A Novel Computational Method for Two-State Transcription Model with Non-Exponential ON and OFF Durations

Manyi Zheng, Zhishan Qiu, Feng Jiao* and Qiwen Sun*

Guangzhou Center for Applied Mathematics, Guangzhou University, Guangzhou 510006, P.R. China

Abstract. The fluctuation of mRNA molecule numbers within an isogenic cell population is primarily attributed to randomly switching between active (ON) and inactive (OFF) periods of gene transcription. In most studies the waiting-times for ON or OFF states are modeled as exponential distributions. However, increasing data suggest that the residence durations at ON or OFF are non-exponential distributed for which the traditional master equations cannot be presented. By combining Kolmogorov forward equations with alternating renewal processes, we present a novel method to compute the average transcription level and its noise by circumventing the bottleneck of master equations under gene ON and OFF switch. As an application, we consider lifetimes of OFF and ON states having Erlang distributions. We show that: (i) multiple steps from OFF to ON force the oscillating transcription while multiple steps from ON to OFF accelerate the transcription, (ii) the increase of steps between ON and OFF rapidly reduces the transcription noise to approach its minimum value. This suggests that a large number of steps between ON and OFF are not needed in the model to capture the stochastic transcription data. Our computation approach can be further used to treat a series of transcription cycles which are non-lattice distributed.

1 Introduction

In both prokaryotes and eukaryotes, gene transcription is the core process in the transmission of genetic information, which flows from DNA to RNA to protein in single cells [22]. The new in vivo RNA detection technique, such as single-cell fluorescence

---

*Corresponding author. *Email addresses:* `jiaof@gzhu.edu.cn` (F. Jiao), `qwsun@gzhu.edu.cn` (Q. Sun)

microscopy and synthetic genetic constructs, has allowed real-time monitoring of transcription events in individual living cells [1, 24, 34]. In experiments, mRNA synthesis is monitored to be random and discontinuous [17, 37]. Randomness in transcription leads to highly variable mRNA distributions, resulting in phenotypic heterogeneity in a cell population [20, 32].

Mathematical models have been built to characterize gene transcription and explore its stochastic regulation [6,30,39,44,47]. In the classical two-state transcription model [30], the promoter is thought to switch randomly between two fundamental states: active and inactive, manifested by the observed transcriptional burst occurring in short-lived active states interspersed by long-lived inactive intervals [27, 37]. To explore the mechanisms that regulate stochastic mRNA production in response to environmental changes, the three-state transcription model instead of a two-state model was proposed to describe the transcription process [4, 39]. In this model, the OFF state is composed of two sub-states connected in series, and both obey the exponential distribution. The assertion was then validated in experiments [11, 37], and was generalized to different models with multiple sub-OFF or sub-ON states [28, 42, 47], and multiple signaling pathways [21, 35, 36]. Also, some complex models and methods were established to estimate system parameters or calculate the probability distribution functions of transcripts [3, 5, 14, 15, 28, 31, 45].

In most models mentioned above, it was assumed that the switching rates between different transcriptional states are constants. In other words, the state switching can be governed by a first-order linear differential equation with a constant coefficient. In the two-state transcription model, the promoter is assumed to switch stochastically between an OFF state and an ON state at rates k_{OFF} and k_{ON}, respectively. The two rates k_{OFF} and k_{ON} are usually defined as

$$k_{OFF} = \frac{1}{T_{ON}}, \quad k_{ON} = \frac{1}{T_{OFF}}, \tag{1.1}$$

where T_{ON} is the average duration of a burst and T_{OFF} is the average time between two consecutive bursts. Since the rates k_{OFF} and k_{ON} are constants, the lifetimes of the OFF and the ON states should have the exponential distribution. From the memoryless property of the exponential distributions, it follows that the two-state transcription process in fact is a continuous-time Markov chain. Other models built in this way also have the Markov property. It is the Markov property that allows us to obtain the chemical master equations of gene transcription.

With the help of real-time monitoring, increasing experiments indicate that the lifetimes of the ON and/or the OFF states are not always exponentially distributed [12, 27, 37, 40]. In recent years, non-Markovian processes have attracted increasing interest [7, 18, 23, 43]. Using single-cell time-lapse bioluminescence imaging, Suter *et al.* [37] monitored transcription kinetics of some genes in mouse fibroblasts and found that the duration of the OFF state should be described by summing two sequential exponential processes. How to bypass the molecular memory in the reaction process to establish an appropriate transcriptional model is crucial for the study of stochastic models.