponentials arising from many single-exponential individual pauses [3, 52, 51], or by a power law ($\sim t^{-3/2}$) for pauses characterized by long lifetime ($> 10\ s$) [81]. In this chapter, we explore a novel fit for the distribution of pausing duration using a simplified underlying model based

on recent experimental data.

4.2.2 Our model

The previous section has mentioned that the short reversible elemental pauses will lead to the lengthier stabilized pauses or possibly to irreversible transcriptional arrest, due to one or more subsequent rearrangement or interactions of the enzyme and the active site. As illustrated in Figure 4.1, when the TEC is at an active non-pausing state, it may either get back on the elongation pathway or enter the elemental pausing state and proceed to a hairpin-stabilized or backtracking pause. Therefore, we present a quantitative model to describe this four-state scenario and we investigate how long it takes to enter and escape the pausing states until the TEC irreversibly resumes the elongation process.

We consider the state change of the TEC as a Markov process as in some other models [80, 82] but we derive analytical expressions for the pause duration from the first passage time. When the TEC system is either in the stabilized pausing state or in the elemental pausing state, it may escape from the pauses into the active non-pausing state or re-enter either one of the pausing state. Once the system reaches the elongation state, it is irreversible, hence we set it as an absorbing boundary. For simplification we denote the above scenario with numbers (see below). We study how long it takes for the TEC to escape the pausing state and resume the elongation process and how the

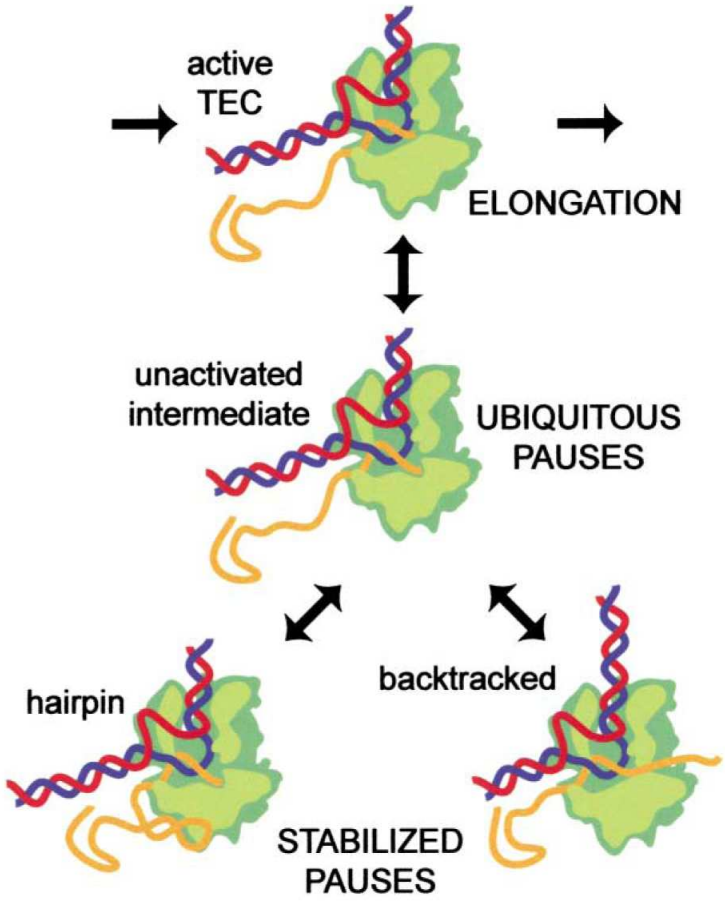


Figure 4.1: Transcriptional elongation pathway (horizontal) and a subset of off-pathway states (vertical transitions). The TEC in the elongation phase of transcription (top) may enter an intermediate state known as elemental pauses (middle), which precedes entry into stabilized pauses associated with either RNA hairpin formation or enzyme backtracking (bottom). (This figure is modified from K. C. Neuman's paper [3].)

probability distribution of the pausing duration changes with various reaction rates. In such case, as we take the elongation state as an absorbing boundary the duration from the pausing state to the elongation state can be therefore considered to be the first passage time.

4.2.3 The mathematical expression

$$\mathrm{P}_{stb} \underset{d_1}{\overset{a_0}{\rightleftharpoons}} \mathrm{P}_{elem} \underset{d_2}{\overset{a_1}{\rightleftharpoons}} \text{Active} \xrightarrow{a_2} \text{Elongation}$$

or

$$0 \underset{d_1}{\overset{a_0}{\rightleftharpoons}} 1 \underset{d_2}{\overset{a_1}{\rightleftharpoons}} 2 \xrightarrow{a_2} 3$$

For the process above, we use $F_{3,m}(t)$ to denote the probability density function of the first passage time when the system starts from either pausing states 0, 1 or the active state 2. According to the discussion in Chapter 2 we can find the probability distribution of the pausing duration as

$$
\begin{cases}
F_{3,0}(t) = \dfrac{a_0a_1a_2e^{s_1t}}{(s_2-s_1)(s_3-s_1)} + \dfrac{a_0a_1a_2e^{s_2t}}{(s_1-s_2)(s_3-s_2)} + \dfrac{a_0a_1a_2e^{s_3t}}{(s_2-s_3)(s_1-s_3)} \\
F_{3,1}(t) = \dfrac{(a_0a_1a_2+a_1a_2s_1)e^{s_1t}}{(s_2-s_1)(s_3-s_1)} + \dfrac{(a_0a_1a_2+a_1a_2s_2)e^{s_2t}}{(s_1-s_2)(s_3-s_2)} \\
\qquad\qquad + \dfrac{(a_0a_1a_2+a_1a_2s_3)e^{s_3t}}{(s_2-s_3)(s_1-s_3)} \\
F_{3,2}(t) = \dfrac{\left[a_0a_1a_2+a_2(a_0+a_1+d_1)s_1+a_2s_1^2\right]e^{s_1t}}{(s_2-s_1)(s_3-s_1)} \\
\qquad\qquad + \dfrac{\left[a_0a_1a_2+a_2(a_0+a_1+d_1)s_2+a_2s_2^2\right]e^{s_2t}}{(s_1-s_2)(s_3-s_2)} \\
\qquad\qquad + \dfrac{\left[a_0a_1a_2+a_2(a_0+a_1+d_1)s_3+a_2s_3^2\right]e^{s_3t}}{(s_2-s_3)(s_1-s_3)}
\end{cases}
$$

where a_0, a_1 are the escape rates from the pausing states, and d_1, d_2 are the entry rates into the pausing states, and a_2 is the rate at which the system resumes elongation. s_1, s_2 and s_3 are variables depending on these reaction rates.

$$
\begin{cases}
s_1 = 2\sqrt[3]{r}\cos(\frac{1}{3}\theta) - A/3 \\
s_2 = 2\sqrt[3]{r}\cos(\frac{\theta}{3} + \frac{2\pi}{3}) - A/3 \\
s_3 = 2\sqrt[3]{r}\cos(\frac{\theta}{3} - \frac{2\pi}{3}) - A/3
\end{cases}
$$

with the constants

$$r = \sqrt{-\frac{p^3}{27}} \quad \text{and} \quad \theta = \arccos\left(-\frac{q}{2r}\right)$$

where $p = B - A^2/3$, $q = C + (2A^3 - 9AB)/27$, and $A = a_0 + a_1 + a_2 + d_1 + d_2$, $B = a_0a_1 + a_1a_2 + a_2a_0 + a_2d_1 + a_0d_2 + d_1d_2$, $C = a_0a_1a_2$. The more detailed calculations to find the first passage time probability distribution of this four-state case are presented in Appendix B.

4.2.4 Validation with simulations

As mentioned previously, the pauses differ in detail from gene to gene and their characteristic parameters, i.e. pause efficiency and pausing duration, are sequence-dependent [3]. We now explore how this can be related to the underlying kinetic parameters using our model. First we determine key parameters based on experimental results for the elemental pauses (which are also termed ubiquitous pauses) with the mean pausing duration $\sim 3\ s$ [3]. In this case, a_0 and d_1 are set to be 0 while the elongation rate a_2 is chosen between $5 \sim 15\ bp\ s^{-1}$ (base-pair per second) under the conditions of 1mM NTPs concentration in single molecule experiments [74, 83, 84, 85]. It has been determined in single-molecule data that the distance between consecutive pauses is approximately 100 bp [50, 52]. That is to say that pauses occur at a roughly constant density of 1 pause per 100 bp, or the probability of pausing occurrence is 1/100, which leads to the relation $d_2 : (d_2 + a_2) = 1 : 100$.

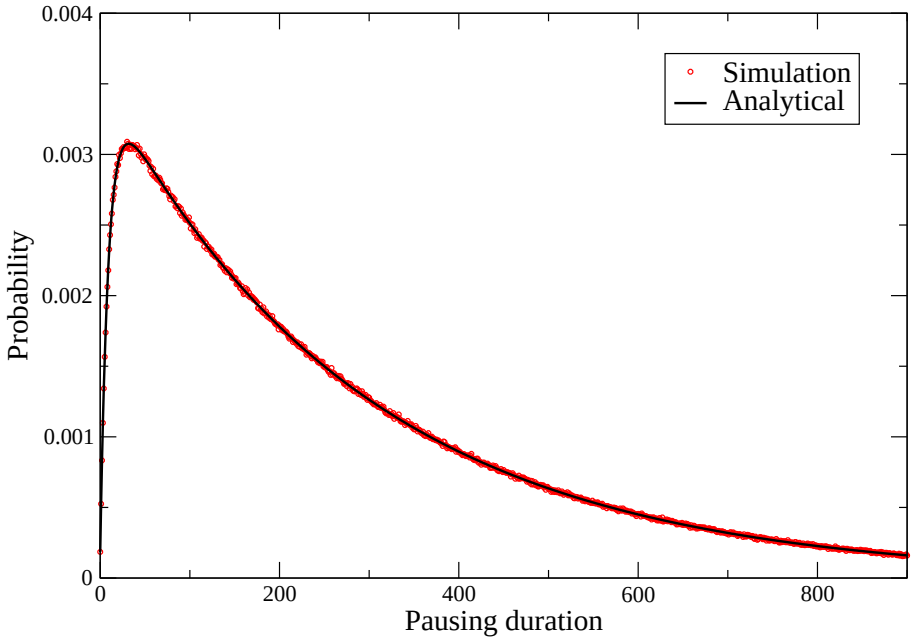


Figure 4.2: Probability distribution of the elemental pausing duration. We select the rate $a_1 = 0.00347\ s^{-1}$. The probability distribution has mean $= 3.00018\ s^{-1}$, std $= 2.9116\ s^{-1}$. The pausing duration is in the units of $0.01\ s$.

Therefore we find $d_2 = 1/9\ s^{-1}$ when $a_2 = 11\ s^{-1}$ (based on available experimental data [3]). We then select the appropriate pausing escape rate a_1 so that the mean pausing duration is in agreement with the data in experiments [3, 50, 56]. As is shown in Figure 4.2, the probability distribution of the elemental pausing duration is shaped like single-exponential distribution and the majority (96.69%) of the distribution is within 10 s, more precisely about 60.10% of the distribution is between $1 \sim 6\ s$ for the parameter choices, consistent with experimental observations. We derive the probability distribution of the pausing duration under different reasonable choices of the reaction rates. On the other hand, we have also matched our analytical results with the stochastic simulations using the Gillespie algorithm [8].

We have explored the quantitative model further with stabilized pauses. Since the ubiquitous pauses are believed not to be a consequence of RNA hairpin [51] or DNA backtracking [3, 62], we fix the previous rates at which the system enters into and escapes from the elemental pausing state and the elongation rate ($d_2,\ a_1,\ a_2$), while we investigate how the duration of the stabilized pause is affected by varying its entry rate d_1 and escape rate a_0. The detailed mechanisms of transitions to and from the stabilized pause to the elemental pause state are still unclear in experimental results [51]. Our results provide a perspective of viewing the issue by looking into the relationship between the pausing duration and change of the kinetic rates by regulatory factors. The shape of the distribution can alter dramatically with different choice of the entry rate d_1 and escape rate a_0. With different values and ratio of d_1 to a_0 the pausing duration can average largely ranging from $> 25\ s$ to $< 10\ s$ (Figure 4.3 and 4.4). In some study, the lifetime of a hairpin-stabilized pause (*his*) has been shown to tremendously decrease due to disruption of the long-lived pause signal, i.e. addition of antisense oligonucleotides that prevent hairpin formation [57]. These are interpreted by our model as factors to change the ratio of $d_1 : a_0$, leading to the change of pausing characteristics as short-lived. As is seen in both Figure 4.3 and Figure 4.4, when $d_1 : a_0 = 5$, the distribution is heavier tailed than when $d_1 : a_0 = 2$, and the system has a higher probability to be stalled in a pause for a longer time either initially starting from an elemental pause or a stabilized pause. Other ratios such as $d_1 : a_0 = m,\ (m = 1/10, 1/9, ..., 1/2, 1, 2, ..., 9, 10)$ have also been used

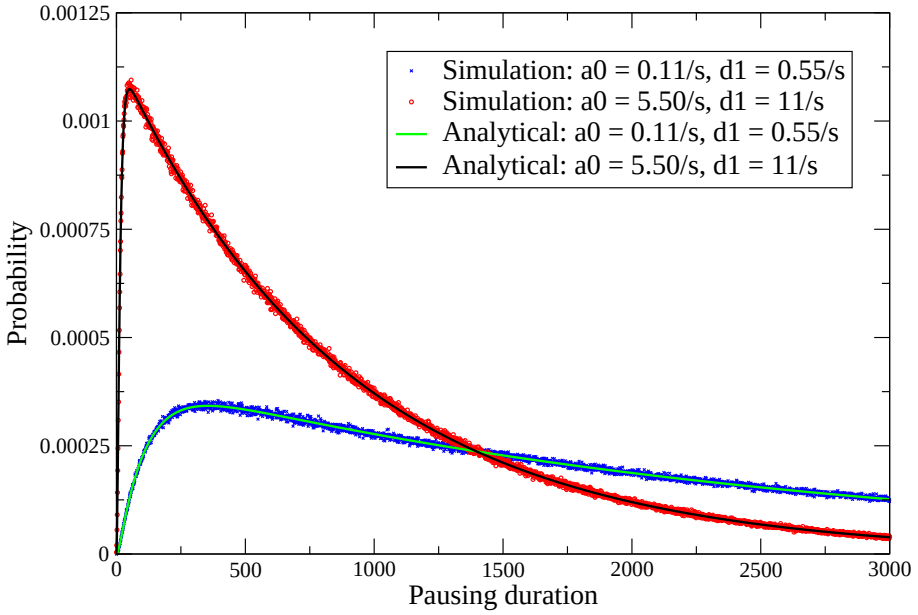


Figure 4.3: Probability distribution of the stabilized pausing duration with initial state as the stabilized pausing state. We take $a_1 = 0.00347\ s^{-1}$. With $a_0 = 0.11\ s^{-1}$, $d_1 = 0.55\ s^{-1}$ the distribution has mean $= 26.637\ s$, std $= 25.532\ s$. With $a_0 = 5.5\ s^{-1}$, $d_1 = 11\ s^{-1}$ the distribution has mean $= 9.001\ s$, std $= 8.851\ s$. The mean pausing duration is in the units of $0.01\ s$.

to verify the analytical results. The values of d_1 and a_0 comparatively have less an impact on the pausing duration, which will be discussed next. We have done stochastic simulations using the Gillespie algorithm for both the examples to validate our analytical results of the cases.

4.2.5 Discussion

After we have confirmed that the probability distribution of the pausing duration varies greatly with the change of reaction rates, we look into the variation of the mean value of different distributions with respect of different

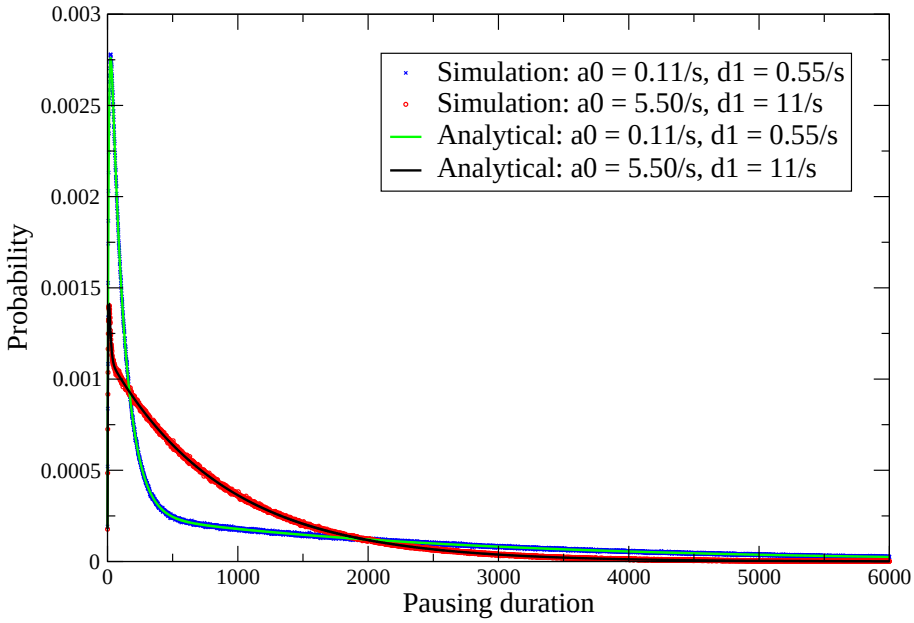


Figure 4.4: Probability distribution of the stabilized pausing duration with initial state as the elemental pausing state. We take $a_1 = 0.00347\ s^{-1}$. With $a_0 = 0.11\ s^{-1}$, $d_1 = 0.55\ s^{-1}$ the distribution has mean $= 17.547\ s$, std $= 23.85833\ s$. With $a_0 = 5.5\ s^{-1}$, $d_1 = 11\ s^{-1}$ the distribution has mean $= 8.819\ s$, std $= 8.850\ s$. The pausing duration is in the units of $0.01\ s$.

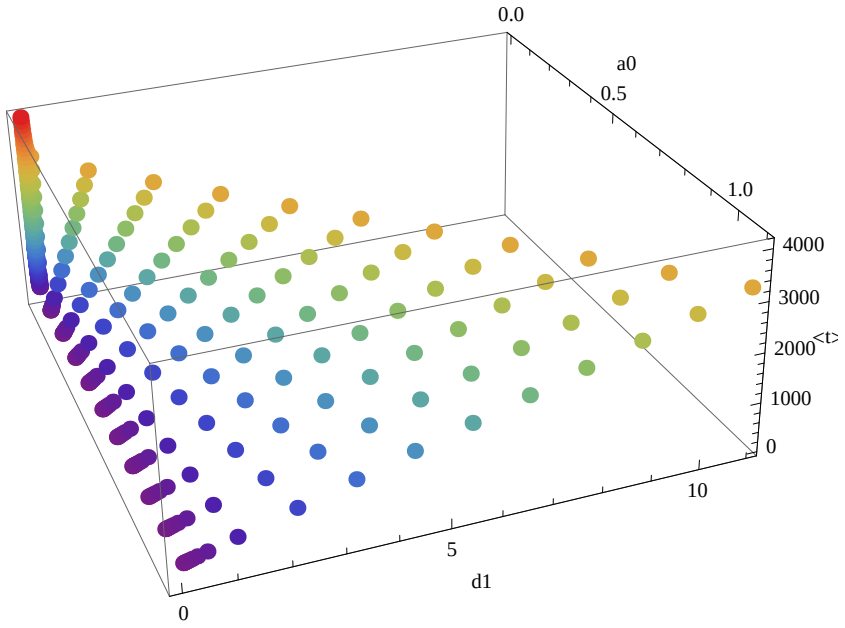


Figure 4.5: Probability distribution of the pausing duration with initial state as the stabilized pausing state in the selected parameter space. a_0 is selected as $a_2/100$, $a_2/90$, $a_2/80$, $\cdots$, a_2, $2a_2$, $3a_2$, $\cdots$, $10a_2$; d_1 is selected as $a_0/10$, $a_0/9$, $a_0/8$, $\cdots$, a_0, $2a_0$, $3a_0$, $\cdots$, $10a_0$.

choices of the rates. We sample the parameter space with a set of entry rate d_1 and escape rate a_0 such that $d_1 : a_0 = a_0/10, a_0/9, a_0/8, \cdots, a_0, 2a_0, 3a_0, \cdots, 10a_0$. The results shown in Figure 4.5 displays an obvious dependence of the mean values of the pausing duration distributions on the ratio $d_1 : a_0$. We select two sets of the parameters with $a_0 = 0.11\ s^{-1}$ and $a_0 = 0.22\ s^{-1}$ to have a closer look at the data in Figure 4.6. The near-linear behavior indicates that how long the system will spend stalling in the pausing states, more specifically the stabilized pausing state, largely relies on the relative value of d_1 and a_0. The actual values of d_1 and a_0 also have an influence on the pausing duration but are much less significant comparatively.

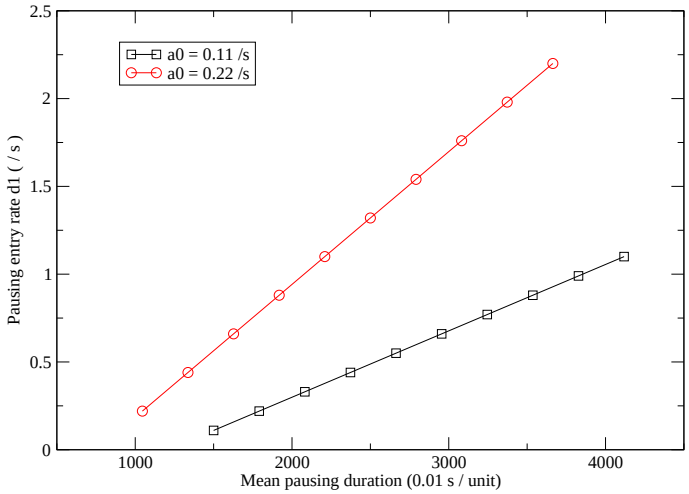


Figure 4.6: Probability distribution of the pausing duration with initial state as the stabilized pausing state. We choose $a_0 = 0.11\ s^{-1}$ (black) or $a_0 = 0.22\ s^{-1}$ (red) and varying entry rates d_1. The mean pausing duration is in the units of 0.01 s.

In the case when the initial state is the elemental pausing state (Figure 4.7 and 4.8) the ratio $d_1 : a_0$ plays an even larger role than the variation of their values. The mean values of pausing duration distributions are almost identical for data with the same ratio even though the values of the variance of the distributions change in different cases.

When the ratio $d_1 : a_0$ is too high or when the escape rate a_0 is close to 0, the pausing state is no longer a transient state but converts into a permanent state instead, which is termed "arrest" [85]. So we can adjust the model into alternative four states: active elongation state, active non-pausing state, reversible pausing intermediate state, arrested state.

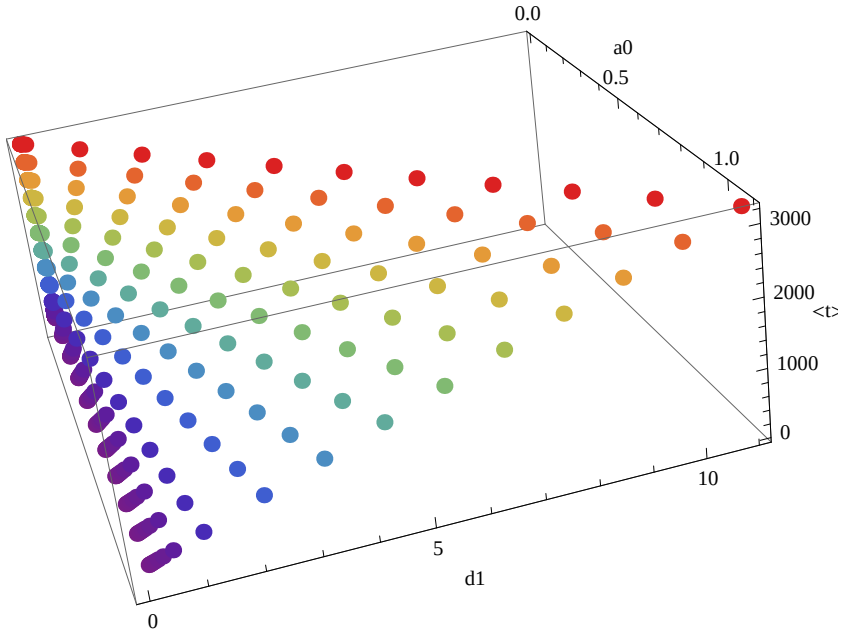


Figure 4.7: Probability distribution of the pausing duration with initial state as the stabilized pausing state in the selected parameter space. a_0 is selected as $a_2/100$, $a_2/90$, $a_2/80$, $\cdots$, a_2, $2a_2$, $3a_2$, $\cdots$, $10a_2$; d_1 is selected as $a_0/10$, $a_0/9$, $a_0/8$, $\cdots$, a_0, $2a_0$, $3a_0$, $\cdots$, $10a_0$.

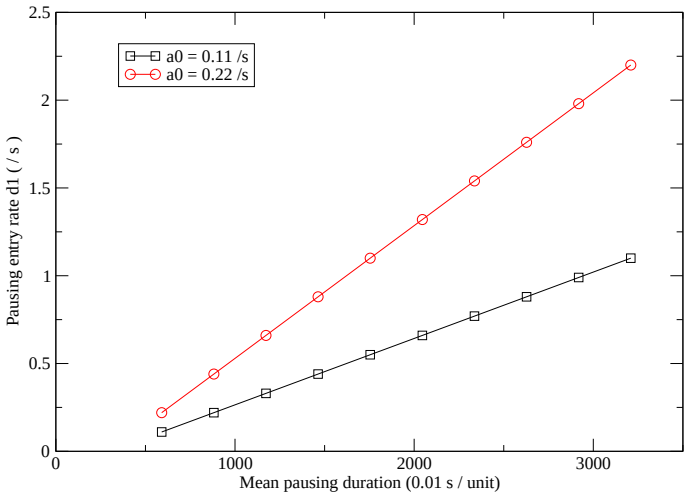


Figure 4.8: Probability distribution of the pausing duration with initial state as the stabilized pausing state. We choose $a_0 = 0.11\ s^{-1}$ (black) or $a_0 = 0.22\ s^{-1}$ (red) and varying entry rates d_1. The mean pausing duration is in the units of $0.01\ s$.

4.3 Outlook

Transcriptional pausing has been studied for almost three decades and its underlying mechanism is still unclear. Our quantitative model provides a fresh insight to understand how the pausing duration varies with the relative rates at which the TEC system enters into and escapes from the pauses.

We anticipate applications for further studies on eukaryotes due to the generality of our model. As mentioned in the previous chapter, Pol II stalling facilitates rapid temporal and spatial changes in gene activity, which reflects two nonexclusive developmental functions: it could activate transcriptional repression, or it could prepare genes for activation at later stages of empryogenesis. Regulated release of paused polymerase has been identified by several studies as a universal mechanism of gene regulation well over their originally emphasized role in the regulation of *Drosophila* heat-shock gene expression [53]. Moreover, genome-wide studies in *Drosophila* and humans indicate that polymerase pausing is widespread in the genome and is strongly associated with genes implicated in development and response to stimuli that include transcription factors and components of multiple cell signaling pathways [54, 55, 68, 69]. There are mounting evidence showing a key regulatory role of Pol II pausing for gene transcription, which leads to strong motivations for applying our model on eukaryotes. Although the pausing pattern of Pol II is a little different from that of *E. coli* RNAP, for instance, Pol II pausing is observed to mostly stabilized into backtracking pauses [86], a better

understanding of the issue is expected with the mounting experimental data besides the *Hsp70* gene of *Drosophila*. As a matter of fact, a recent discovery reveals that c-Myc releases Pol II from pauses on a large subset of genes in mouse embryonic stem cell, which indicates a significant progress of the issue because the release of Pol II from the promoter region into the body of gene is recognized as a rate-limiting step in the control of gene expression might be effective in halting the proliferation of cancer cells. We can soon anticipate more advances in understanding the pausing dynamics with the help of our model.

Finally we note that the model developed also corresponds to the three-step transcription initiation model of McClure [87, 88, 89, 90], which includes RNAP recruitment, closed and open complex formation and irreversible promoter escape leading to transcriptional elongation. Transcription initiation is a multistep process that begins with the binding of RNA polymerase to the appropriate promoter sequence in DNA, forming an open promoter complex in which the DNA is locally melted to form a transcription bubble and the bases of the template-strand DNA is exposed. After the RNA transcription starts, the transcriptional factors escape from the promoter region, perhaps accompany the releasing of initiation factors in the process. Such process is also called promoter clearance. The whole transcriptional initiation phase is therefore considered to be a three-step process which our model will fit perfectly. Thus the results derived will also provide insights into mechanisms of transcription initiation.

Chapter 5

Single-molecule enzyme kinetics

In this chapter we study the fluctuations of substrate in single-molecule enzyme kinetics. We look into the effect of the substrate fluctuation around the steady-state result according to the Michaelis-Menten equation.

5.1 Introduction

The study of enzyme kinetics has great significance in understanding and predicting how enzymes work in living organisms, their roles to control metabolism, and how their activities can be controlled or inhibited by drugs. Recent progresses in microscopic technology have made it possible to study the turnovers of enzymatic reactions on single-molecule level. Therefore, people have been investigating stochastic models for the catalytic mechanism of a single enzyme molecule to supplement the steady-state theory according to the Michaelis-Menten (MM) model.

5.1.1 Enzyme kinetics

Enzymes are needed in almost all biochemical or cellular processes and it has yet been a challenge to fully understand what an enzyme might do or how it will respond to environmental changes in a cell. The studies of enzymatic mechanism show that enzymes are selective to substrates and their catalyzed reaction rates largely depend on the substrate concentration, i.e. number of substrate molecules. As shown in an example progress curve (Figure 5.1) the enzyme converts substrate into product at an approximately linear rate at the beginning of the reaction, but the rate slows down as the reaction proceeds and substrate is consumed. Most enzyme kinetics studies focus on the initial, near-linear part of enzyme reactions. Experiments show that in a single-substrate enzyme catalysis the initial reaction rate increases with substrate concentration and asymptotically approaches a maximum when the substrate concentration is relatively high and almost all the free enzymes are converted into the substrate-bound form. This saturatable kinetics is well interpreted by the MM model.

5.1.2 Michaelis-Menten mechanism

The Michaelis-Menten (MM) kinetics is named after the German chemist Leonor Michaelis and his Canadian postdoc Maud Leonora Menten. It has been the leading quantitative theory of single-substrate enzyme kinetics for a century. In the MM model the enzymatic mechanism is described as one

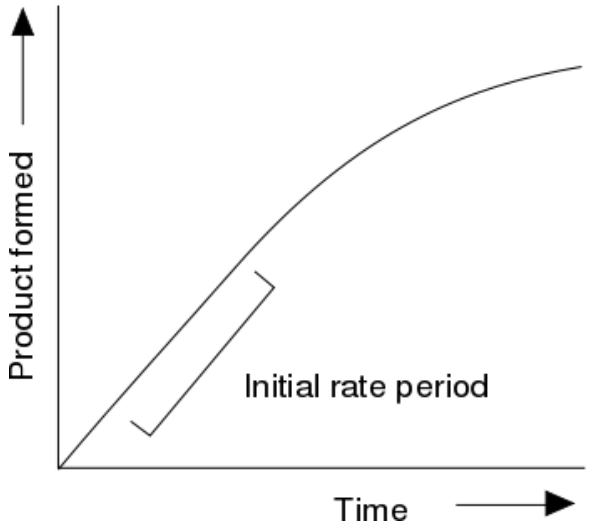


Figure 5.1: An example progress curve for an enzyme reaction. The slope of the curve is approximately constant at the beginning. This slope namely the initial rate of reaction varies with different concentration of substrate (Michaelis-Menten kinetics). (This figure is modified from wikipedia website.)

reversible step of forming the enzyme-substrate complex and one irreversible step of producing the product with constant rates (i.e. there is no catalytic intermediate or product inhibition, and no allostericity or cooperativity).

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_2} E + P$$

where k_i are the reaction rates, E and S denotes enzyme and substrate molecules, ES is an enzyme-substrate complex and P is a product molecule. The equations of rate change for their concentrations are

$$\begin{cases} \dfrac{d[ES]}{dt} = k_1[E][S] - (k_{-1} + k_2)[ES] \\ \dfrac{d[E]}{dt} = -k_1[E][S] + (k_{-1} + k_2)[ES] \\ \dfrac{d[P]}{dt} = k_2[ES] \end{cases}$$

Note that the MM kinetics is based on two key assumptions. The first assumption is the quasi-steady-state approximation, which states that the concentration of the substrate-bound enzyme changes much more slowly than those of the product and substrate, thus $d[ES]/dt = 0$. The second assumption is that the total enzyme concentration keeps constant, therefore $[E]_{\text{tot}} = [E] + [ES] = \text{const}$. Combining these two equations we have

$$\begin{aligned} \frac{d[ES]}{dt} &= k_1[S]\left([E]_{\text{tot}} - [ES]\right) - (k_{-1} + k_2)[ES] = 0 \\ [S][E]_{\text{tot}} &= [S][ES] + [ES](k_{-1} + k_2)/k_1 \\ [ES] &= \frac{[S][E]_{\text{tot}}}{K_M + [S]} \end{aligned}$$

The initial reaction rate is therefore

$$v_0 \equiv \frac{d[P]}{dt} = k_2[ES] \qquad \text{or} \qquad v_0 = \frac{V_{\max}[S]}{K_M + [S]} \tag{5.1}$$

where the constants are $V_{\max} = k_2[E]_{\text{tot}}$ and $K_M \equiv \dfrac{k_{-1} + k_2}{k_1}$. The above Eq. (5.1) is called the Michaelis-Menten equation which gives a quantitative guide of how the initial rate of the enzyme-catalyzed reaction changes with

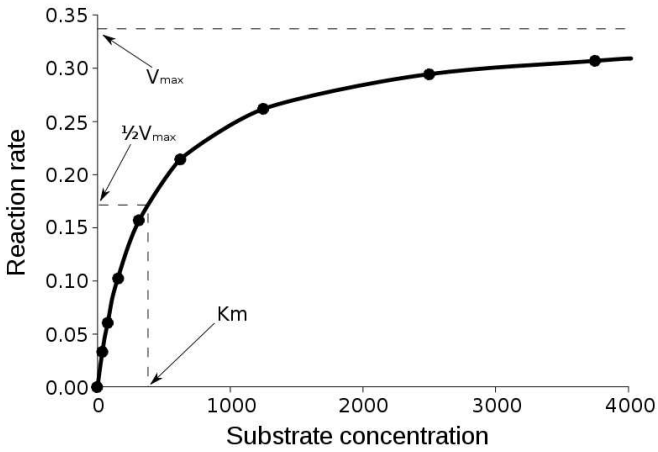


Figure 5.2: The plot of the initial reaction rate v_0 vs. substrate concentration [S] shows v_0 is saturated at high [S]. (This figure is modified from wikipedia website.)

the substrate concentration. We can see from Figure 5.2 as well as the MM equation that v_0 increases non-linearly with [S] and display saturation when the substrate is excessive.

- When $[\mathrm{S}] \ll K_M$, very little ES complex is formed, so $[E]_{\mathrm{tot}} \simeq [E]$ and the product generation rate is dependent both on the enzyme concentration and the substrate concentration. $v_0 \approx V_{\max}[S]/K_M = (k_2/K_M)[E][S]$.

- When $[\mathrm{S}] = K_M$, $v_0 = \frac{1}{2}V_{\max}$. So the constant K_M is experimentally defined as the substrate concentration at which the reaction rate is half of its maximum. Each enzyme has a characteristic K_M for a given substrate to show how tight the binding of the substrate is to the enzyme.

- When $[\mathrm{S}] \gg K_M$, $v_0 \approx V_{\max}$, the product generation rate reaches maximum (saturation) at high substrate concentration.

5.1.3 Single-molecule enzymology

Only in the past decade, however, have the experimental techniques been improved sufficiently to permit the study of single-molecule enzymology on how enzymes interact with their substrates and the differences between individual enzymes in a population [91, 92, 93]. This progress on the experimental front calls for the revival of the earlier theoretical works and their integration with the existing literature on single-channel recording. For example, the stochastic trajectories of a single-molecule property, i.e. enzymatic turnovers of a few motor protein systems, have been recorded in real time and detailed dynamical information has been extracted through statistical analysis such as the fluctuations in enzyme activities [94, 95, 96]. This newly-revealed dynamic disorder behavior is beyond the scope of conventional chemical kinetics that only accounts for the ensemble-average behaviors. The MM model that arises from deterministic chemical rate equations under quasi-steady-state approximation is believed to fail in the situations in which enzyme or substrate concentrations are very low and concentration fluctuations are non-negligible. Therefore supplementary theoretical models are needed for the analysis including the dynamic variation of reaction rates in single-molecule enzymatic kinetics.

5.1.4 The turnover time probability distribution

From a single-molecule perspective, we consider the evolution of probabilities for the state the single enzyme molecule is in. The basic MM model of enzymatic reaction is given by

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_2} E + P$$

We replace the chemical rate equations with the following equations for the probabilities

$$\frac{dP_{ES}}{dt} = k_1[S]P_E - (k_{-1} + k_2)P_{ES} \tag{5.2}$$

$$\frac{dP_E}{dt} = -k_1[S]P_E + (k_{-1} + k_2)P_{ES} \tag{5.3}$$

$$\frac{dP_{prod}}{dt} = k_2 P_{ES} \tag{5.4}$$

Note that P_{prod} denotes the probability that a turnover occurs before time t, $P(\tau \leq t)$, when τ denotes the turnover time. Then P_{prod} is the cumulative probability distribution for $P(\tau)$ and from Eq. (5.4) we can get its probability density function (pdf). The pdf of turnover time with a constant substrate concentration has been derived [97] with the initial conditions $P_E(0) = 1$, $P_{ES}(0) = 0$, $P_{prod}(0) = 0$.

$$P(\tau) = \frac{k_1 k_2[S]}{2A}\left\{e^{(A+B)\tau} - e^{(B-A)\tau}\right\} \tag{5.5}$$

where

$$A = \sqrt{(k_1[S] + k_{-1} + k_2)^2/4 - k_1[S]k_2}, \quad B = -(k_1[S] + k_{-1} + k_2)/2$$

The average turnover rate is also the inverse of the average turnover time.

$$v = \frac{1}{< \tau >} = \frac{1}{\int_0^\infty \tau P(\tau) d\tau}$$

or

$$v = \frac{k_2[S]}{[S] + K_M}$$

which recovers the MM equation (5.1).

5.2 The model

Due to recent advances in studying the kinetics at the level of single molecule, the question has been raised as to whether ensemble-average-based MM equation is adequate to describe the fluctuations in single-molecule enzyme activities. In this section we present a stochastic model to supplement the MM mechanism to study the effect of substrate fluctuations on enzymatic turnovers. We consider that the number of the substrate molecules fluctuates around the mean value of its density distribution so that when we study the probability distribution of the turnover time $P(\tau)$, we use the average number $< n >$ for $[S]$ in Eq. (5.5). Besides, we assume that the substrate

has an constant influx of rate k_{in} into the volume around the enzyme at steady-state.

$$nS \xrightarrow{k_{in}} (n+1)S \tag{5.6}$$

In this case the enzymatic reactions are represented as

$$E + nS \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES + (n-1)S \tag{5.7}$$

and

$$ES \xrightarrow{k_2} E + P \tag{5.8}$$

The above schemes lead to the following master equations:

$$\begin{aligned}
\frac{dP_E(n)}{dt} &= k_{in}P_E(n-1) + k_{-1}P_{ES}(n-1) + k_2P_{ES}(n) - (k_{in} + k_1 n)P_E(n) \\
\frac{dP_{ES}(n)}{dt} &= k_{in}P_{ES}(n-1) + k_1(n+1)P_E(n+1) - (k_{in} + k_{-1} + k_2)P_{ES}(n)
\end{aligned}$$

One way to analyze the above equations is to convert them into probability-generating functions. In Chapter 2 we have mentioned that the generating function has the relationship with the probability density function such as

$$G(z,t) = \sum_{n=0}^{\infty} P(n,t)z^n$$

Therefore we express the master equations as

$$\frac{\partial G_E(z,t)}{\partial t} = k_{in} z G_E + k_{-1} z G_{ES} + k_2 G_{ES} - k_{in} G_E - k_1 z \frac{\partial G_E}{\partial z} \quad (5.9)$$

$$\frac{\partial G_{ES}(z,t)}{\partial t} = k_{in} z G_{ES} + k_1 \frac{\partial G_E}{\partial z} - (k_{in} + k_{-1} + k_2) G_{ES} \quad (5.10)$$

5.2.1 Steady-state substrate fluctuations

Now we denote E as state 0 when the active enzyme site is vacant and ES as state 1 when it is occupied. For the steady states of the system $\partial G_E(z,t)/\partial t = \partial G_{ES}(z,t)/\partial t = 0$ we have

$$0 = k_{in} z G_0 + k_{-1} z G_1 + k_2 G_1 - k_{in} G_0 - k_1 z \frac{\partial G_0}{\partial z} \quad (5.11)$$

$$0 = k_{in} z G_1 + k_1 \frac{\partial G_0}{\partial z} - (k_{in} + k_{-1} + k_2) G_1 \quad (5.12)$$

We rearrange Eq. (5.12) and get

$$k_1 \frac{\partial G_0}{\partial z} = (k_{in} + k_{-1} + k_2) G_1 - k_{in} z G_1$$

Then we substitute the above equation into Eq. (5.11) so that

$$(z-1)[k_{in} G_0 - (k_2 - k_{in} z) G_1] = 0$$

or

$$G_1 = \frac{k_{in}}{k_2 - k_{in} z} G_0 \quad (5.13)$$

next we plug the above equation back into Eq. (5.10) to solve for $G_0(z)$

$$\frac{\partial G_0}{\partial z} = \left[\frac{k_{in}(k_{in}+k_{-1})}{k_1(k_2-k_{in}z)} + \frac{k_{in}}{k_1}\right] G_0$$

The solution will be

$$\begin{aligned} G_0(z) &= C \times \exp\left[\frac{k_{in}+k_{-1}}{-k_1}\ln(k_1k_2-k_1k_{in}z) + \frac{k_{in}}{k_1}z\right] \\ &= C \times (k_1k_2-k_1k_{in}z)^{-\frac{k_{in}+k_{-1}}{k_1}} \exp\left[\frac{k_{in}}{k_1}z\right] \end{aligned} \tag{5.14}$$

Since $P_E + P_{ES} = 1$, we have the boundary condition $G_0(1,t) + G_1(1,t) = 1$. Then the constant C will be found as

$$C = \frac{k_2-k_{in}}{k_2}(k_1k_2-k_1k_{in})^{\frac{k_{in}+k_{-1}}{k_1}} \exp\left[-\frac{k_{in}}{k_1}\right]$$

and therefore the solutions are

$$\begin{cases} G_0(z) = \left(\dfrac{k_2-k_{in}}{k_2}\right)\left(\dfrac{k_2-k_{in}}{k_2-k_{in}z}\right)^{\frac{k_{in}+k_{-1}}{k_1}} \times \exp\left[\frac{k_{in}}{k_1}(z-1)\right] \\ G_1(z) = \dfrac{k_{in}}{k_2}\left(\dfrac{k_2-k_{in}}{k_2-k_{in}z}\right)^{\frac{k_{in}+k_{-1}}{k_1}+1} \times \exp\left[\frac{k_{in}}{k_1}(z-1)\right] \end{cases} \tag{5.15}$$

In Chapter 2 we have introduced that the generating function provides us a way to obtain the quantities of interest without deriving the actual probability

distribution.

Since

$$\left[\frac{\partial \ln G(z)}{\partial z}\right]_{z=1} = \left[\frac{1}{G(z)}\frac{\partial G(z)}{\partial z}\right]_{z=1} = < n >$$

$$\left[\frac{\partial^2 \ln G(z)}{\partial z^2}\right]_{z=1} = \left[\frac{\partial}{\partial z}\left(\frac{1}{G}\frac{\partial G}{\partial z}\right)\right]_{z=1} = \left[\frac{1}{G}\frac{\partial^2 G}{\partial z^2} - \frac{1}{G^2}\left(\frac{\partial G}{\partial z}\right)^2\right]_{z=1}$$

$$= < n(n-1) > - < n >^2 = < n^2 > - < n >^2 - < n >$$

we can find the mean and the variance of the distribution as

$$< n >= \sum n \left[P_0(n) + P_1(n)\right] = \left[\frac{\partial G_0(z)}{\partial z} + \frac{\partial G_1(z)}{\partial z}\right]_{z=1}$$

$$\sigma^2 =< n^2 > - < n >^2 = \left[\frac{\partial^2 G_0(z)}{\partial z^2} + \frac{\partial^2 G_1(z)}{\partial z^2}\right]_{z=1} + < n > - < n >^2$$

thus

$$\begin{cases} < n > = \dfrac{k_{in}(k_{-1}+k_2)}{k_1 k_2} + \dfrac{k_{in}^2(k_{-1}+k_2+k_1)}{k_1 k_2(k_2-k_{in})} = \dfrac{k_{in}(k_{-1}k_2 + k_2^2 + k_{in}k_1)}{k_1 k_2(k_2-k_{in})} \\ \sigma^2 = \dfrac{k_{in}^2\left[(k_{-1}+k_2)^2 + k_1(k_{in}+k_{-1})\right]}{k_1^2 k_2(k_2-k_{in})} \\ \quad + \dfrac{k_{in}^3\left[(k_{-1}+k_2)^2 + k_1(2k_1+3k_{-1}+2k_2+k_{in})\right]}{k_1^2 k_2(k_2-k_{in})^2} \\ \quad + \dfrac{k_{in}(k_{-1}k_2 + k_2^2 + k_{in}k_1)}{k_1 k_2(k_2-k_{in})} - \left[\dfrac{k_{in}(k_{-1}k_2 + k_2^2 + k_{in}k_1)}{k_1 k_2(k_2-k_{in})}\right]^2 \\ \quad = \dfrac{k_{in}^2(k_{in}k_1 - k_2^2)}{k_1 k_2^2(k_2-k_{in})} + \dfrac{k_2}{k_2-k_{in}} < n > \end{cases}$$

5.2.2 Simulations

As we mentioned in the previous section, in our model we substitute the constant value for the substrate concentration $[S]$ with the average number of the substrate molecules $< n >$ in the turnover time probability distribution $P(\tau)$ Eq. (5.5).

$$P(\tau) = \frac{k_1 k_2 < n >}{2A} \left\{ e^{(A+B)\tau} - e^{(B-A)\tau} \right\}$$

where

$$A = \sqrt{\frac{(k_1 < n > + k_{-1} + k_2)^2}{4} - k_1 k_2 < n >}$$

$$B = -\frac{k_1 < n > + k_{-1} + k_2}{2}$$

and

$$< n >= \frac{k_{in}(k_{-1}k_2 + k_2^2 + k_{in}k_1)}{k_1 k_2 (k_2 - k_{in})}$$

To validate this theoretical results, we have also tried a diverse choices of parameters for simulations with the consideration of substrate fluctuations. The numerical results of turnover rate probability distributions have shown rough agreement with the analytical expressions in cases of both slow steady-state and fast steady-state turnover rates, however there are also clear deviations. Figure 5.3 represents a case of slow steady-state turnover, i.e. $v/k_2 = 0.05$,

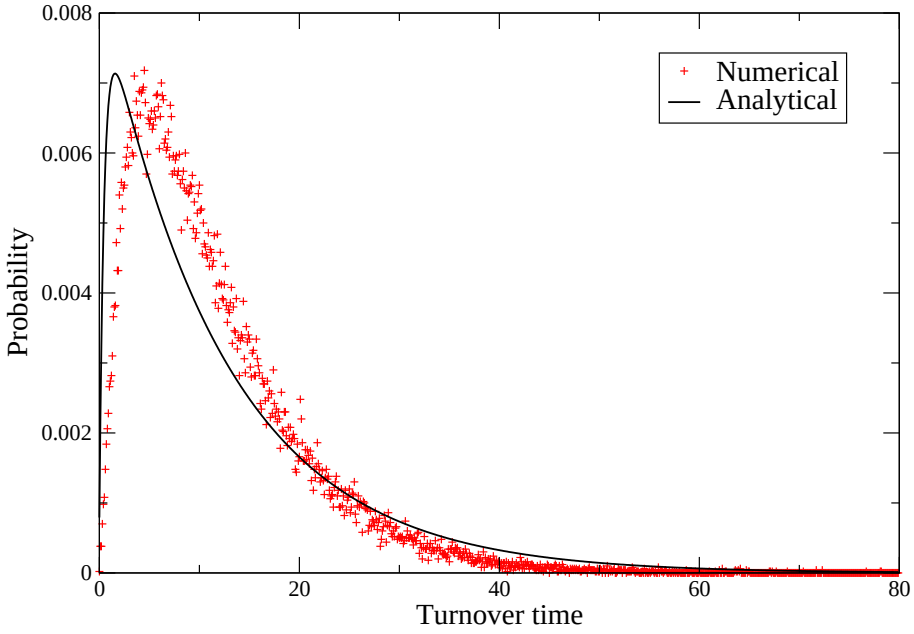


Figure 5.3: Probability distribution of the turnover time for slow steady-state rates. We choose the parameters as $k_1 = 1.0\ s^{-1}$, $k_{-1} = 1.0\ s^{-1}$, $k_2 = 1.0\ s^{-1}$ and $k_{in} = 0.05\ s^{-1}$, in which case $< n >= 0.108$.

that is related to the turnovers 5% of the maximum rate; and a case of fast steady-state turnover, i.e. $v/k_2 = 0.7$, corresponding to the turnovers 70% of the maximum rate is shown in Figure 5.4. The discrepancy indicates that the effect of substrate fluctuations cannot be adequately captured by a single average rate of substrate binding (using only the mean substrate levels), as also seen in previous work [98]. Therefore, we need to look into the actual probability distribution rather than the mean value, i.e. taking into account the fluctuation of substrate concentrations.

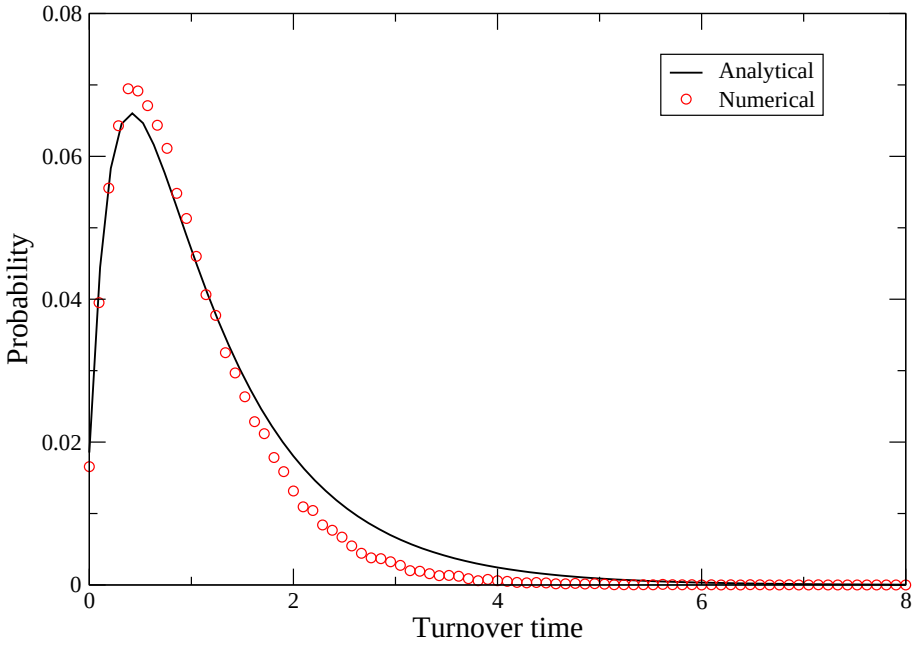


Figure 5.4: Probability distribution of the turnover time fast steady-state rates. We choose the parameters as $k_1 = 1.0\ s^{-1}$, $k_{-1} = 0.001\ s^{-1}$, $k_2 = 1.0\ s^{-1}$ and $k_{in} = 0.7\ s^{-1}$, in which case $< n >= 3.969$.

5.2.3 The probability distribution of the enzyme state

It is possible to analytically derive the actual probability distributions of p_0 and p_1 from $G_0(z)$ and $G_1(z)$, if we denote $\gamma = \dfrac{k_{in} + k_{-1}}{k_1}$ and rewrite Eq. (5.15) as follows.

$$\begin{cases} G_0(z) = \left(1 - \dfrac{k_{in}}{k_2}\right)^{\gamma+1} e^{-\frac{k_{in}}{k_1}} \sum_n \left(\dfrac{k_{in}}{k_1}\right)^n \dfrac{z^n}{n!} \left(1 - \dfrac{k_{in}}{k_2} z\right)^{-\gamma} \\ G_1(z) = \dfrac{k_{in}}{k_2} \left(1 - \dfrac{k_{in}}{k_2}\right)^{\gamma+1} e^{-\frac{k_{in}}{k_1}} \sum_n \left(\dfrac{k_{in}}{k_1}\right)^n \dfrac{z^n}{n!} \left(1 - \dfrac{k_{in}}{k_2} z\right)^{-\gamma-1} \end{cases}$$

The standard procedure to find the probability distribution $p(m)$ is that

$$\oint_l \frac{G(z)}{z^{m+1}} dz = \oint_l \frac{\sum z^n p(n)}{z^{m+1}} dz = p(m) \oint_l \frac{dz}{z} = 2\pi i \; p(m)$$

hence

$$p(m) = \frac{1}{2\pi i} \oint_l \frac{G(z)}{z^{m+1}} dz = \sum \text{Res} \left[\frac{G(z)}{z^{m+1}} \right]$$

To find the residues of the above form we need to find the singularities. $\gamma = 1.0$ for the choice of $k_1 = 1.0 \; s^{-1}$, $k_{-1} = 1.0 \; s^{-1}$ and $k_{in} = 0.05 \; s^{-1}$; $\gamma = 10.0$ for the choice of $k_1 = 1.0 \; s^{-1}$, $k_{-1} = 10.0 \; s^{-1}$ and $k_{in} = 0.7 \; s^{-1}$. In such cases, $G_0(z)$ has poles of $(m - n + 1)^{th}$ order $z = 0$ when $n < m + 1$, and poles of γ^{th} order $z = k_2/k_{in}$, but $G_0(z)$ only has poles $z = k_2/k_{in}$ when $n \geq m + 1$.

In the cases when γ is an integer

$$
\begin{aligned}
& \operatorname{Res}\left[\frac{G_0(z)}{z^{m+1}}\right]_{z=0} \\
= & \lim_{z\to 0}\frac{1}{(m-n)!}\frac{d^{m-n}}{dz^{m-n}}\left[\frac{z^{m-n+1}G_0(z)}{z^{m+1}}\right] \\
= & \left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\sum_{n=0}^{m}\left(\frac{k_{in}}{k_1}\right)^{n}\frac{e^{-\frac{k_{in}}{k_1}}}{n!(m-n)!}\lim_{z\to 0}\frac{d^{m-n}}{dz^{m-n}}\left[\left(1-\frac{k_{in}}{k_2}z\right)^{-\gamma}\right] \\
= & \left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\sum_{n=0}^{m}\left(\frac{k_{in}}{k_1}\right)^{n}\frac{e^{-\frac{k_{in}}{k_1}}}{n!(m-n)!}\frac{(\gamma+m-n-1)!}{(\gamma-1)!}\left(\frac{k_{in}}{k_2}\right)^{m-n} \\
= & \left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\sum_{n=0}^{m}\begin{pmatrix} m-n+\gamma-1 \\ \gamma-1 \end{pmatrix}\left(\frac{k_2}{k_1}\right)^{n}\left(\frac{k_{in}}{k_2}\right)^{m}\frac{e^{-\frac{k_{in}}{k_1}}}{n!} \\
= & \left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\left(\frac{k_{in}}{k_2}\right)^{m}\frac{(\gamma+m-1)!\,e^{-\frac{k_{in}}{k_1}}}{m!(\gamma-1)!}\,F_1(-m,1-\gamma-m,\frac{k_2}{k_1})
\end{aligned}
$$

where $F_1(-m,1-\gamma-m,\frac{k_2}{k_1})$ is a hypergeometric function,

$$
\begin{aligned}
& \operatorname{Res}\left[\frac{G_0(z)}{z^{m+1}}\right]_{z=\frac{k_2}{k_{in}}} \\
= & \lim_{z\to\frac{k_2}{k_{in}}} \frac{1}{(\gamma-1)!}\frac{d^{\gamma-1}}{dz^{\gamma-1}}\left[\frac{(z-\frac{k_2}{k_{in}})^{\gamma}G_0(z)}{z^{m+1}}\right] \\
= & \left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\sum_n\left(\frac{k_{in}}{k_1}\right)^n\frac{e^{-\frac{k_{in}}{k_1}}}{n!(\gamma-1)!}\lim_{z\to\frac{k_2}{k_{in}}}\frac{d^{\gamma-1}}{dz^{\gamma-1}}\left[z^{n-m-1}\left(-\frac{k_2}{k_{in}}\right)^{\gamma}\right] \\
= & \left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\sum_n\left(\frac{k_{in}}{k_1}\right)^n\frac{e^{-\frac{k_{in}}{k_1}}}{n!(\gamma-1)!}\frac{(n-m-1)!}{(n-m-\gamma)!}\left(\frac{k_2}{k_{in}}\right)^{n-m-\gamma}\left(-\frac{k_2}{k_{in}}\right)^{\gamma} \\
= & \left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\sum_n\begin{pmatrix} n-m-1 \\ \gamma-1 \end{pmatrix}\left(\frac{k_2}{k_1}\right)^n\left(\frac{k_{in}}{k_2}\right)^m\frac{e^{-\frac{k_{in}}{k_1}}}{n!}(-1)^{\gamma} \\
= & \left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\left(\frac{k_{in}}{k_2}\right)^m\frac{(-1)^{\gamma}(m+1)!e^{-\frac{k_{in}}{k_1}}}{(m+\gamma)!(\gamma-1)!}F_1(-m,1-\gamma-m,\frac{k_2}{k_1})
\end{aligned}
$$

Then the probability density function will be

$$
\begin{aligned}
p_0(m) &= \sum \operatorname{Res}\left[\frac{G_0(z)}{z^{m+1}}\right] \\
&= \left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\left(\frac{k_{in}}{k_2}\right)^m e^{-\frac{k_{in}}{k_1}} \\
&\qquad \left[\frac{(\gamma+m-1)!}{m!(\gamma-1)!}+\frac{(-1)^{\gamma}(m+1)!}{(m+\gamma)!(\gamma-1)!}\right]F_1(-m,1-\gamma-m,\frac{k_2}{k_1})
\end{aligned}
$$

Similarly, $G_1(z)$ has poles of $(m-n+1)^{th}$ order $z=0$, and poles of $(\gamma+1)^{th}$ order $z=\frac{k_2}{k_{in}}$.

$$
\begin{aligned}
& \operatorname{Res}\left[\frac{G_1(z)}{z^{m+1}}\right]_{z=0} \\
= & \lim_{z\to 0}\frac{1}{(m-n)!}\frac{d^{m-n}}{dz^{m-n}}\left[\frac{z^{m-n+1}G_1(z)}{z^{m+1}}\right] \\
= & \frac{k_{in}}{k_2}\left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\sum_{n=0}^{m}\left(\frac{k_{in}}{k_1}\right)^n\frac{e^{-\frac{k_{in}}{k_1}}}{n!(m-n)!}\lim_{z\to 0}\frac{d^{m-n}}{dz^{m-n}}\left[\left(1-\frac{k_{in}}{k_2}z\right)^{-(\gamma+1)}\right] \\
= & \frac{k_{in}}{k_2}\left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\sum_{n=0}^{m}\left(\frac{k_{in}}{k_1}\right)^n\frac{e^{-\frac{k_{in}}{k_1}}}{n!(m-n)!}\frac{(\gamma+m-n)!}{\gamma!}\left(\frac{k_{in}}{k_2}\right)^{m-n} \\
= & \frac{k_{in}}{k_2}\left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\sum_{n=0}^{m}\binom{m-n+\gamma}{\gamma}\left(\frac{k_2}{k_1}\right)^n\left(\frac{k_{in}}{k_2}\right)^m\frac{e^{-\frac{k_{in}}{k_1}}}{n!} \\
= & \left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\left(\frac{k_{in}}{k_2}\right)^{m+1}\frac{(m+\gamma)!e^{-\frac{k_{in}}{k_1}}}{m!\,\gamma!}\,F_1(-m,-m-\gamma,\frac{k_2}{k_1})
\end{aligned}
$$

$$
\begin{aligned}
& \operatorname{Res}\left[\frac{G_1(z)}{z^{m+1}}\right]_{z=\frac{k_2}{k_{in}}} \\
= & \lim_{z\to\frac{k_2}{k_{in}}} \frac{1}{\gamma!}\frac{d^\gamma}{dz^\gamma}\left[\frac{(z-\frac{k_{in}}{k_2})^{\gamma+1}G_1(z)}{z^{m+1}}\right] \\
= & \frac{k_{in}}{k_2}\left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\sum_n\left(\frac{k_{in}}{k_1}\right)^n\frac{e^{-\frac{k_{in}}{k_1}}}{n!\gamma!}\lim_{z\to\frac{k_2}{k_{in}}}\frac{d^\gamma}{dz^\gamma}\left[z^{n-m-1}\left(-\frac{k_2}{k_{in}}\right)^{\gamma+1}\right] \\
= & \frac{k_{in}}{k_2}\left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\sum_n\left(\frac{k_{in}}{k_1}\right)^n\frac{e^{-\frac{k_{in}}{k_1}}}{n!\gamma!}\frac{(n-m-1)!}{(n-m-\gamma-1)!}\left(\frac{k_2}{k_{in}}\right)^{n-m-\gamma-1}\left(-\frac{k_2}{k_{in}}\right)^{\gamma+1} \\
= & \frac{k_{in}}{k_2}\left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\sum_n\binom{n-m-1}{\gamma}\left(\frac{k_2}{k_1}\right)^n\left(\frac{k_{in}}{k_2}\right)^m\frac{e^{-\frac{k_{in}}{k_1}}}{n!}(-1)^{\gamma+1} \\
= & \left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\left(\frac{k_{in}}{k_2}\right)^{m+1}\frac{(-1)^\gamma(m+1)!e^{-\frac{k_{in}}{k_1}}}{(m+\gamma+1)!\,\gamma!}\,F_1(-m,-m-\gamma,\frac{k_2}{k_1})
\end{aligned}
$$

The probability density function will be

$$
\begin{aligned}
p_1(m) &= \sum \operatorname{Res}\left[\frac{G_1(z)}{z^{m+1}}\right] \\
&= \left(1-\frac{k_{in}}{k_2}\right)^{\gamma+1}\left(\frac{k_{in}}{k_2}\right)^{m+1}e^{-\frac{k_{in}}{k_1}} \\
&\qquad \left[\frac{(m+\gamma)!}{m!\,\gamma!}+\frac{(-1)^\gamma(m+1)!}{(m+\gamma+1)!\,\gamma!}\right]F_1(-m,-m-\gamma,\frac{k_2}{k_1})
\end{aligned}
$$

Now that we have derived the mathematical expressions for the probability distributions of a single enzyme molecule with respect to the substrate

concentration, either in a free state or a substrate-bound state, we are able to gain further insight of the turnover dynamics and its dependence on the substrate fluctuation. We expect to continue this work along multiple paths in the future, e.g. the time-dependent enzymatic kinetics, which can be obtained from Eq. (5.9) and Eq. (5.10). Besides, for trajectories with a limited number of turnovers (limited statistics), it is helpful to evaluate the width of the distribution using the randomness parameter, $(n- < n >)^2/ < n >$, a measure of the relative magnitudes of the variance and the mean of the corresponding probability distributions. Another sensitive way to evaluate the dynamic disorder in the catalytic step of the enzymatic reaction is the auto-correlation function, $c(m) =< \Delta\tau(0)\Delta\tau(m) > / < \Delta\tau >$, where m is the index number of turnovers and $\Delta\tau(m) = \tau_m - < \tau >$.

5.3 Discussion

In this chapter, we have discussed how single-molecule enzymatic experiments can provide us measurements of the probability distribution of the waiting times for individual turnovers. More information than the ensemble-average data has emerged and awaits better interpretation and modeling that account for the apparent fluctuations observed in enzymatic kinetics. Under such circumstances, the MM equation for a deterministic rate evaluation under a quasi-steady-state approximation is believed to be inadequate. In particular, we have seen that substrate fluctuations can give rise to deviations in

the turnover time distribution when compared with the same distributions obtained using average substrate concentrations, as is usually done. In the presence of dynamic disorder, the single-molecule enzymatic turnover exhibits a dependence on the fluctuation of substrate concentration and provides more dynamic information than the steady-state ensemble kinetics. Therefore, we propose a model for single-molecule enzymatic turnovers, which leads to a quantitative understanding of how the steady-state fluctuation of substrate concentration affects the enzyme kinetics. We have derived the analytical expressions for the probability distributions with respect to different numbers of substrate molecules in steady-state and the mean and variance of the distributions. We have also evaluated the dependence of turnover time probability distribution upon the substrate concentration using analytical approaches and simulations. Further analysis of the effect of substrate fluctuations can be carried out by obtaining the time-dependent probability distribution by solving the master equation with time evolution. The expressions for the corresponding generating function can then be used to obtain additional quantities of interest such as autocorrelation functions. An important input for such analytical calculations is the steady-state distribution of substrate fluctuations which has been obtained in this work. Future work building on these results can thus provide more insights of relevance to experimental measurements of single-molecule enzyme kinetics.

Chapter 6

Conclusions

In this research we have analyzed three nucleation-based stochastic processes with quantitative models that are validated by stochastic simulations. We believe that the results we derived have provided insights to the relevant issues and will be useful for future research in the corresponding fields.

Min protein oscillation in *Escherichia coli*

In Chapter 3 we have found out that the self-organized behavior in *E. coli* cell division can be explained from a statistical point of view. The precise localization of the division site, that is strongly related to the temporal formation of MinD proteins at the cell polar zones, can be a natural result from the coordination of biological processes such as diffusion, membrane attachment and nucleotide exchange. Through our simplified computational model that is focused on the key processes, we have shown that the most likely location

for MinD:ATP membrane binding is at the bare pole where the new MinD polar zone starts to form as observed in the experiments. Thus the combination of the basic processes explored in our model can potentially explain polar localization of MinD in *E. coli* cell division. More broadly, the nucleation-based localization model proposed in this work can serve as a general paradigm for other instances of regulated protein subcellular localization.

DNA transcriptional pausing

In Chapter 4 we have presented a quantitative model that describes the transcriptional pausing phenomena as a four-state nucleation process. We have studied the distribution of times for the system to get back onto the elongation pathway from a pausing state, and how the pausing duration changes with different entry and escape rates of the pauses. We have derived the probability distribution of the pausing duration and compared the mean values of different distributions with selected reaction rates of choice. The exact analytical expressions derived have been validated by stochastic simulations and will serve as useful inputs for single-molecule studies of transcriptional pausing and initiation.

Single molecule enzyme kinetics

In Chapter 5 we have analyzed single molecule enzymatic kinetics based on the Michaelis-Menten model. We have studied the effect of substrate fluc-

tuations on turnovers of the enzymatic reactions. Steady-state results have been derived for the stochastic model under study. The extension of these results to dynamical properties will provide analytical expressions for dynamic properties of interest.

Appendix A

Calculations of the roots of a cubic function

The following method to solve a standard cubic equation was first published by Gerolamo Cardano in 1545.

For every cubic equation with real coefficients

$$s^3 + as^2 + bs + c = 0 \tag{A.1}$$

we can eliminate the quadratic term by variable change $t = s + \frac{a}{3}$,

$$t^3 + pt + q = 0 \tag{A.2}$$

where

$$p = b - \frac{a^2}{3} \quad \text{and} \quad q = c + \frac{2a^3 - 9ab}{27}$$

Eq.A.2 is called a *depressed cubic*. To solve it we introduce another two variables u and v linked by $t = u + v$. Eq.A.2 can therefore be transformed into

$$u^3 + v^3 + (3uv + p)(u + v) + q = 0$$

If we impose the condition

$$3uv + p = 0 \quad \text{or} \quad v = -\frac{p}{3u}$$

we can derive a quadratic equation in u^3

$$u^6 + qu^3 - \frac{p^3}{27} = 0 \tag{A.3}$$

from which u is simply solved as

$$u = \sqrt[3]{-\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \tag{A.4}$$

Note that u can take six possible values here

$$u = \begin{cases} \sqrt[3]{-\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \\ \left(-\frac{1}{2} + i\frac{\sqrt{3}}{2}\right) \sqrt[3]{-\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \\ \left(-\frac{1}{2} - i\frac{\sqrt{3}}{2}\right) \sqrt[3]{-\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \end{cases}$$

Since v follow the equivalent procedure and u, v must satisfy

$$u^3 + v^3 + q = 0 \quad \text{and} \quad uv = -\frac{p}{3}$$

one can show that

$$u = \begin{cases} \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \\ \left(-\frac{1}{2} + i\frac{\sqrt{3}}{2}\right) \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \\ \left(-\frac{1}{2} - i\frac{\sqrt{3}}{2}\right) \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \end{cases}$$

$$v = \begin{cases} \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \\ \left(-\frac{1}{2} + i\frac{\sqrt{3}}{2}\right) \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \\ \left(-\frac{1}{2} - i\frac{\sqrt{3}}{2}\right) \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \end{cases}$$

Then there are three combinations of u and v to make $3uv = -p$ valid, so that

$$\begin{aligned} t &= u + v \\ &= \begin{cases} \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \\ \left(-\frac{1}{2} + i\frac{\sqrt{3}}{2}\right) \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \left(-\frac{1}{2} - i\frac{\sqrt{3}}{2}\right) \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \\ \left(-\frac{1}{2} - i\frac{\sqrt{3}}{2}\right) \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \left(-\frac{1}{2} + i\frac{\sqrt{3}}{2}\right) \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \end{cases} \end{aligned}$$

Consequently we can derive the roots of the original cubic equation by $s = t - a/3$. Here let us distinguish the different cases for the roots.

If we set

$$D = \frac{q^2}{4} + \frac{p^3}{27}$$

then

1. If D is positive then there is one real and two complex roots.
2. If D is negative then there are three real roots (Casus irreducibilis).
3. If $D = 0$ then there is one real root (a triple root) or two real roots (a single and a double root).

We can show the above statements more clearly if we put u^3 into polar coordinates.

$$u^3 = -\frac{q}{2} \pm i\sqrt{-D} = (r, \pm\theta)$$

where

$$r = \sqrt{\frac{q^2}{4} - D} = \sqrt{-\frac{p^3}{27}} \quad \text{and} \quad \theta = \arccos(-\frac{q}{2r})$$

Now solve for u with six possibilities

$$(\sqrt[3]{r}, \pm\frac{\theta}{3}),\ (\sqrt[3]{r}, \pm\frac{\theta}{3}) \times (1, \frac{2\pi}{3}) \ \text{ and } \ (\sqrt[3]{r}, \pm\frac{\theta}{3}) \times (1, -\frac{2\pi}{3})$$

and equivalently v can take six corresponding possible values under the as-

sumptions of $u^3 + v^3 = -q$ and $3uv = -p$.

$$(\sqrt[3]{r}, \mp\frac{\theta}{3}),\ (\sqrt[3]{r}, \mp\frac{\theta}{3}) \times (1, -\frac{2\pi}{3}) \ \text{ and } \ (\sqrt[3]{r}, \mp\frac{\theta}{3}) \times (1, \frac{2\pi}{3})$$

Therefore s can be solved from $s = u + v - a/3$. It clearly shows below that when $D < 0$ there exist three real roots for the cubic equation.

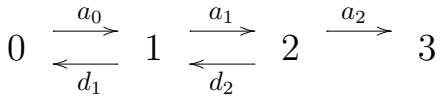
$$\begin{cases} s_1 = \sqrt[3]{r}\ e^{i\frac{\theta}{3}} + \sqrt[3]{r}\ e^{-i\frac{\theta}{3}} - \dfrac{a}{3} \\ s_2 = \sqrt[3]{r}\ e^{i(\frac{\theta}{3}+\frac{2\pi}{3})} + \sqrt[3]{r}\ e^{-i(\frac{\theta}{3}+\frac{2\pi}{3})} - \dfrac{a}{3} \\ s_3 = \sqrt[3]{r}\ e^{i(\frac{\theta}{3}-\frac{2\pi}{3})} + \sqrt[3]{r}\ e^{-i(\frac{\theta}{3}-\frac{2\pi}{3})} - \dfrac{a}{3} \end{cases}$$

or

$$\begin{cases} s_1 = 2\sqrt[3]{r}\cos(\frac{1}{3}\theta) - \dfrac{a}{3} \\ s_2 = 2\sqrt[3]{r}\cos(\frac{\theta}{3} + \frac{2\pi}{3}) - \dfrac{a}{3} \\ s_3 = 2\sqrt[3]{r}\cos(\frac{\theta}{3} - \frac{2\pi}{3}) - \dfrac{a}{3} \end{cases}$$

Appendix B

Calculations of the first passage time probability distribution: generalized four-state case



To solve the backward master equations for the first passage time probability $F_{3,m}(t)$ in the scenario of four-state, we first apply the boundary conditions.

1. For the lower boundary, state 0 is a reflecting boundary, so $\mu_0 = 0$. Plus state -1 is never reached, so $F_{3,-1}(t) = 0$ and

$$\frac{dF_{3,0}(t)}{dt} = a_0\ F_{3,1}(t) - a_0\ F_{3,0}(t)$$

2. For the upper boundary, state 3 is an absorbing boundary, so $F_{3,3}(t) = 0$

and

$$\frac{dF_{3,2}(t)}{dt} = -(a_2 + d_2)\ F_{3,2}(t) + d_2\ F_{3,1}(t)$$

Therefore we derive the equations of the system to be solved as

$$\begin{cases} \dfrac{dF_{3,0}(t)}{dt} = a_0\ F_{3,1}(t) - a_0\ F_{3,0}(t) \\ \dfrac{dF_{3,1}(t)}{dt} = a_1\ F_{3,2}(t) - (a_1 + d_1)\ F_{3,1}(t) + d_1\ F_{3,0}(t) \\ \dfrac{dF_{3,2}(t)}{dt} = -(a_2 + d_2)\ F_{3,2}(t) + d_2\ F_{3,1}(t) \end{cases}$$

We take the Laplace Transforms of the above equations while using the initial conditions as $F_{3,1}(0) = 0$, $F_{3,2}(0) = 0$ and $F_{3,2}(0) = a_2$. Now we have

$$\begin{cases} f_{3,0}(s) = \dfrac{1}{\dfrac{s + a_2 + d_2}{a_2}\left(\dfrac{s + a_1 + d_1}{a_1}\dfrac{s + a_0}{a_0} - \dfrac{d_1}{a_1}\right) - \dfrac{d_2}{a_2}\dfrac{s + a_0}{a_0}} \\ f_{3,1}(s) = \dfrac{s + a_0}{a_0}\ f_{3,0}(s) \\ f_{3,2}(s) = \left(\dfrac{s + a_1 + d_1}{a_1}\dfrac{s + a_0}{a_0} - \dfrac{d_1}{a_1}\right) f_{3,0}(s) \end{cases}$$

First we solve for $f_{3,0}(s)$. It can also be rearranged and expressed as

$$f_{3,0}(s) = \frac{a_0 a_1 a_2}{(s - s_1)(s - s_2)(s - s_3)}$$

Note that we factorize the denominator of $f_{3,0}(s)$ in order to derive the inverse

Laplace Transform $F_{3,0}(t)$ later. The values of s_1, s_2 and s_3 can be found by solving the equation

$$s^3 + As^2 + Bs + C = 0$$

where $A = a_0 + a_1 + a_2 + d_1 + d_2$, $B = a_0a_1 + a_1a_2 + a_2a_0 + a_2d_1 + a_0d_2 + d_1d_2$ and $C = a_0a_1a_2$. We have shown the standard procedure to solve for a cubic equation following Cardano's methods in the previous section Appendix A. When s_1, s_2 and s_3 are all real numbers, the results are

$$\begin{cases} s_1 = 2\sqrt[3]{r}\cos(\frac{1}{3}\theta) - A/3 \\ s_2 = 2\sqrt[3]{r}\cos(\frac{\theta}{3} + \frac{2\pi}{3}) - A/3 \\ s_3 = 2\sqrt[3]{r}\cos(\frac{\theta}{3} - \frac{2\pi}{3}) - A/3 \end{cases}$$

with the constants

$$r = \sqrt{-\frac{p^3}{27}} \quad \text{and} \quad \theta = \arccos\left(-\frac{q}{2r}\right)$$

where

$$p = B - \frac{A^2}{3} \quad \text{and} \quad q = C + \frac{2A^3 - 9AB}{27}$$

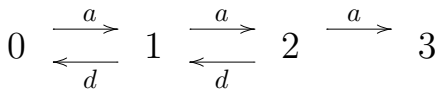
Now that we have solved for $f_{3,0}(s)$, $f_{3,1}(s)$ and $f_{3,2}(s)$, we can find the inverse

Laplace transforms to obtain all the three probability density functions of the first passage time $F_{3,0}(t)$, $F_{3,1}(t)$ and $F_{3,2}(t)$ as follows.

$$
\begin{cases}
F_{3,0}(t) = \dfrac{a_0a_1a_2e^{s_1t}}{(s_2-s_1)(s_3-s_1)} + \dfrac{a_0a_1a_2e^{s_2t}}{(s_1-s_2)(s_3-s_2)} + \dfrac{a_0a_1a_2e^{s_3t}}{(s_2-s_3)(s_1-s_3)} \\
F_{3,1}(t) = \dfrac{(a_0a_1a_2+a_1a_2s_1)e^{s_1t}}{(s_2-s_1)(s_3-s_1)} + \dfrac{(a_0a_1a_2+a_1a_2s_2)e^{s_2t}}{(s_1-s_2)(s_3-s_2)} \\
\qquad\qquad + \dfrac{(a_0a_1a_2+a_1a_2s_3)e^{s_3t}}{(s_2-s_3)(s_1-s_3)} \\
F_{3,2}(t) = \dfrac{\left[a_0a_1a_2+a_2(a_0+a_1+d_1)s_1+a_2s_1^2\right]e^{s_1t}}{(s_2-s_1)(s_3-s_1)} \\
\qquad\qquad + \dfrac{\left[a_0a_1a_2+a_2(a_0+a_1+d_1)s_2+a_2s_2^2\right]e^{s_2t}}{(s_1-s_2)(s_3-s_2)} \\
\qquad\qquad + \dfrac{\left[a_0a_1a_2+a_2(a_0+a_1+d_1)s_3+a_2s_3^2\right]e^{s_3t}}{(s_2-s_3)(s_1-s_3)}
\end{cases}
$$

Appendix C

Calculations of the first passage time probability distribution: a special case



Consider the above four-state random walk process where all the forward rates are equal to a and all the backward rates are d. State 0 is a reflecting boundary of the system and state 3 is an absorbing boundary. In this chapter we demonstrate the calculations for the first passage time probability distribution using both the backward master equations and transition rate matrix approach.

C.1 Approach 1: using backward master equations

We use $F_{3,m}$ to denote the probability density function of the first passage time for a random walker that starts from state m ($m = 0, 1, 2$) and ends at state 3. The backward master equations are applied below with the boundary conditions.

$$\begin{cases} \dfrac{dF_{3,0}(t)}{dt} = a\ F_{3,1}(t) - a\ F_{3,0}(t) \\ \dfrac{dF_{3,1}(t)}{dt} = a\ F_{3,2}(t) - (a+d)\ F_{3,1}(t) + d\ F_{3,0}(t) \\ \dfrac{dF_{3,2}(t)}{dt} = -(a+d)\ F_{3,2}(t) + d\ F_{3,1}(t) \end{cases}$$

From the Laplace Transform $f_{3,m}(s) = \displaystyle\int_0^\infty e^{-st} F_{3,m}(t)dt$

we have $\displaystyle\int_0^\infty e^{-st}\frac{dF_{3,m}(t)}{dt}dt = s\int_0^\infty e^{-st}F_{3,m}(t)dt - F_{3,m}(0) = sf_{3,m}(s) - F_{3,m}(0)$. Thus the equations are rewritten as

$$\begin{cases} sf_{3,0}(s) - F_{3,0}(0) = a\ f_{3,1}(s) - a\ f_{3,0}(s) \\ sf_{3,1}(s) - F_{3,1}(0) = a\ f_{3,2}(s) - (a+d)\ f_{3,1}(s) + d\ f_{3,0}(s) \\ sf_{3,2}(s) - F_{3,2}(0) = -(a+d)\ f_{3,2}(s) + d\ f_{3,1}(s) \end{cases}$$

Then we rearrange and solve for the above equations with the initial conditions $F_{3,0}(0) = 0$, $F_{3,1}(0) = 0$ and $F_{3,2}(0) = a$.

$$f_{3,0}(s) = \frac{1}{\frac{s+a+d}{a}(\frac{s+a+d}{a}\frac{s+a}{a} - \frac{d}{a}) - \frac{s+a}{a}\frac{d}{a}}$$
$$f_{3,1}(s) = \frac{s+a}{a}\ f_{3,0}(s)$$
$$f_{3,2}(s) = (\frac{s+a+d}{a}\frac{s+a}{a} - \frac{d}{a})\ f_{3,0}(s)$$

We can rewrite $f_{3,0}(s)$ in the form of

$$\begin{aligned} f_{3,0}(s) &= \frac{a^3}{s^3 + (3a+2d)s^2 + [3a^2+2ad+d^2]s + a^3} \\ &= \frac{a^3}{(s-s_1)(s-s_2)(s-s_3)} \end{aligned} \tag{C.1}$$

Apparently s_1, s_2 and s_3 are the solutions of the cubic equation $s^3 + As^2 + Bs + C = 0$, where $A = 3a + 2d$, $B = 3a^2 + 2ad + d^2$ and $C = a^3$. Following the method introduced in the previous section (A), we first calculate all the following related variables.

$$p = B - \frac{A^2}{3} = -2ad - \frac{1}{3}d^2$$

$$q = C + \frac{2A^3 - 9AB}{27} = \frac{1}{3}ad^2 - \frac{2}{27}d^3$$

and

$$D = \frac{q^2}{4} + \frac{p^3}{27} = -\frac{8}{27}a^3d^3 - \frac{13}{108}a^2d^4 - \frac{1}{27}ad^5$$

Since $D < 0$ in this case, we know that there are three real roots for s

$$\begin{cases} s_1 = 2\sqrt[3]{r}\cos(\frac{1}{3}\theta) - (a + \frac{2}{3}d) \\ s_2 = 2\sqrt[3]{r}\cos(\frac{\theta}{3} + \frac{2\pi}{3}) - (a + \frac{2}{3}d) \\ s_3 = 2\sqrt[3]{r}\cos(\frac{\theta}{3} - \frac{2\pi}{3}) - (a + \frac{2}{3}d) \end{cases}$$

where

$$r = \sqrt{\frac{q^2}{4} - D} = \sqrt{(\frac{2}{3}ad + \frac{1}{9}d^2)^3}$$

and

$$\theta = \arccos\left(-\frac{q}{2r}\right) = \arccos\left(\frac{-\frac{1}{6}ad^2 + \frac{1}{27}d^3}{r}\right)$$

Now that we have solved for s_1, s_2 and s_3, we can find the inverse Laplace transforms of $f_{3,m}(s)$ $(m = 0, 1, 2)$.

$$\begin{cases} F_{3,0}(t) = -\dfrac{a^3 e^{s_1 t}}{(s_1 - s_2)(s_3 - s_1)} - \dfrac{a^3 e^{s_2 t}}{(s_1 - s_2)(s_2 - s_3)} - \dfrac{a^3 e^{s_3 t}}{(s_2 - s_3)(s_3 - s_1)} \\ F_{3,1}(t) = -\dfrac{\left(a^3 + a^2 s_1\right) e^{s_1 t}}{(s_1 - s_2)(s_3 - s_1)} - \dfrac{\left(a^3 + a^2 s_2\right) e^{s_2 t}}{(s_1 - s_2)(s_2 - s_3)} - \dfrac{\left(a^3 + a^2 s_3\right) e^{s_3 t}}{(s_2 - s_3)(s_3 - s_1)} \\ F_{3,2}(t) = -\dfrac{\left[a^3 + (2a^2 + ad)s_1 + a s_1^2\right] e^{s_1 t}}{(s_1 - s_2)(s_3 - s_1)} - \dfrac{\left[a^3 + (2a^2 + ad)s_2 + a s_2^2\right] e^{s_2 t}}{(s_1 - s_2)(s_2 - s_3)} \\ \qquad\qquad - \dfrac{\left[a^3 + (2a^2 + ad)s_3 + a s_3^2\right] e^{s_3 t}}{(s_2 - s_3)(s_3 - s_1)} \end{cases} \tag{C.2}$$

C.2 Approach 2: using the transition rate matrix

In Chapter 2 we have discussed how a phase-type distribution represents a group of inter-related stochastic processes, and demonstrated how to obtain the probability density function of the random variable, time t, for a random process from starting until reaching the absorbing phase. In this specific case here where all the forward rates are the same as a and all the backward rates are equal to d, we first construct the transition rate matrix T

$$T = \begin{bmatrix} -a & a & 0 \\ d & -(a+d) & a \\ 0 & d & -(a+d) \end{bmatrix}$$

We know from Chapter 2 Eq. 2.13 that the first passage time probability density function is derived as

$$F(t) = \vec{\alpha}\, e^{t\,T}\nu = \vec{\alpha}\, P e^{t\,M} P^{-1}\nu$$

where $\vec{\alpha}$ denotes the initial condition ($m = 0, 1, 2$, is the initial position),

$$\vec{\alpha} = (\delta_{m,0},\ \delta_{m,1}\ \delta_{m,2}) \tag{C.3}$$

ν denotes all the transition from each state to the absorbing state,

$$\nu = -T\ \vec{1} = (0,\ 0,\ a) \tag{C.4}$$

and M is a diagonalized similar matrix of the transition rate matrix T. One common way to find $M = P^{-1}TP$ is to compose a matrix P with all the eigenvectors of T. To find the eigenvalues, we can use

$$\det|T - \lambda I| = \begin{vmatrix} -a-\lambda & a & 0 \\ d & -(a+d)-\lambda & a \\ 0 & d & -(a+d)-\lambda \end{vmatrix} = 0$$

Thus

$$-(a+\lambda)(a+d+\lambda)^2 + ad(a+\lambda) + ad(a+d+\lambda) = 0 \qquad \text{(C.5)}$$

Rearrange the equation according to the order of λ we have

$$\lambda^3 + (3a+2d)\lambda^2 + (3a^2+2ad+d^2)\lambda + a^3 = 0 \qquad \text{(C.6)}$$

Note that the L.H.S of Eq. (C.6) is the same as the denominator of $f_{3,0}(s)$ in Eq. (C.1). We already solved the cubic equation and derived the three roots as

$$\begin{cases} \lambda_1 = 2\sqrt[3]{r}\cos(\frac{1}{3}\theta) - (a + \frac{2}{3}d) \\ \lambda_2 = 2\sqrt[3]{r}\cos(\frac{\theta}{3} + \frac{2\pi}{3}) - (a + \frac{2}{3}d) \\ \lambda_3 = 2\sqrt[3]{r}\cos(\frac{\theta}{3} - \frac{2\pi}{3}) - (a + \frac{2}{3}d) \end{cases}$$

where $r = \sqrt{(\frac{2}{3}ad + \frac{1}{9}d^2)^3}$ and $\theta = \arccos\left(\dfrac{-\frac{1}{6}ad^2 + \frac{1}{27}d^3}{r}\right)$

We can thus derive all the three eigenvectors to compose matrix, $P = (\mathbf{x_1},\ \mathbf{x_2},\ \mathbf{x_3})$,

$$P = \begin{pmatrix} \frac{a}{a+\lambda_1} & \frac{a}{a+\lambda_2} & \frac{a}{a+\lambda_3} \\ 1 & 1 & 1 \\ \frac{d}{a+d+\lambda_1} & \frac{d}{a+d+\lambda_2} & \frac{d}{a+d+\lambda_3} \end{pmatrix}$$

In order to find P^{-1} we need

$$\begin{aligned} \det|P| &= \frac{ad}{(a+\lambda_1)(a+d+\lambda_3)} + \frac{ad}{(a+\lambda_2)(a+d+\lambda_1)} \\ &+ \frac{ad}{(a+\lambda_3)(a+d+\lambda_2)} - \frac{ad}{(a+\lambda_1)(a+d+\lambda_2)} \\ &- \frac{ad}{(a+\lambda_2)(a+d+\lambda_3)} - \frac{ad}{(a+\lambda_3)(a+d+\lambda_1)} \\ &= \frac{ad(\lambda_3-\lambda_2)(a+\lambda_1)}{(a+\lambda_1)(a+\lambda_2)(a+\lambda_3)(a+d+\lambda_1)} \\ &+ \frac{ad(\lambda_1-\lambda_3)(a+\lambda_2)}{(a+\lambda_1)(a+\lambda_2)(a+\lambda_3)(a+d+\lambda_2)} \\ &+ \frac{ad(\lambda_2-\lambda_1)(a+\lambda_3)}{(a+\lambda_1)(a+\lambda_2)(a+\lambda_3)(a+d+\lambda_3)} \end{aligned}$$

From the characteristic equation (C.5) we have

$$\frac{ad(a+\lambda_i)}{a+d+\lambda_i} = (a+d+\lambda_i)(a+\lambda_i) - ad = \lambda_i^2 + (2a+d)\lambda_i + a^2 \qquad \text{(C.7)}$$

Therefore

$$\begin{aligned} \det|P| &= \frac{(\lambda_3-\lambda_2)\lambda_1^2 + (\lambda_1-\lambda_3)\lambda_2^2 + (\lambda_2-\lambda_1)\lambda_3^2}{(a+\lambda_1)(a+\lambda_2)(a+\lambda_3)} \\ &= \frac{(\lambda_1-\lambda_2)(\lambda_2-\lambda_3)(\lambda_3-\lambda_1)}{(a+\lambda_1)(a+\lambda_2)(a+\lambda_3)} \end{aligned} \qquad \text{(C.8)}$$

Besides we find

$$
\begin{aligned}
\mathrm{adj}(P) &= \begin{pmatrix}
\frac{d}{a+d+\lambda_3} - \frac{d}{a+d+\lambda_2} & \frac{ad}{(a+\lambda_3)(a+d+\lambda_2)} - \frac{ad}{(a+\lambda_2)(a+d+\lambda_3)} & \frac{a}{a+\lambda_2} - \frac{a}{a+\lambda_3} \\
\frac{d}{a+d+\lambda_1} - \frac{d}{a+d+\lambda_3} & \frac{ad}{(a+\lambda_1)(a+d+\lambda_3)} - \frac{ad}{(a+\lambda_3)(a+d+\lambda_1)} & \frac{a}{a+\lambda_3} - \frac{a}{a+\lambda_1} \\
\frac{d}{a+d+\lambda_2} - \frac{d}{a+d+\lambda_1} & \frac{ad}{(a+\lambda_2)(a+d+\lambda_1)} - \frac{ad}{(a+\lambda_1)(a+d+\lambda_2)} & \frac{a}{a+\lambda_1} - \frac{a}{a+\lambda_2}
\end{pmatrix} \\
&= \begin{pmatrix}
\frac{d\,(\lambda_2-\lambda_3)}{(a+d+\lambda_3)(a+d+\lambda_2)} & \frac{ad}{(a+\lambda_3)(a+d+\lambda_2)} - \frac{ad}{(a+\lambda_2)(a+d+\lambda_3)} & \frac{a\,(\lambda_3-\lambda_2)}{(a+\lambda_2)(a+\lambda_3)} \\
\frac{d\,(\lambda_3-\lambda_1)}{(a+d+\lambda_1)(a+d+\lambda_3)} & \frac{ad}{(a+\lambda_1)(a+d+\lambda_3)} - \frac{ad}{(a+\lambda_3)(a+d+\lambda_1)} & \frac{a\,(\lambda_1-\lambda_3)}{(a+\lambda_3)(a+\lambda_1)} \\
\frac{d\,(\lambda_1-\lambda_2)}{(a+d+\lambda_2)(a+d+\lambda_1)} & \frac{ad}{(a+\lambda_2)(a+d+\lambda_1)} - \frac{ad}{(a+\lambda_1)(a+d+\lambda_2)} & \frac{a\,(\lambda_2-\lambda_1)}{(a+\lambda_1)(a+\lambda_2)}
\end{pmatrix}
\end{aligned}
$$

Now we are ready to calculate the probability density function of the first passage time.

$$
\begin{aligned}
F(t) &= \vec{\alpha}\, P e^{t\,M} P^{-1} \nu \\
&= \vec{\alpha}\, P \begin{pmatrix} e^{\lambda_1 t} & 0 & 0 \\ 0 & e^{\lambda_2 t} & 0 \\ 0 & 0 & e^{\lambda_3 t} \end{pmatrix} \frac{\mathrm{adj}(P)}{\det|P|} \begin{pmatrix} 0 \\ 0 \\ a \end{pmatrix} \\
&= \frac{\vec{\alpha}}{\det|P|} \begin{pmatrix} \frac{a}{a+\lambda_1} & \frac{a}{a+\lambda_2} & \frac{a}{a+\lambda_3} \\ 1 & 1 & 1 \\ \frac{d}{a+d+\lambda_1} & \frac{d}{a+d+\lambda_2} & \frac{d}{a+d+\lambda_3} \end{pmatrix} \begin{pmatrix} \frac{(\lambda_3-\lambda_2)a^2 e^{\lambda_1 t}}{(a+\lambda_2)(a+\lambda_3)} \\ \frac{(\lambda_1-\lambda_3)a^2 e^{\lambda_2 t}}{(a+\lambda_3)(a+\lambda_1)} \\ \frac{(\lambda_2-\lambda_1)a^2 e^{\lambda_3 t}}{(a+\lambda_1)(a+\lambda_2)} \end{pmatrix} \\
&= \frac{\vec{\alpha}}{\det|P|} \begin{pmatrix} \frac{(\lambda_3-\lambda_2)\,a^3\,e^{\lambda_1 t}}{(a+\lambda_1)(a+\lambda_2)(a+\lambda_3)} + \frac{(\lambda_1-\lambda_3)\,a^3\,e^{\lambda_2 t}}{(a+\lambda_2)(a+\lambda_3)(a+\lambda_1)} + \frac{(\lambda_2-\lambda_1)\,a^3\,e^{\lambda_3 t}}{(a+\lambda_3)(a+\lambda_1)(a+\lambda_2)} \\ \frac{(\lambda_3-\lambda_2)\,a^2\,e^{\lambda_1 t}}{(a+\lambda_2)(a+\lambda_3)} + \frac{(\lambda_1-\lambda_3)\,a^2\,e^{\lambda_2 t}}{(a+\lambda_3)(a+\lambda_1)} + \frac{(\lambda_2-\lambda_1)\,a^2\,e^{\lambda_3 t}}{(a+\lambda_1)(a+\lambda_2)} \\ \frac{(\lambda_3-\lambda_2)\,a^2\,de^{\lambda_1 t}}{(a+d+\lambda_1)(a+\lambda_2)(a+\lambda_3)} + \frac{(\lambda_1-\lambda_3)\,a^2\,de^{\lambda_2 t}}{(a+d+\lambda_2)(a+\lambda_3)(a+\lambda_1)} + \frac{(\lambda_2-\lambda_1)\,a^2\,de^{\lambda_3 t}}{(a+d+\lambda_3)(a+\lambda_1)(a+\lambda_2)} \end{pmatrix}
\end{aligned}
$$

We then plug in the value of $\vec{\alpha}$ (C.3) and $\det|P|$ (C.8) to obtain the final result

$$\begin{pmatrix} F_{3,0}(t) \\ F_{3,1}(t) \\ F_{3,2}(t) \end{pmatrix} = \begin{pmatrix} -\frac{a^3 e^{\lambda_1 t}}{(\lambda_1-\lambda_2)(\lambda_3-\lambda_1)} - \frac{a^3 e^{\lambda_2 t}}{(\lambda_1-\lambda_2)(\lambda_2-\lambda_3)} - \frac{a^3 e^{\lambda_3 t}}{(\lambda_2-\lambda_3)(\lambda_3-\lambda_1)} \\ -\frac{a^2\ (a+\lambda_1) e^{\lambda_1 t}}{(\lambda_1-\lambda_2)(\lambda_3-\lambda_1)} - \frac{a^2\ (a+\lambda_2) e^{\lambda_2 t}}{(\lambda_1-\lambda_2)(\lambda_2-\lambda_3)} - \frac{a^2\ (a+\lambda_3) e^{\lambda_3 t}}{(\lambda_2-\lambda_3)(\lambda_3-\lambda_1)} \\ -\frac{a^2 d\ (a+\lambda_1) e^{\lambda_1 t}}{(a+d+\lambda_1)(\lambda_1-\lambda_2)(\lambda_3-\lambda_1)} - \frac{a^2 d\ (a+\lambda_2) e^{\lambda_2 t}}{(a+d+\lambda_2)(\lambda_1-\lambda_2)(\lambda_2-\lambda_3)} \\ -\frac{a^2 d\ (a+\lambda_3) e^{\lambda_3 t}}{(a+d+\lambda_3)(\lambda_2-\lambda_3)(\lambda_3-\lambda_1)} \end{pmatrix}$$

If we apply the relationship in (C.7)

$$ad\ (a+\lambda_i)/(a+d+\lambda_i) = \lambda_i^2 + (2a+d)\lambda_i + a^2$$

on $F_{3,2}(t)$, we can see that the results derived with the transition rate matrix are exactly the same as those derived with the backward master equations in the previous section (C.2).

Bibliography

[1] D. M. Raskin and P. A. J. de Boer. Minde-dependent pole-to-pole oscillation of division inhibitor minc in escherichia coli. *J. Bacteriol.*, 181(20):6419–6424, October 1999.

[2] J. Lutkenhaus. Assembly dynamics of the bacterial mincde system and spatial regulation of the z ring. *Annual Review of Biochemistry*, 76(1):539–562, 2007.

[3] K. C. Neuman, E. A. Abbondanzieri, R. Landick, J. Gelles, and S. M. Block. Ubiquitous transcriptional pausing is independent of rna polymerase backtracking. *Cell*, 115(4):437–447, November 2003.

[4] N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry, Third Edition (North-Holland Personal Library)*. North Holland, 3 edition, May 2007.

[5] C. Gardiner. *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences (Springer Series in Synergetics)*. Springer, 3rd edition, April 2004.

[6] Narendra S. Goel and N. Richter-Dyn. *Stochastic Models in Biology*. The Blackburn Press.

[7] Christian Commault. Phase-type distributions and representations: some results and open problems for system theory. *International Journal of Control*, 76(6):566–580, April 2003.

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

[9] J. Collier and L. Shapiro. Spatial complexity and control of a bacterial cell cycle. *Current Opinion in Biotechnology*, 18(4):333–340, August 2007.

[10] Thomas A. Leonard, P. Jonathan Butler, and Jan Löwe. Bacterial chromosome segregation: structure and dna binding of the soj dimer–a conserved biological switch. *The EMBO journal*, 24(2):270–282, January 2005.

[11] P. Deboer. A division inhibitor and a topological specificity factor coded for by the minicell locus determine proper placement of the division septum in e. coli. *Cell*, 56(4):641–649, February 1989.

[12] P. A. de Boer, R. E. Crossley, and L. I. Rothfield. Roles of minc and mind in the site-specific septation block mediated by the mincde system of escherichia coli. *J. Bacteriol.*, 174(1):63–70, January 1992.

[13] E. Bi and J. Lutkenhaus. Ftsz ring structure associated with division in escherichia coli. *Nature*, 354(6349):161–164, November 1991.

[14] E. Bi and J. Lutkenhaus. Cell division inhibitors sula and mincd prevent formation of the ftsz ring. *J. Bacteriol.*, 175(4):1118–1125, February 1993.

[15] J. Huang, C. Cao, and J. Lutkenhaus. Interaction between ftsz and inhibitors of cell division. *J. Bacteriol.*, 178(17):5080–5085, September 1996.

[16] David M. Raskin and Piet A. J. de Boer. Rapid pole-to-pole oscillation of a protein required for directing division to the middle of escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 96(9):4971–4976, 1999.

[17] Zonglin Hu, Edward P. Gogol, and J. Lutkenhaus. Dynamic assembly of mind on phospholipid vesicles regulated by atp and mine. *Proceedings of the National Academy of Sciences of the United States of America*, 99(10):6761–6766, 2002.

[18] Xuan-Chuan Yu and W. Margolin. Ftsz ring clusters in *min* and partition mutants: role of both the min system and the nucleoid in regulating ftsz ring localization. *Molecular Microbiology*, 32(2):315–326, 1999.

[19] Z. Hu, A. Mukherjee, S. Pichoff, and J. Lutkenhaus. The minc component of the division site selection system in escherichia coli interacts with

ftsz to prevent polymerization. *Proceedings of the National Academy of Sciences of the United States of America*, 96(26):14819–14824, 1999.

[20] S. Pichoff and J. Lutkenhaus. Escherichia coli division inhibitor mincd blocks septation by preventing z-ring formation. *J. Bacteriol.*, 183(22):6630–6635, November 2001.

[21] C. A. Hale, H. Meinhardt, and P. A. de Boer. Dynamic localization cycle of the cell division regulator mine in escherichia coli. *The EMBO journal*, 20(7):1563–1572, April 2001.

[22] P. A. de Boer, R. E. Crossley, A. R. Hand, and L. I. Rothfield. The mind protein is a membrane atpase required for the correct placement of the escherichia coli division site. *The EMBO journal*, 10(13):4371–4380, December 1991.

[23] David M. Raskin and Piet A. J. de Boer. The mine ring: An ftsz-independent cell structure required for selection of the correct division site in e. coli. *Cell*, 91(5):685–694, November 1997.

[24] Z. Hu and J. Lutkenhaus. Topological regulation of cell division in *escherichia coli* involves rapid pole to pole oscillation of the division inhibitor minc under the control of mind and mine. *Molecular Microbiology*, 34(1):82–90, 1999.

[25] Z. Hu, C. Saez, and J. Lutkenhaus. Recruitment of minc, an inhibitor of z-ring formation, to the membrane in escherichia coli: Role of mind and mine. *J. Bacteriol.*, 185(1):196–203, January 2003.

[26] S. L. Rowland, X. Fu, M. A. Sayed, Y. Zhang, W. R. Cook, and L. I. Rothfield. Membrane redistribution of the escherichia coli mind protein induced by mine. *J. Bacteriol.*, 182(3):613–619, February 2000.

[27] Z. Hu and J. Lutkenhaus. Topological regulation of cell division in e. colispatiotemporal oscillation of mind requires stimulation of its atpase by mine and phospholipid. *Molecular Cell*, 7(6):1337–1343, June 2001.

[28] X. Fu, Yu L. Shih, Y. Zhang, and Lawrence I. Rothfield. The mine ring required for proper placement of the division site is a mobile structure that changes its cellular location during the escherichia coli division cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 98(3):980–985, 2001.

[29] H. Meinhardt and Piet A. J. de Boer. Pattern formation in escherichia coli: A model for the pole-to-pole oscillations of min proteins and the localization of the division site. *Proceedings of the National Academy of Sciences of the United States of America*, 98(25):14202–14207, 2001.

[30] M. Howard, A. D. Rutenberg, and S. de Vet. Dynamic compartmentalization of bacteria: Accurate division in e. coli. *Physical Review Letters*, 87(27):278102+, December 2001.

[31] M. Howard and A. D. Rutenberg. Pattern formation inside bacteria: Fluctuations due to the low copy number of proteins. *Physical Review Letters*, 90(12):128102+, Mar 2003.

[32] M. Howard. A mechanism for polar protein localization in bacteria. *Journal of Molecular Biology*, 335(2):655–663, January 2004.

[33] F. Tostevin and M. Howard. A stochastic model of min oscillations in escherichia coli and min protein segregation during cell division. *Physical Biology*, 3(1):1+, March 2006.

[34] K. C. Huang, Y. Meir, and Ned S. Wingreen. Dynamic structures in escherichia coli: Spontaneous formation of mine rings and mind polar zones. *Proceedings of the National Academy of Sciences of the United States of America*, 100(22):12724–12728, 2003.

[35] K. C. Huang and N. S. Wingreen. Min-protein oscillations in round bacteria. *Physical Biology*, 1(4):229+, December 2004.

[36] K. Kruse. A dynamic model for determining the middle of escherichia coli. *Biophysical journal*, 82(2):618–627, February 2002.

[37] G. Meacci and K. Kruse. Min-oscillations in escherichia coli induced by interactions of membrane-bound proteins. *Physical Biology*, 2(2):89+, June 2005.

[38] R. A. Kerr, H. Levine, T. J. Sejnowski, and W. J. Rappel. Division accuracy in a stochastic model of min oscillations in escherichia coli.

Proceedings of the National Academy of Sciences of the United States of America, 103(2):347–352, 2006.

[39] D. A. Drew, M. J. Osborn, and L. I. Rothfield. A polymerization-depolymerization model that accurately generates the self-sustained oscillatory system involved in bacterial division site placement. *Proceedings of the National Academy of Sciences of the United States of America*, 102(17):6114–6118, 2005.

[40] N. Pavin, H. Č. Paljetak, and V. Krstić. Min-protein oscillations in escherichia coli with spontaneous formation of two-stranded filaments in a three-dimensional stochastic reaction-diffusion model. *Physical Review E*, 73(2):021904+, Feb 2006.

[41] D. Fange and J. Elf. Noise-induced min phenotypes in e. coli. *PLoS Comput Biol*, 2(6):e80+, June 2006.

[42] V. Krstić, Ž. Maglica, H. Č. Paljetak, B. Podobnik, and N. Pavin. Min-protein oscillations in e. coli: Three-dimensional off-lattice stochastic reaction-diffusion model. *Journal of Statistical Physics*, 128(1/2):5–20, July 2007.

[43] E. N. Cytrynbaum and B. D. Marshall. A multistranded polymer model explains minde dynamics in e. coli cell division. *Biophysical journal*, 93(4):1134–1150, August 2007.

[44] R. V. Kulkarni, K. C. Huang, M. Kloster, and N. S. Wingreen. Pattern formation within escherichia coli: Diffusion, membrane attachment, and self-interaction of mind molecules. *Physical Review Letters*, 93(22):228103+, November 2004.

[45] G. Meacci, J. Ries, E. Fischer-Friedrich, N. Kahya, P. Schwille, and K. Kruse. Mobility of min-proteins in escherichia coli measured by fluorescence correlation spectroscopy. *Physical Biology*, 3(4):255+, December 2006.

[46] Sidney Redner. *A Guide to First-Passage Processes*. Cambridge University Press, August 2001.

[47] K. Suefuji, R. Valluzzi, and D. RayChaudhuri. Dynamic assembly of mind into filament bundles modulated by atp, phospholipids, and mine. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26):16776–16781, December 2002.

[48] Yu-Ling Shih, T. Le, and L. Rothfield. Division site selection in escherichia coli involves dynamic redistribution of min proteins within coiled structures that extend between the two cell poles. *Proceedings of the National Academy of Sciences of the United States of America*, 100(13):7865–7870, June 2003.

[49] B. Houchmandzadeh, E. Wieschaus, and S. Leibler. Precise domain specification in the developing drosophila embryo. *Physical Review E*,

72(6):061920+, Dec 2005.

[50] K. Adelman, A. La Porta, T. J. Santangelo, J. T. Lis, J. W. Roberts, and M. D. Wang. Single molecule analysis of rna polymerase elongation reveals uniform kinetic behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 99(21):13538–13543, October 2002.

[51] R. Dalal, M. Larson, K. Neuman, J. Gelles, R. Landick, and S. Block. Pulling on the nascent rna during transcription does not alter kinetics of elongation or ubiquitous pausing. *Molecular Cell*, 23(2):231–239, July 2006.

[52] K. Herbert, A. Laporta, B. Wong, R. Mooney, K. Neuman, R. Landick, and S. Block. Sequence-resolved detection of pausing by single rna polymerase molecules. *Cell*, 125(6):1083–1094, June 2006.

[53] A. Saunders, L. J. Core, and J. T. Lis. Breaking barriers to transcription elongation. *Nature Reviews Molecular Cell Biology*, 7(8):557–567, August 2006.

[54] G. W. Muse, D. A. Gilchrist, S. Nechaev, R. Shah, J. S. Parker, S. F. Grissom, J. Zeitlinger, and K. Adelman. Rna polymerase is poised for activation across the genome. *Nature genetics*, 39(12):1507–1511, December 2007.

[55] J. Zeitlinger, A. Stark, M. Kellis, Joung-Woo 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):1512–1516, December 2007.

[56] R. Landick. The regulatory roles and mechanism of transcriptional pausing. *Biochemical Society Transactions*, 34:1062+, 2006.

[57] I. Artsimovitch and R. Landick. Pausing by bacterial rna polymerase is mediated by mechanistically distinct classes of signals. *Proceedings of the National Academy of Sciences of the United States of America*, 97(13):7090–7095, June 2000.

[58] K. M. Herbert, W. J. Greenleaf, and S. M. Block. Single-molecule studies of rna polymerase: Motoring along. *Annual Review of Biochemistry*, 77(1):149–176, 2008.

[59] D. Erie. The many conformational states of rna polymerase elongation complexes and their roles in the regulation of transcription. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1577(2):224–239, September 2002.

[60] I. Toulokhonov and R. Landick. The flap domain is required for pause rna hairpin inhibition of catalysis by rna polymerase and can modulate intrinsic termination. *Molecular Cell*, 12(5):1125–1136, November 2003.

[61] N. Komissarova and M. Kashlev. Rna polymerase switches between inactivated and activated states by translocating back and forth along the dna and the rna. *Journal of Biological Chemistry*, 272(24):15329–15338, June 1997.

[62] J. W. Shaevitz, E. A. Abbondanzieri, R. Landick, and S. M. Block. Backtracking by single rna polymerase molecules observed at near-base-pair resolution. *Nature*, 426(6967):684–687, November 2003.

[63] B. Ring, W. Yarnell, and J. Roberts. Function of e. coli rna polymerase factor- in promoter-proximal pausing. *Cell*, 86(3):485–493, August 1996.

[64] H. Tang, Y. Liu, L. Madabusi, and D. S. Gilmour. Promoter-proximal pausing on the hsp70 promoter in drosophila melanogaster depends on the upstream regulator. *Molecular and cellular biology*, 20(7):2569–2580, April 2000.

[65] I. Artsimovitch. The transcriptional regulator rfah stimulates rna chain synthesis after recruitment to elongation complexes by the exposed non-template dna strand. *Cell*, 109(2):193–203, April 2002.

[66] M. L. Kireeva, B. Hancock, G. H. Cremona, W. Walter, V. M. Studitsky, and M. Kashlev. Nature of the nucleosomal barrier to rna polymerase ii. *Molecular Cell*, 18(1):97–108, April 2005.

[67] C. Lee, X. Li, A. Hechmer, M. Eisen, M. D. Biggin, B. J. Venters, C. Jiang, J. Li, B. F. Pugh, and D. S. Gilmour. Nelf and gaga factor

are linked to promoter-proximal pausing at many genes in drosophila. *Molecular and cellular biology*, 28(10):3290–3300, May 2008.

[68] S. Nechaev, D. C. Fargo, G. dos Santos, L. Liu, Y. Gao, and K. Adelman. Global analysis of short rnas reveals widespread promoter-proximal stalling and arrest of pol ii in drosophila. *Science*, 327(5963):335–338, January 2010.

[69] V. M. Weake and J. L. Workman. Inducible gene expression: diverse regulatory mechanisms. *Nature Reviews Genetics*, 11(6):426–437, April 2010.

[70] A. N. Boettiger and M. Levine. Synchronous and stochastic patterns of gene activation in the drosophila embryo. *Science*, 325(5939):471–473, July 2009.

[71] B. M. Peterlin and D. H. Price. Controlling the elongation phase of transcription with p-tefb. *Molecular Cell*, 23(3):297–305, August 2006.

[72] David S. Gilmour. Promoter proximal pausing on genes in metazoans. *Chromosoma*, 118(1):1–10, February 2009.

[73] P. B. Rahl, C. Y. Lin, A. C. Seila, R. A. Flynn, S. McCuine, C. B. Burge, P. A. Sharp, and R. A. Young. c-myc regulates transcriptional pause release. 141(3):432–445, April 2010.

[74] D. A. Schafer, J. Gelles, M. P. Sheetz, and R. Landick. Transcription by single molecules of rna polymerase observed by light microscopy. *Nature*, 352(6334):444–448, August 1991.

[75] E. A. Abbondanzieri, W. J. Greenleaf, J. W. Shaevitz, R. Landick, and S. M. Block. Direct observation of base-pair stepping by rna polymerase. *Nature*, 438(7067):460–465, November 2005.

[76] R. Zhu, A. S. Ribeiro, D. Salahub, and S. A. Kauffman. Studying genetic regulatory networks at the molecular level: Delayed reaction stochastic models. *Journal of Theoretical Biology*, 246(4):725–745, June 2007.

[77] A. S. Ribeiro. Delayed stochastic model of transcription at the single nucleotide level. *Journal of Computational Biology*, 16(4):539–553, April 2009.

[78] T. Rajala, A. Häkkinen, S. Healy, O. Yli-Harja, and A. S. Ribeiro. Effects of transcriptional pausing on gene expression dynamics. *PLoS computational biology*, 6(3):e1000704+, March 2010.

[79] A. S. Ribeiro, A. Häkkinen, S. Healy, and O. Yli-Harja. Dynamical effects of transcriptional pause-prone sites. *Computational Biology and Chemistry*, May 2010.

[80] M. Voliotis. Fluctuations, pauses, and backtracking in dna transcription. *Biophysical Journal*, 94(2):334–348, January 2008.

[81] E. A. Galburt, S. W. Grill, A. Wiedmann, L. Lubkowska, J. Choy, E. Nogales, M. Kashlev, and C. Bustamante. Backtracking determines the force sensitivity of rnapii in a factor-dependent manner. *Nature*, 446(7137):820–823, March 2007.

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

[83] H. Yin, R. Landick, and J. Gelles. Tethered particle motion method for studying transcript elongation by a single rna polymerase molecule. *Biophysical journal*, 67(6):2468–2478, December 1994.

[84] M. D. Wang, M. J. Schnitzer, H. Yin, R. Landick, J. Gelles, and S. M. Block. Force and velocity measured for single molecules of rna polymerase. *Science*, 282(5390):902+, October 1998.

[85] R. J. Davenport, G. J. L. Wuite, R. Landick, and C.s Bustamante. Single-molecule study of transcriptional pausing and arrest by e. coli rna polymerase. *Science*, 287(5462):2497–2500, March 2000.

[86] M. L. Kireeva and M. Kashlev. Mechanism of sequence-specific pausing of bacterial rna polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, 106(22):8900–8905, June 2009.

[87] D. Hawley. Mechanism of activation of transcription initiation from the prm promoter. *Journal of Molecular Biology*, 157(3):493–525, May 1982.

[88] H. Buc and W. R. McClure. Kinetics of open complex formation between escherichia coli rna polymerase and the lac uv5 promoter. evidence for a sequential mechanism involving three steps. *Biochemistry*, 24(11):2712–2723, May 1985.

[89] W. R. McClure. Mechanism and control of transcription initiation in prokaryotes. *Annual Review of Biochemistry*, 54(1):171–204, 1985.

[90] N. Mitarai, I. B. Dodd, M. T. Crooks, and K. Sneppen. The generation of promoter-mediated transcriptional noise in bacteria. *PLoS Comput Biol*, 4(7):e1000109+, July 2008.

[91] D. B. Craig, E. A. Arriaga, J. C. Y. Wong, H. Lu, and N. J. Dovichi. Studies on single alkaline phosphatase molecules: reaction rate and activation energy of a reaction catalyzed by a single molecule and the effect of thermal denaturationthe death of an enzyme. *Journal of the American Chemical Society*, 118(22):5245–5253, January 1996.

[92] X. S. Xie and H. P. Lu. Single-molecule enzymology. *Journal of Biological Chemistry*, 274(23):15967–15970, June 1999.

[93] S. Weiss. Fluorescence spectroscopy of single biomolecules. *Science*, 283(5408):1676+, March 1999.

[94] T. Funatsu, Y. Harada, M. Tokunaga, K. Saito, and T. Yanagida. Imaging of single fluorescent molecules and individual atp turnovers by single

myosin molecules in aqueous solution. *Nature*, 374(6522):555–559, April 1995.

[95] L. Edman, Z. Foldespapp, S. Wennmalm, and R. Rigler. The fluctuating enzyme: a single molecule approach. *Chemical Physics*, 247(1):11–22, August 1999.

[96] Q. Xue and E. S. Yeung. Differences in the chemical reactivity of individual molecules of an enzyme. *Nature*, 373(6516):681–683, February 1995.

[97] H. P. Lu, L. Xun, and X. S. Xie. Single-molecule enzymatic dynamics. *Science*, 282(5395):1877–1882, December 1998.

[98] A. Molski. Single-molecule michaelismenten kinetics: Effect of substrate fluctuations. *Chemical Physics*, 352(1-3):276–280, September 2008.