Stochastic Transcription Elongation via Rule Based Modelling

Masahiro Hamano[1,2]

Graduate School of Information Science and Technology
University of Tokyo, JAPAN

Abstract

Transcription elongation is the mechanism by which RNA polymerase (RNAP) moves along template un-zipped DNA and synthesizes a complementary single-stranded RNA. During the elongation, RNAP forms a stable transcription elongation complex (TEC) with the template DNA and the nascent RNA. The mechanism involves back-tracked and forward-tracked modes of TEC and the polymerisation and depolymerisation of RNA. To capture the stochasticity of the elongation, we describe the mechanism in terms of rule-based modelling through the TEC's local window frame of adjacent active sites. In this way, we can uniformly derive the variations of known kinetic pathways for various interaction combinations of TEC's active sites. From the compact interactions at local sites, we find abstracted rules for the elongation. As the semantic counterpart, we derive quasi-steady state approximations to the chemical master equations. The stochastic models are thermodynamically interpreted as the free energy distributions of agents with variant configurations.

Keywords: RNA polymerase, Transcription Elongation Complex, Master Equation, Rule-Based Modelling, Brownian Ratchet Model, Equilibrium Kinetics, Chemical Equilibrium, Steady State Dynamics, Quasi-Steady State, Michaelis-Menten Kinetics, Free Energy, Boltzmann Distribution

1 Introduction

The stochasticity of gene regulation and expression, which is intrinsic to discreteness and small numbers of molecules partaking in and regulating biological events [19], has been successfully modelled by stochastic process calculi such as stochastic π [22] and rule based κ [7,11]. Such models represent each constituent of the biochemical process by stochastic interactions among agents. Agents interact at biochemical sites through channels at rates modelled by exponential time distributions. As a syntactical rule, this description yields a semantic counterpart of a Markov process, such that each single-step interaction has a Poisson waiting time. Similar modelling has been applied in queue theory [10] and more generally in the theory of stochastic petri nets [9].
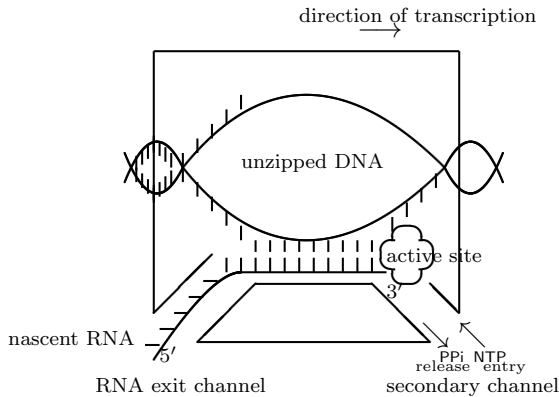
---

direction of transcription

Fig. 1. Structure of TEC

In this paper, we examine whether stochastic rule-based modelling captures multi-step molecular interactions. As biochemical interactions are non-uniform and occur at different rates, they are characterized by stochastic fluctuations, non-Poisson time dynamics and decay of particular biological events. For this purpose, we model *transcription*, which is a typical molecular multi-step interaction. As the first phase of gene expression, transcription involves a tiny minority of the molecules reacting in cells. Thus, for modelling transcription, a rule-based stochastic method is preferred over the deterministic method of mass action laws. Rule-based modelling directly describes the stochastic and discrete nature of chemical reactions in transcriptional elongation. The primitive agents are the hundreds (or thousands) of base pairs contained in a single gene, and a similar number of nucleoside triphosphates (NTPs) that join the transcript via RNA polymerase. The present author applied a similar nucleotide-based modelling concept to multiple branching processes in RNA interference [15].

Transcription (mRNA synthesis) from the corresponding unzipped single-stranded DNA template proceeds in three sequential stages: *Initiation*, *Elongation* and *Termination*. The main stage is elongation. RNAP is a mechano-chemical coupling mechanism that converts the chemical energy derived from NTP hydrolysis into mechanical work, together with random Brownian motion involving back-tracked and forward-tracked modes. The RNAP kinetics and mechano-chemical motions are mediated by the *Transcription Elongation Complex* (TEC) [18], formed by the combination of RNAP, the template DNA and the nascent RNA (cf. Fig 1). The principal kinetic feature is the mechano-chemical cycle of NTP binding: Nucleotide incorporation induces the pre-translocation state of TEC. The TEC advances by 1 bp along the DNA, thereby returning to post-translocation. Initiation is much faster than elongation; moreover, during the elongation phase, the mRNA synthesis is often disrupted by transcriptional pauses, which are linked to the reverse translocation of RNA along the DNA. Therefore, stochastic fluctuations and time delays are characteristic of transcriptional regulation.

Based on recent single-molecule experiments (such as magnetic and optical tweezers and single-molecule fluorescence), researchers have proposed theoretical biochemical models to explain the elongation kinetics in terms of the individual

biochemical reactions that dominate the behaviour [1,2,6,14,21,23,27,33]. Theoretical transcription models have focused on the RNA-mediated coupling of chemical catalysis energy to translocation and mechanical work. The present paper surmises that rule-based modelling conveniently captures the fundamentals of the dynamism; namely, the *thermodynamic stability* of TEC as the elongation proceeds [2,33]. Our rule-based framework represents biochemical interactions through the adjacent active sites of TEC, and captures the stability of multiple variant forms of TEC arising from interactions during mechano-chemical movement. As a semantic counterpart to our rule-based description, we uniformly derive systems of stochastic master equations that describe stochastic models [17,24,28,30,31] of elongation in special cases. Moreover, we naturally derive the probability distribution of the various TEC forms. Consequently, the thermodynamical stability of TEC along DNA and RNA is given by an energy function.

The remainder of this paper is organised as follows. Section 2 explains how syntactical rule-based modelling uniformly describes the biochemical pathways of elongation. The description considers enzyme reactions through certain active sites and the binding/unbinding interactions in polymerisation/depolymerisation. Section 3 develops the semantic counterpart of Section 2. In this section, the time-evolutions of probabilities among agents with varying configurations are modelled by systems of stochastic master equations. The constructed models elucidate the thermodynamical meaning of the kinetic pathways in the mechano-chemical reactions by means of the probability distributions and convergence to equilibrium. Moreover, we characterize the biochemically popular quasi-steady state approximation [16,26], and compare it with the semantics abstracted from the master equation. The approximation suggests that the finer semantics can be retrieved from coarser grained semantics arising in the rule-based abstraction. In Section 4, our stochastic rule-based method is shown to incorporate Michaelis-Menten enzyme kinetics under chemical equilibrium conditions. Moreover, our agent model is shown to naturally provide a thermodynamical and kinetic modelling in terms of the free energy distribution.

2 Elongation Pathways as Site Interactions inside TEC

First, we introduce the minimal common mechanism that underlies various elongation pathways mediated by TEC. This mechanism is crucial for describing various biochemical kinetic pathways in terms of the NTP incorporation cycle, which proceeds as follows:

Pre- and post-translocation [2,3,18]

Elongation is the polymerisation step, which adds a nucleotide to the $3'$ end of the nascent RNA. This reaction is catalysed by the active site of RNAP. After the polymerisation, TEC immediately enters the *pre-translocation* state, and translates forward along the DNA template until the active site is correctly positioned to catalyse the next nucleotide addition. *Pre-translocation* is the state immediately
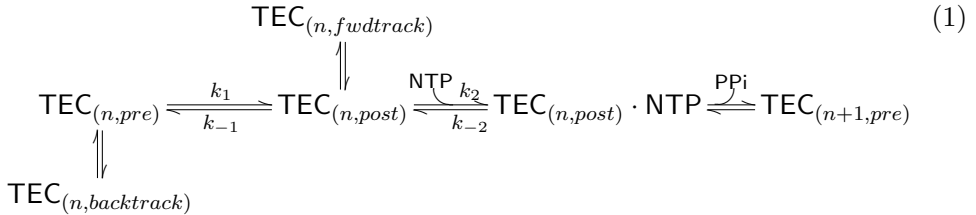
following the forward location of TEC and polymerisation of the next nucleotide. Only when the TEC is post-translocated, the RNAP's active site is adjacent to the $3'$-end of the nascent RNA, enabling binding and incorporation of the next nucleotide.

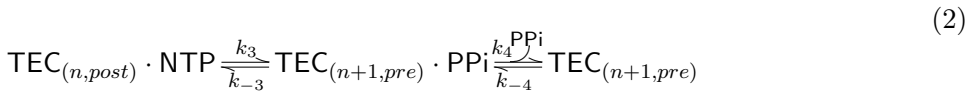**Back- and forward-tracking [2,3,12]**

The forward polymerisation of the RNA transcript is disrupted by pauses of the RNAP (known as transcriptional pausing [18]). The major cause of pausing is *back-track*, in which the active site of RNAP moves backward along the DNA template and RNA transcript without depolymerising the transcript. The opposite movement is *forward-track* (also called *hyper translocation*) towards the $5'$ end of the template DNA.

Using the above terminologies, we develop a a fundamental scheme for the elongation mechanism. Let $\mathsf{TEC}_{(n,*)}$ represent a TEC with transcript size n at stage $*$ (hereafter, we abbreviate the terms pre- and post-translocation to pre and post, respectively). The pathway converts chemical energy into the mechanical energy required to move RNAP in the backward and forward directions

Brownian Ratchet Model [2,3]

$$
\begin{array}{c}
\mathsf{TEC}_{(n,fwdtrack)} \hspace{3cm} (1)\\[1ex]
\Big\Uparrow\Big\downarrow\\[1ex]
\mathsf{TEC}_{(n,pre)} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathsf{TEC}_{(n,post)} \overset{\mathsf{NTP}\; k_2}{\underset{k_{-2}}{\rightleftharpoons}} \mathsf{TEC}_{(n,post)}\cdot \mathsf{NTP} \overset{\mathsf{PPi}}{\rightleftharpoons} \mathsf{TEC}_{(n+1,pre)}\\[1ex]
\Big\Uparrow\Big\downarrow\\[1ex]
\mathsf{TEC}_{(n,backtrack)}
\end{array}
$$

The first reaction is translocation and the central reaction is NTP binding and recognition, when the active site locates at the growing tip of the mRNA. The right-most reaction combines two reactions; NTP hydrolysis and the release of pyrophosphate PPi:

$$
\mathsf{TEC}_{(n,post)}\cdot \mathsf{NTP} \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} \mathsf{TEC}_{(n+1,pre)}\cdot \mathsf{PPi} \overset{k_4\,\mathsf{PPi}}{\underset{k_{-4}}{\rightleftharpoons}} \mathsf{TEC}_{(n+1,pre)} \hspace{1.5cm} (2)
$$

This combination assumes that the NTP hydrolysis rates are much larger than the PPi release rate; thus, the separate steps can be treated as one step with a single effective rate.

The pathway (1) is described by a rule-based modelling using κ calculus. The agents and rules are specified as follows:

Agents

A *nucleotide* $N(\ell, r)$ is a monomeric subunit with phosphate bonds ℓ and r. Binding of both sites is represented by a common superscript.

The agent W represents a *window frame* of TEC's multi-partite active sites (P) and (A), where the (P) and (A) sites hold the RNA 3'-nucleotide and the substrate NTP

[18], respectively. W possesses a site for PPi association. As the window frame, the agent W slides back and forth along the template DNA and RNA transcript, mimicking the movement of the active site. The *Null agent* is denoted by $\emptyset$. The substrate agent NTP is parallel to N but lacks a bonding site. See Fig 2.
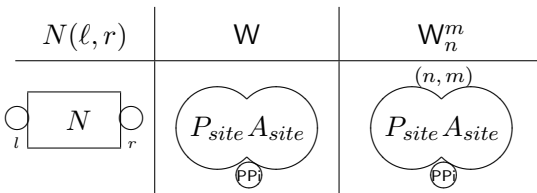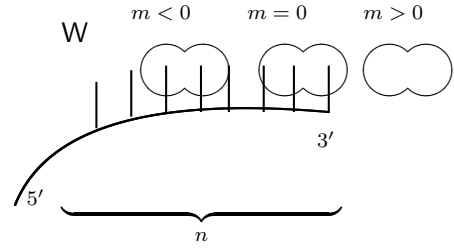

Fig. 2. Agents

Fig. 3. Position (n, m) of W

These agents discriminate among post-translocation, pre-translocation and loading of TEC as follows:

post-translocation: $\quad \mathsf{W}\left[N(r), \emptyset\right] \quad$ schematically $\left(N_{\mathrm{OH}}\right)$

pre-translocation: $\quad \mathsf{W}\left[N(r^1), N(\ell^1, r)\right]$ schematically $\left(N{-}N_{\mathrm{OH}}\right)$

loading: $\quad \mathsf{W}\left[N(r), \mathsf{NTP}\right] \quad$ schematically $\left(N_{\mathrm{OH}}\,\mathsf{NTP}\right)$

The OH subscript of N designates the $3'$ OH in the next nucleotide reaction that extends the RNA chain. In other words, OH indicates that the r site of N is not bounded. In the pre-translocation state, the line joining N and N_{OH} designates a phosphate bond between the r and ℓ sites respectively of N and N_{OH}.

An additional site is augmented for counting the position $(n, m) \in \mathbb{N} \times \mathbb{Z}$, where n indicates the length of the nascent RNA (the position of either the last nucleotide or the loaded NTP) and m indicates the back ($m < 0$) and forward ($m > 1$) track positions. Although redundant for $m = 0$ and 1, the position $m = 0$ (resp. $m = 1$) indicates a pre-translocated (resp. post-translocated) position. See Fig 3.

We next introduce three kinds of rules, $\mathsf{Elong}_{\mathbb{N}}^{\mathbb{Z}}$, $\mathsf{Elong}_{\mathbb{N}}$ and Elong, which dictate the degree of abstraction of agent W. That is, forgetting the counter, we abstract W_n^m first to W_n, and then to W. Let us start with the finest grained rules $\mathsf{Elong}_{\mathbb{N}}^{\mathbb{Z}}$.
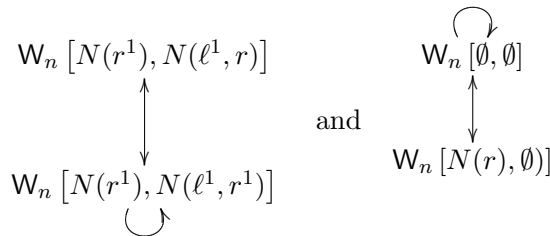
Rules of $\mathsf{Elong}_{\mathbb{N}}^{\mathbb{Z}}$

$$\vdots$$

$$\mathsf{W}_n^3\,[\emptyset, \emptyset]$$

$$\mathsf{W}_n^2\,[\emptyset, \emptyset]$$

$$\mathsf{W}_n\,\big[N(r^1), N(\ell^1, r)\big] \longleftrightarrow \mathsf{W}_n\,[N(r), \emptyset] \longleftrightarrow \quad \diamondsuit \quad \text{of the bottom}$$

$$\mathsf{W}_n^{-1}\,\big[N(r^1), N(\ell^1, r^1)\big]$$

$$\mathsf{W}_n^{-2}\,\big[N(r^1), N(\ell^1, r^1)\big]$$

These rules directly correspond to pathway (1), so the vertical and horizontal rules respectively correspond to back/forward-tracking and the main reactions of the pathway.

$$\vdots$$

The rules $\diamondsuit$ correspond directly to the pathway (2):
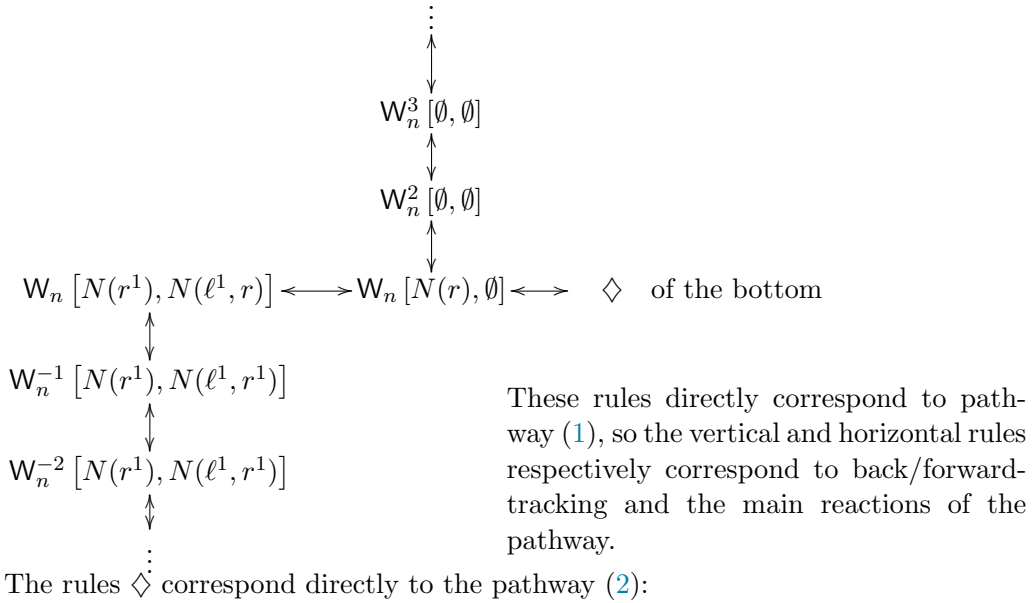
$$\diamondsuit \qquad \mathsf{W}_{n+1}\,[N(r), \mathsf{NTP}] \longleftrightarrow \mathsf{W}_{n+1}^{\mathsf{PPi}}\,\big[N(r^1), N(\ell^1, r)\big] \longleftrightarrow \mathsf{W}_{n+1}\,\big[N(r^1), N(\ell^1, r)\big]$$

Fig. 5. $\mathsf{Elong}_{\mathbb{N}}^{\mathbb{Z}}$

Forgetting the counter m of W_n^m for W_n, the horizontal pathways of forward- and back-tracking respectively reduce to:

Rules of $\mathsf{Elong}_{\mathbb{N}}$

$$\mathsf{W}_n\,\big[N(r^1), N(\ell^1, r)\big] \qquad\qquad \mathsf{W}_n\,[\emptyset, \emptyset]$$

$$\text{and}$$

$$\mathsf{W}_n\,[N(r), \emptyset)]$$

$$\mathsf{W}_n\,\big[N(r^1), N(\ell^1, r^1)\big]$$

Finally, if agent W is expressed in its simplest form without the counter, the above rules are abstracted as follows. These rules explicitly express the *mechano-chemical cycle* of NTP during the elongation stage:

Rules of Elong

A complex version of the above simple Brownian ratchet model allows NTP binding to both post-translocated and pre-translocated positions (**Complex Brownian Ratchet Model with an Auxiliary Binding Site** [3]).

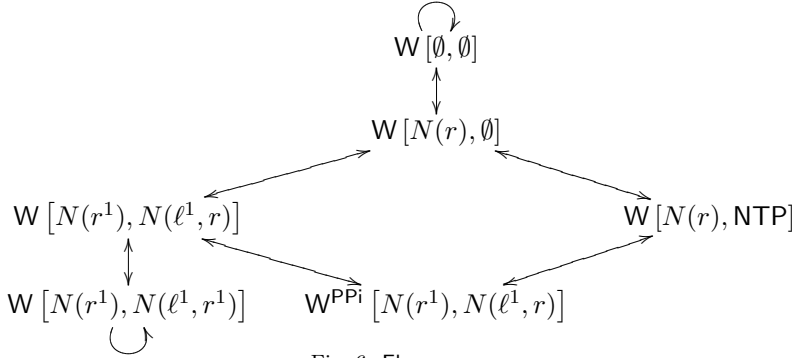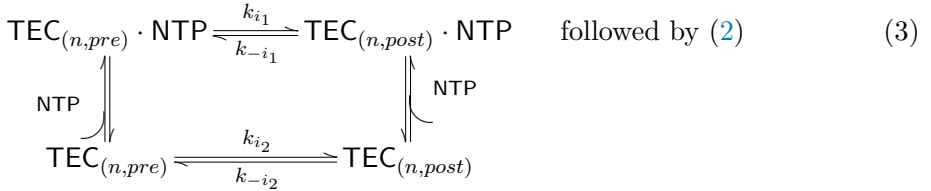$$\mathsf{W}\,[\emptyset, \emptyset]$$

$$\mathsf{W}\,[N(r), \emptyset]$$

$$\mathsf{W}\,\big[N(r^1), N(\ell^1, r)\big] \qquad\qquad \mathsf{W}\,[N(r), \mathsf{NTP}]$$

$$\mathsf{W}\,\big[N(r^1), N(\ell^1, r^1)\big] \qquad \mathsf{W}^{\mathsf{PPi}}\,\big[N(r^1), N(\ell^1, r)\big]$$

Fig. 6. Elong

$$\mathsf{TEC}_{(n,pre)} \cdot \mathsf{NTP} \underset{k_{-i_1}}{\overset{k_{i_1}}{\rightleftharpoons}} \mathsf{TEC}_{(n,post)} \cdot \mathsf{NTP} \qquad \text{followed by (2)} \qquad\qquad (3)$$

$$\mathsf{NTP}\Big\Downarrow \qquad\qquad\qquad \Big\Downarrow \mathsf{NTP}$$

$$\mathsf{TEC}_{(n,pre)} \underset{k_{-i_2}}{\overset{k_{i_2}}{\rightleftharpoons}} \mathsf{TEC}_{(n,post)}$$

When an (E) site for NTP-entry can be augmented to W, the complex ratchet model is realised by the following rules. Thereby, the complex pathway (3) is realised by adding one auxiliary site to the agent W.

(Agent)	(Rules)
$\mathsf{W} = \big(P_{site}\,A_{site}\,E_{site}\big)$	$\mathsf{W}_n\,\big[N(r^1), N(\ell^1, r), \mathsf{NTP}\big] \longleftrightarrow \mathsf{W}_n\,[N(r), \mathsf{NTP}, \emptyset]$ $\mathsf{W}_n\,\big[N(r^1), N(\ell^1, r), \emptyset\big] \longleftrightarrow \mathsf{W}_n\,[N(r), \emptyset, \emptyset]$

3 Semantics and Abstraction

This section constructs the semantics of the rule-based description of the elongation presented in Section 2. Using master equations, three semantics are constructed for $\mathsf{Elong}_{\mathbb{N}}^{\mathbb{Z}}$, $\mathsf{Elong}_{\mathbb{N}}$ and Elong. We proceed from the finest grained semantics of $\mathsf{Elong}_{\mathbb{N}}^{\mathbb{Z}}$, through the coarser grained semantics of $\mathsf{Elong}_{\mathbb{N}}$ and finally to the coarsest semantics of Elong, by progressively simplifying (abstracting) the construction. Our stochastic semantics of the chemical master equations describes the time evolution of the agent Ws with various configurations, governed by the rules of Section 2. By deriving the master equations in tight accordance with the rule description, we obtain a uniform comprehension of existing mathematical elongation models [17,24,28,30,31].

We emphasize here that all rules of Section 2 are local; the chemical interactions are independent of the global conditions of agents lying outside the TEC's active sites. In other words, we need not specify the contextual details of agents outside the window frame represented by W. (Such details include the polymerisation/depolymerisation of contiguous nucleotides downstream from W to the 5′-end of RNA, and their hybridisation to the template DNA.) Local interactions via adjacent sites inside the window W effectively capture the biochemical pathways. This
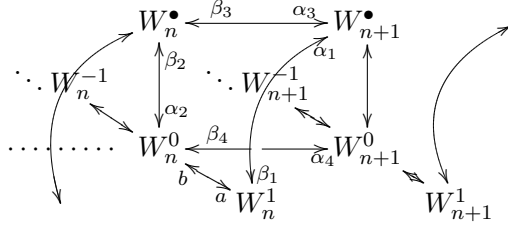
Fig. 7. Local state transitions of $\mathsf{Elong}_{\mathbb{N}}^{\mathbb{Z}_-}$

locality allows us to construct a compact system of master equations, avoiding combinatorial blow-up of the number of transition-states, and ensuring a tractable master equation for extracting the means and approximations [24]. General master equations are difficult to solve directly, but our equations have finite dimensions because the TEC dynamics involve small populations of molecules. Biologically, the locality reflects the thermodynamic stability of TEC with respect to the DNA template and the RNA transcript [32].

3.1 Master Equation Semantics

We start this subsection with a convention imposed by the locality of the rules: The agents Ws at the back ($m > 2$) and forward ($m < 0$) tracked positions are sufficiently determined by their counting states at site (n, m) of W_n^m. Similarly, in the main reaction pathway, we stipulate that $m = 1$ (resp. $m = 0$) designates pre-translocation (resp. post-translocation) as follows:

$$W_n^0 = \mathsf{W}_n\left[N(r^1), N(\ell^1, r)\right] \qquad W_n^1 = \mathsf{W}_n\left[N(r), \emptyset\right] \qquad W_n^\bullet = \mathsf{W}_n\left[N(r), \mathsf{NTP}\right]$$

These W_n^j's with varying (n, j) constitute the state transitions representing TEC's local interactions during elongation.

The state transitions are locally described by the stereographic diagram of Fig 7 [3]. The transition probabilities are identified by the rate constants of their corresponding reactions in pathways (1) and (2); that is, $a = k_1$ $b = k_{-1}$, $\alpha_1 = k_2$, $\beta_1 = k_{-2}$, $\alpha_2 = k_4 k_3$, and $\beta_2 = k_{-3} k_{-4}$.

In the stereographic diagram, the vertical and horizontal arrows represent loading and unloading of NTP and polymerisation/depolymerisation, respectively. The arrows projecting out of and into the plane of the paper represent back-track and forward-track, respectively. Because TEC is biologically stable, we assume that depolymerisation occurs only in the post-translocation state.

Fig 8 is the global projection of Fig 7 onto a plane. The boundary states are indicated.

The finest grained semantics for the finest rule $\mathsf{Elong}_{\mathbb{N}}^{\mathbb{Z}_-}$ of Section 2 is now derived as master equations:

---

[3]   To simplify the model construction, we hereafter omit the forward-tracking pathway, and consider $\mathsf{Elong}_{\mathbb{N}}^{\mathbb{Z}_-}$ (rather than $\mathsf{Elong}_{\mathbb{N}}^{\mathbb{Z}}$) where $\mathbb{Z}_-$ is the set of negative integers.
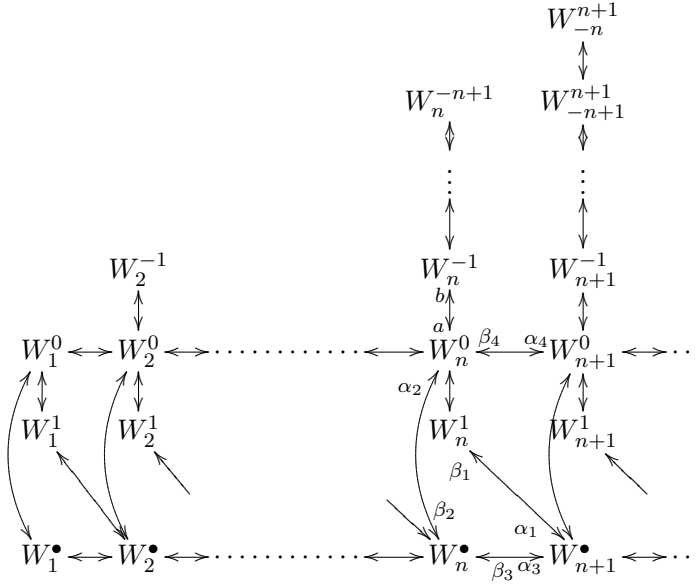
Fig. 8. Global diagram of Fig 7 for $\mathsf{Elong}_{\mathbb{N}}^{\mathbb{Z}-}$

Semantics for $\mathsf{Elong}_{\mathbb{N}}^{\mathbb{Z}-}$

For $j \in \{0, 1, \bullet\}$, let $w_n^{[j]}$ be the probability $\pi(\mathsf{W}_n^j, t)$ of finding W_n^j at time t. The master equation [29] is the gain-loss equation describing the probabilities $w_n^{[j]}$ among separate states W_n^j. In the following equations, the first term on the right-hand side is the gain probability that the system arrives at the state from the left. The second right-hand term is the loss probability that the system leaves the state for the left.

Evolution along the main pathway

$$\frac{d}{dt}w_n^{[0]} = \left(aw_n^{[-1]} + \alpha_4 w_{n-1}^{[0]} + \beta_4 w_{n+1}^{[0]} + \alpha_2 w_n^{[\bullet]} + bw_n^{[1]} \right)$$
$$- (\alpha_4 + \beta_2 + a + b + \beta_4)w_n^{[0]} \tag{4}$$

$$\frac{d}{dt}w_n^{[\bullet]} = \left(\alpha_3 w_{n-1}^{[\bullet]} + \beta_3 w_{n+1}^{[\bullet]} + \beta_2 w_n^{[0]} + \alpha_1 w_{n-1}^{[1]} \right)$$
$$- (\alpha_3 + \beta_3 + \alpha_2 + \beta_1)w_n^{[\bullet]} \tag{5}$$

$$\frac{d}{dt}w_n^{[1]} = \left(\beta_1 w_{n+1}^{[\bullet]} + aw_n^{[0]} \right) - (b + \alpha_1)w_n^{[1]} \tag{6}$$

Evolution along the back-track pathway

$$\frac{d}{dt}w_n^{[-(n-1)]} = bw_n^{[-(n-2)]} - aw_n^{[-(n-1)]} \tag{7}$$

$$\frac{d}{dt}w_n^{[-j]} = \left(aw_n^{[-(j+1)]} + bw_n^{[-(j-1)]} \right) - (a+b)w_n^{[-j]} \qquad \text{for } 1 \le j < n - 1 \tag{8}$$

Normalization constraint

$$\sum_{(n,j)} \left(w_n^{[0]} + w_n^{[1]} + w_n^{[\bullet]} + w_n^{[-j]} \right) = 1 \tag{9}$$

$$\text{where} \quad (n, -j) \in \mathbb{N} \times \mathbb{Z}_- \quad s.t. 1 \leq j \leq n - 1 \tag{10}$$

Semantics for $\mathsf{Elong}_{\mathbb{N}}$

In the rules of abstracted $\mathsf{Elong}_{\mathbb{N}}$, we drop the superscripts $-j$'s of the W_n^{-j}s for $j \geq 1$. Then, by a similar convention to that specified at the beginning of Section 3.1, we have $W_n^- = \mathsf{W}_n \left[N(r^1), N(\ell^1, r^1) \right]$, where the abstracted W_n^{-j}'s are expressed as W_n^-. The probability $w_n^{[-]}$ of finding W_n^- at time t and the corresponding master equation are then given as

$$w_n^{[-]} = \sum_{1 \leq j \leq n-1} w_n^{[-j]} \quad \text{and} \quad \frac{d}{dt} w_n^{[-]} = \sum_{1 \leq j \leq n-1} \frac{d}{dt} w_n^{[-j]} = b w_n^{[0]} - a w_n^{[-1]} \tag{11}$$

The time evolution of $w_n^{[-]}$ of (11) is obtained by summing equations (7) and (8) describing the back-track pathways. By replacing (7) and (8) with (11), we extract the semantics of $\mathsf{Elong}_{\mathbb{N}}$ from those of $\mathsf{Elong}_{\mathbb{N}}^{\mathbb{Z}_-}$.

Because the abstracted $w_n^{[-]}$ together with $w_n^{[-1]}$ constitute a random walk of the forward/backward rates a/b, we can define $w_n^{[-1]}$ in terms of $w_n^{[-]}$ as $w_n^{[-1]} = (a'/a)\, w_n^{[-]}$, where $a' := (a - b)/(1 - (b/a)^{n-1})$.

Semantics for Elong

In obtaining the coarsest rules of Elong, we forget the subscript n of the W_ns, so the agents can be any of W^0, W^1, $W^\bullet$ and W^-. Then, for $j \in \{0, 1, \bullet, -\}$, the probability $w^{[j]}$ of finding W^j at time t is obtained by summing $w_n^{[j]}$ over the ns; i.e., $w^{[j]} = \sum_n w_n^{[j]}$. The master equations for the probabilities $w^{[\bullet]}$, $w^{[0]}$, $w^{[1]}$ and $w^{[-]}$, under the normalization constraint $1 = w^{[\bullet]} + w^{[0]} + w^{[1]} + w^{[-]}$, are correspondingly obtained by summing equations (4), (5) and (6) over the ns, respectively.

$$\frac{d}{dt} w^{[0]} = a w^{[-1]} + \alpha_2 w^{[\bullet]} + b w^{[1]} - (\beta_2 + a + b) w^{[0]} \tag{12}$$

$$\frac{d}{dt} w^{[\bullet]} = \beta_2 w^{[0]} + \alpha_1 w^{[1]} - (\alpha_2 + \beta_1) w^{[\bullet]} \tag{13}$$

$$\frac{d}{dt} w^{[1]} = \beta_1 w^{[\bullet]} + a w^{[0]} - (b + \alpha_1) w^{[1]} \tag{14}$$

The master equations (12), (13), and (14) together with (11) can be characterised by the state-transitions shown in Fig 9.

We now examine the equilibrium convergence of these equations, as depicted in Fig 9. Biologically, the equilibrium assumption corresponds to rapid equilibrium kinetics of translocations [14] and NTP binding for polymerisation [2].

One characteristic of thermal equilibrium is the principle of *detailed balance* [8,10,16,20], which asserts that forward and reverse reactions must proceed with equal frequency. The word "detailed" conveys that balance must be satisfied for
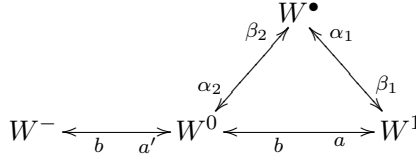
$$W^\bullet$$

Fig. 9. State transitions for Elong

each individual reaction; hence, detailed balance strengthens the globally balanced condition of the *stationary distribution* (i.e., invariance). Detailed balance characterises the time reversibility of the probability flow, and imposes the following conditions on the probability distributions $w_*^{[j]}$.

Detailed balance

$$w_*^{[0]} a = w_*^{[1]} b \qquad w_*^{[1]} \alpha_1 = w_*^{[\bullet]} \beta_1 \qquad w_*^{[\bullet]} \alpha_2 = w_*^{[0]} \beta_2$$

Moreover, in detailed balance, the reaction rates are constrained by the *Wegscheider condition* [8,16]. The Wegscheider condition, used by chemists as an algorithm to check for detailed balance, stipulates that the free energy change along the triangular path of Fig 9 is zero when the rates are thermodynamically described as in Section 4:

Wegscheider condition

$$(a/b)\,(\alpha_1/\beta_1)\,(\alpha_2/\beta_2) = 1$$

Under this condition, the master equations of Elong converge to the equilibrium probabilities $w_*^{[j]}$ with a relaxation time of τ:

Equilibrium

$$w_*^{[0]} = b\beta_1\tau \qquad w_*^{[1]} = a\beta_1\tau \qquad w_*^{[\bullet]} = a\alpha_1\tau \qquad \tau = 1/(b\beta_1 + a\beta_1 + a\alpha_1) \text{ (15)}$$

We end this subsection with a short remark on the average behaviour, similarly to [30,31].

Average chemical master equation of Elong

The mean position of the RNAP tip is given by

$$\langle x(t) \rangle = \sum_n n \left(w_n^{[0]} + w_n^{[1]} + w_n^{[\bullet]} + w_n^{[-j]} \right).$$

By summing (11), (12), (13) and (14) n-times, the time evolution of the average is simply given by

$$\frac{d}{dt}\langle x(t)\rangle = (\alpha_4 - \beta_4)w^{[0]} + (\alpha_3 - (\beta_1 + \beta_3))w^{[\bullet]} + \alpha_1 w^{[1]}$$

3.2   Quasi-Steady State Approximation: From Elong *Back to* Elong$_{\mathbb{N}}^{\mathbb{Z}}$

In this short subsection, we accommodate the quasi-steady state approximation (QSSA) into our rule-based modelling. The QSSA, firstly employed in enzymatic kinetic modelling, is widely used in biochemical network modelling and considerably simplifies the modelling [16,26]. QSSA stems from the empirical observation that the reaction intermediates are short-lived relative to the other reactants. In our QSSA, by requiring separation of time scales between the fast and slow reacting interactions, we can approximately retrieve the finest grained rules from the coarsest grained ones, conversely of the abstract rules from Elong$_{\mathbb{N}}^{\mathbb{Z}}$ to Elong. Hence the QSSA is adjoint to the rule abstraction of Section 3.1 between the coarsest and finest grained rules.

The quasi-steady state approximation relaxes the probabilistic constraint among $w_n^{[j]}$'s with respect to the time scale of elongation. Assuming that polymerisation/depolymerisation is much slower than translocation; that is, $\alpha_i, \beta_i \ll a, b$, the two constants a and b may be omitted over the time scale of our interest [2,14,30], and the observable agents can be expressed as W_n. This yields a quasi-steady state approximation of the semantics, where w_n is the probability of finding W_n at time t.

Quasi-steady state approximation

For $n \in \mathbb{N}$ and $j \in \{0, 1, \bullet, -\}$, we have

$$w_n^{[j]}(t) \approx w_*^{[j]} \times w_n(t) \qquad \text{where} \quad w_n = \sum_{j \in \{1,0,-,\bullet\}} w_n^{[j]}$$

Then, summing equations (4), (5), (6) and (11), we obtain a simple birthdeath master equation with effective polymerisation/depolymerisation rates θ_+/θ_- :

$$\frac{d}{dt} w_n(t) = \theta_+ w_{n-1} - (\theta_+ + \theta_-) w_n + \theta_- w_{n+1}$$

in which $\theta_+ = \alpha_4 b \beta_1 + \alpha_3 a \alpha_1 + \alpha_1 a \beta_1$ and $\theta_- = \beta_4 b \beta_1 + \beta_3 a \alpha_1 + \beta_1 a \alpha_1$.

Together with the equilibrium solution (15) of Elong, this simple biased random walk derived from the quasi-steady state approximation provides the semantics of the finest rules Elong$_{\mathbb{N}}^{\mathbb{Z}-}$.

4   Energy and Chemical Equilibrium

This section discusses the thermodynamical meaning of elongation. We discuss the free energy profile associated with an agent W of variant configurations binding to the active sites of TEC. At equilibrium, the probability function π from a set of configurations w_n^m's is given by an energy function $\mathcal{E}$ [8,9]:
(Boltzmann distribution)

$$w_n^{[j]} = \pi(W_n^j) = \frac{1}{Z_n} \exp\left(\frac{-\mathcal{E}(W_n^j)}{k_B T}\right) \qquad \text{with} \qquad Z_n = \sum_j \exp\left(\frac{-\mathcal{E}(W_n^j)}{k_B T}\right) \quad (16)$$

where k_B is the Boltzmann constant, T is the temperature and Z_n is the partition function of the transcription length n.

Back-track and forward-track pathways in chemical equilibrium

Let us consider the translocation reaction in $\mathsf{Elong}^{\mathbb{Z}}_{\mathbb{N}}$. The free energy $\mathcal{E}$ of a reaction is given by $\mathcal{E} = \mathcal{E}^0 + k_B T \log Q$, where $\mathcal{E}^0$ is the free energy change in the reaction and Q is the reaction quotient, derived from the law of mass action. At equilibrium, $\mathcal{E} = 0$, Q equals the equilibrium constant K_{eq}; hence we have $\mathcal{E}^0 = -k_B T \log K_{eq}$. That is,

$$K_{eq} = \frac{a}{b} = \exp\left(\frac{\mathcal{E}(W_n^0) - \mathcal{E}(W_n^1)}{k_B T}\right) = \exp\left(\frac{\mathcal{E}(W_n^{i-1}) - \mathcal{E}(W_n^i)}{k_B T}\right) \qquad (17)$$

According to this equation, the equilibrium constant K_{eq} is solely determined by the free energy difference of the reaction. Note that (17) is equivalent to the detailed balance condition between the transition rates a and b at equilibrium, but involves the Boltzmann distribution.

The main reaction pathway follows Michaelis-Menten kinetics

In the presence of a competitive inhibitor, the whole horizontal pathway of (1) follows Michaelis-Menten enzyme kinetics [2,14]. In our agent based framework, the enzyme kinetics is directly derivable only under the probabilistic assumption of the rapid convergence of the stationary distribution. We do not require the two traditional assumptions of rapid equilibrium; namely, the Henri-MechaelisMenten assumption (or its more general quasi-steady state version, the Briggs-Haldane assumption [5]) and the mass balance of the total concentration [25].

Let p_1, $p_\bullet$ and p_0 be the probabilities $w_n^{[1]}$, $w_{n+1}^{[\bullet]}$ and $w_n^{[0]}$, respectively. The transition matrix Q in the Kolmogorov backward equation $p'(t) = Q\,p(t)$ for $p = (p_i)_{i=1,\bullet,0}$ is then given by $Q = \begin{pmatrix} -(q_{\bullet 1}+q_{01}) & q_{1\bullet} & q_{10} \\ q_{\bullet 1} & -q_{1\bullet} & 0 \\ q_{01} & 0 & -q_{10} \end{pmatrix}$. Setting $Q\,p(t) = 0$, we obtain the stationary distribution, in which $p_\bullet$, p_1 and p_0 are given by

$$p_\bullet = \frac{1}{\frac{q_{1\bullet}}{q_{\bullet 1}}(1+\frac{q_{01}}{q_{10}})+1} \qquad p_1 = \frac{q_{1\bullet}}{q_{\bullet 1}}\,p_\bullet \qquad p_0 = \frac{q_{1\bullet}}{q_{\bullet 1}}\frac{q_{01}}{q_{10}}\,p_\bullet$$

The probability $q_{1\bullet}$ is associated with the unbinding rate k_{-2} of NTP from $\mathsf{TEC}_{(n,post)}$, and the reverse binding NTP probability is proportional to the substrate concentration $[\mathsf{NTP}]$. Therefore, the first equation becomes

$$p_\bullet = \frac{[\mathsf{NTP}]}{\frac{k_{-2}}{k_2}(1+\frac{q_{01}}{q_{10}}) + [\mathsf{NTP}]} \qquad \text{by } q_{1\bullet} = k_{-2} \text{ and } q_{\bullet 1} = k_2[\mathsf{NTP}]$$

Let v denote the velocity of the formation rate of $\mathsf{TEC}_{(n+1,pre)}$. v is maximized at $v_{max} = k_{3'}[\mathsf{TEC}_{(n,pre)}]$ and the ratio $v/v_{max} = p_\bullet$; thus v becomes

$$v = \frac{v_{max}[\mathsf{NTP}]}{\frac{k_{-2}}{k_2}(1+\frac{q_{01}}{q_{10}}) + [\mathsf{NTP}]}$$

This equation describes Michaelis-Menten kinetics. For the simple translocation reaction given by the ratio a/b in (17), $\frac{q_{01}}{q_{10}}$ becomes the equilibrium constant K_{eq}.

Similarly, the maximum reaction velocity of the complex pathway (3) becomes

$$v = \frac{v_{max}[\mathsf{NTP}]}{\frac{k_{-2}}{k_2}\left(\frac{1+(k_{-i_1}/k_{i_1})}{1+(k_{-i_2}/k_{i_2})}\right) + [\mathsf{NTP}]},$$

which describes the Michaelis-Menten kinetics of the complex pathway.

On the other hand, when the translocations in the back/forward track pathways are in equilibrium, the steady state elongation rate follows global Michaelis-Menten kinetics, given by:

$$v = \frac{v_{max}[\mathsf{NTP}]}{K_n + [\mathsf{NTP}]} \quad \text{with} \quad K_n = \frac{k_{-2}}{k_2}\log\prod_j \frac{w_n^{[j]}}{w_n^{[0]}}$$

where the range of j ensures that state W_n^j is accessible from state W_n^0. Note that by the energy/probability correspondence of (16), the log term equals the sum of the exponential energy differences $\sum_j \exp\left((\mathcal{E}(W_n^0) - \mathcal{E}(W_n^j))/k_B T\right)$.

5   Conclusion and Future Work

This paper presents rule-based modelling of transcription elongation, considering the stochastic interactions among the active sites of TEC. Our modelling provides a uniform insight into the syntactical (i) and semantic (ii) nature of elongation:
(i) Variations in chemical pathways arise by combinations of interacting sites on the agents. The mechano-chemical pathways are described by rules abstracted by ignoring certain sites on the agents.
(ii) Our descriptive rules yield a stochastic model of the chemical master equations, so abstractions are naturally derivable from the model. The quasi-steady state approximation is understood as the retrieval of the finest grained rules from the coarser grained ones, which supplements the model abstraction.
Finally, our rule-based modelling provides the thermodynamics of the elongation via the Boltzmann distribution, in which Michaelis-Menten kinetics is well incorporated.

In future work, we will focus on two themes:

Elaborated details of the Brownian ratchet mechanism
The rule-based modelling may be applicable to Brownian movement; in particular, it may describe the conformational details of various pawls in TEC, such as the stationary and reciprocating pawls explained in [4]. These conformations include the G-loop configuration that controls bending/straightening of the F-helix. As the thermodynamical counterpart, the generalized Michaelis-Menten kinetics [13] needs to be examined in terms of our rule-based modelling.

Generalized diffusion with drift
Given that RNAP is a molecular motor, we will augment our rule-based description of Brownian motion with thermal drift (thermal force). In this case, the probability flux is determined by the drift plus Fickian diffusion. Therefore, in

terms of κ, we need to express the probability flux as a Fokker-Planck equation. This theme is important because the Boltzmann distribution is an equilibrium solution of the FP equation. Our future work will provide a more energy-efficient description of elongation (accommodating the energy landscape and other terms used by chemists).

Acknowledgement

References

[1] I.Artsimovitch, and R. Landick, *Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals*, Proc Natl Acad Sci USA, 20, 97(13), 7090-5 (2000)

[2] L. Bai, A. Shundrovsky, MD. Wang, *Sequence-dependent kinetic model for transcription elongation by RNA polymerase*, J Mol Biol. **19;344(2)**, 335-49 (2004)

[3] L. Bai, A. Shundrovsky, and MD. Wang, *Kinetic modeling of transcription elongation*, In RNA polymerases as molecular motors by the Royal Society of Chemistry (2009)

[4] G.Bar-Nahum, V.Epshtein, AE.Ruckenstein, R.Rafikov, A.Mustaev and E.Nudler, *A Ratchet Mechanism of Transcription Elongation and its Control*, Cell **120(28)** 183-93 (2005)

[5] G.E. Briggs and J.B.S. Haldane, *A note on the kinematics of enzyme action*, Biochem J **19 (2)**: 338339. (1925)

[6] P. R. Costa, M. L. Acencio, and N. Lemke, *Cooperative RNA Polymerase Molecules Behavior on a Stochastic Sequence-Dependent Model for Transcription Elongation*, PLoS One. **8(2)**, e57328. (2013)

[7] Vincent Danos, Jérôme Feret, Walter Fontana, Russell Harmer and Jean Krivine, *Rule-based modelling of cellular signalling*, (CONCUR 2007), LNCS **4730**, Springer (2007)

[8] Vincent Danos, Nicolas Oury, *Equilibrium and Termination*, DCM'10, EPTCS **26**, 75-84 (2010)

[9] Vincent Danos, Nicolas Oury, *Energy and Termination II: the case of Petri nets*, Mathematical Structures in Computer Science **23(2)**: 290-307 (2013)

[10] Rick Durrett, *Essentials of Stochastic Processes*, Springer Texts in Statistics (1999)

[11] Jérôme Feret, Vincent Danos, Jean Krivine, Russ Harmer and Walter Fontana, *Internal coarse-graining of molecular systems*, Proc. Natl. Acad. Sci. USA, **106**(16) 6453-6458 (2009)

[12] J. Gelles, R. Landick, *RNA polymerase as a molecular motor*, Cell **93.1** 13-6 (1998)

[13] A.N. Gorban and M. Shahzad, *The Michaelis-Menten-Stueckelberg Theorem*, Entropy, **13** 966-1019 (2011)

[14] R. Guajardo, R. Sousa, *A model for the mechanism of polymerase translocation*, J Mol Biol. **10;265(1)**, 8-19 (1997)

[15] Masahiro Hamano, *Sustainability of RNA-interference in Rule-Based Modelling*, In Proc. of the Third International Workshop on Static Analysis and Systems Biology (SASB 2012) Electronic Notes in Theoretical Computer Science **313**, 6577 (2015)

[16] R. Heinrich, S. Schuster, *The Regulation of Cellular Systems*, Chapman Hall, (1996)

[17] Frank Julicher and Robijn Bruinsma, *Motion of RNA Polymerase along DNA: A Stochastic Model*, Biophysical Journal **74** 11691185 (1998)

[18] R. Landick, *The regulatory roles and mechanism of transcriptional pausing*, Biochemical Society Transactions **34**, part 6, (2006)

[19] HH. McAdams, A. Arkin, *Stochastic mechanisms in gene expression*, Proc Natl Acad Sci USA. **4;94(3)** 814-9 (1997)

[20] J. R. Norris, *Markov Chains*, Cambridge University Press, Cambridge Series in Statistical and Probabilistic Mathematics (1998)

[21] E. Nudler, *Transcription elongation: structural basis and mechanisms*, Journal of Molecular Biology, **288(1)** 1-12. (1999)

[22] C. Priami, A. Regev, E. Shapiro, W. Silverman, *Application of a stochastic name-passing calculus to representation and simulation of molecular processes*, Information Processing Letters **80** 25-31 (2001)

[23] AS. Ribeiro, OP. Smolander, T. Rajala, A. Hkkinen, O. Yli-Harja, *Delayed stochastic model of transcription at the single nucleotide level*, J Comput Biol. **16(4)**, 539-53 (2009)

[24] Marc R. Roussel and Rui Zhu, *Stochastic kinetics description of a simple transcription model*, Bull. Math. Biol. **68**, 1681-1713 (2006)

[25] IH. Segel, *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*, Wiley, (1993)

[26] L. A. Segel, *Modelling dynamic phenomena in molecular and cellular biology*, Cambridge Univ. Press, Cambridge (1984)

[27] V. Tadigotla, D. Maoileigidh, A. Sengupta, V. Epshtein, R. Ebright, et al, *Thermodynamic and kinetic modeling of transcriptional pausing*, Proc. Natl. Acad. Sci. USA. **103**, 819 (2006)

[28] T. Tripathi, D. Chowdhury, *Interacting RNA polymerase motors on DNA track: effects of traffic congestion and intrinsic noise on RNA synthesis*, Physical Review E**77 (1)**, 011921 (2008)

[29] N.G. Van Kampen, *Stochastic Processes in Physics and Chemistry (3rd edn)* North Holland (2007)

[30] M. Voliotis, N. Cohen, C. Molina-París, and T. B. Liverpool, *Fluctuations, pauses, and backtracking in DNA transcription* Biophys. J. **94**, 334348 (2008)

[31] Hong-Yun Wang, Tim Elston, Alexander Mogilner, George Oster, *Force Generation in RNA Polymerase*, Biophysical Journal **74**, 11861202 (1998)

[32] K. Wilson, C. Conant and P. von Hippel, *Determinants of the Stability of Transcription Elongation Complexes: Interactions of the Nascent RNA with the DNA Template and the RNA Polymerase*, J. Mol. Biol. **289**, 1179-1194 (1999)

[33] Thomas D. Yager , Peter H. Von Hippel, *A thermodynamic analysis of RNA transcript elongation and termination in Escherichia coli*, Biochemistry **30 (4)**, 10971118 (1991)