RIMS Kôkyûroku Bessatsu
B41 (2013), 101–123

Ultradiscrete Modeling and Simulation for Gene Transcription

By

Yoshihiro Ohta * and Sigeo Ihara **

Abstract

RNA polymerase II (RNAPII) is responsible for transcription, that is a central cellular process. Genome-wide studies show that transcription by RNAPII is dynamically regulated. Due to the experimental difficulty in molecular biological approach, the picture of the gene transcription remains snapshot rather than dynamical views. Therefore, to reveal the principles of transcription, the mathematical modeling and simulation, by fusing spatial-temporal deep analysis of real data, are crucial. From the simulations of low density RNAPII in the *SAMD4A* gene, we found that the RNAPII molecules move as a free flow state, though there exist regions of reduced velocity, as far as the time interval between nearest RNAPII molecules is larger than the time required for an RNAPII passing the exclusion length in the reduction region. On the other hand, if the reduction is strong enough to reach a certain threshold, at the maximally reductive velocity region, a transition occurs from the free flow state to the states with congested and repetitive flows.

§ 1. Introduction

§ 1.1. Transcription dynamics

Transcription by RNA polymerase II (RNAPII) [1] is at the core of gene expression and hence is the basis of all cellular activities. To generate a mature messenger RNA (mRNA) [2] RNAPII traverses a transcription cycle; this involves recruitment to an activated promoter, initiation, escape into the gene, elongation, and termination [1].

Received October 30, 2012. Revised January 28, 2013.
2000 Mathematics Subject Classification(s): 2000 Mathematics Subject Classification(s):
*RCAST, The University of Tokyo, Tokyo 153-8904, Japan
e-mail: ohta@genome.rcast.u-tokyo.ac.jp
**RCAST, The University of Tokyo, Tokyo 153-8904, Japan
e-mail: ihara@genome.rcast.u-tokyo.ac.jp

[1] Enzyme found in eukaryotic cells that catalyzes the transcription of DNA to synthesize precursors of mRNA.

[2] Molecule of RNA encoding a chemical blueprint for a protein product.

Processing of the nascent transcript - that can include capping, splicing, and poly (A) addition - is coupled to polymerization, and the C-terminal domain (CTD) of the polymerase acts as a scaffold for the binding of many of the factors involved [2, 3, 4].

Based on the most recent results on the transcription in the literature and the deep analysis of real data of our own, we will construct a mathematical modeling and perform simulations to reveal the principles of gene transcription.

§ 1.2. Related Research

Since the transcription process is comprised of an ingenious form of cooperativity of proteins, the dynamical transport of RNAPII is still controversial [5, 6, 7, 8, 9]. Our previous experiments [5] indicated a fine spatiotemporal structure (10 bp and 7.5 min) of gene transcription but finer time resolution experiments are quite difficult to perform from the point of view of molecular biological methodology. Thus, this paper reports an application of a cellular automaton model [10] of transcription, in which finer time resolution is performed so as to take correlation between RNAPIIs into account.

In the previous report, the simulation of transcription dynamics has been done on bacterial RNAPs [11]. By applying a mathematical method of traffic analysis to high density RNAPII motion, Tripathi and Chowdhury [12, 13] showed that the closely associated RNAPII molecules exhibit a local correlation [7, 14], and they studied the effects of traffic congestion on mRNA synthesis. Their investigation mainly concerns the congested sites in prokaryotes where RNAPII direct interaction controls the RNAPII flow without a blockade.

Also, the nucleosome modifications, insulator proteins, and blockade sites, as well as the clearly distinct genomic regions in eukaryotic cells, pose issues that were not addressed by earlier models of cooperative transcription [12, 13]. For example, what is the effect of there being two distinct regions in which RNAPII moves at two different average velocities on the spatial distribution of the RNAPIIs? This inhomogeneity, that is distinct from the sequence inhomogeneity of the template DNA, occurs on much longer length scales [5, 15, 16].

In understanding the organization of transcription in higher eukaryotic cell and in estimating the still unknown interaction between RNAPIIs in the transcription factories, knowing how often RNAPIIs are injected into the gene in a cell and how the dynamical character changes are quite important. By applying a cellular-automaton model [10, 17, 18, 19, 20, 21, 22, 23], we handle this problem.

§ 1.3. Modeling and Simulation of Transcription

In our previous time-course experimental studies of RNAPII protein binding to DNA by ChIP-sequencing in cultured human umbilical vein cells [5], we collected the

gene response data of more than 100 kpb-long gene after drug *stimulation* (hereafter, we use italic form for this word so as not to confuse 'simulation'), tumor necrosis factor α (TNFα), a potent cytokine that causes the inflammatory response by sequentially activating the expression of more than 6,000 genes. At different times after *stimulation* with TNFα, total nuclear RNA was purified and hybridized to a tiling microarray bearing oligonucleotides complementary to *SAMD4A*, a long gene of 221 kbp; signals were normalized using an algorithm.

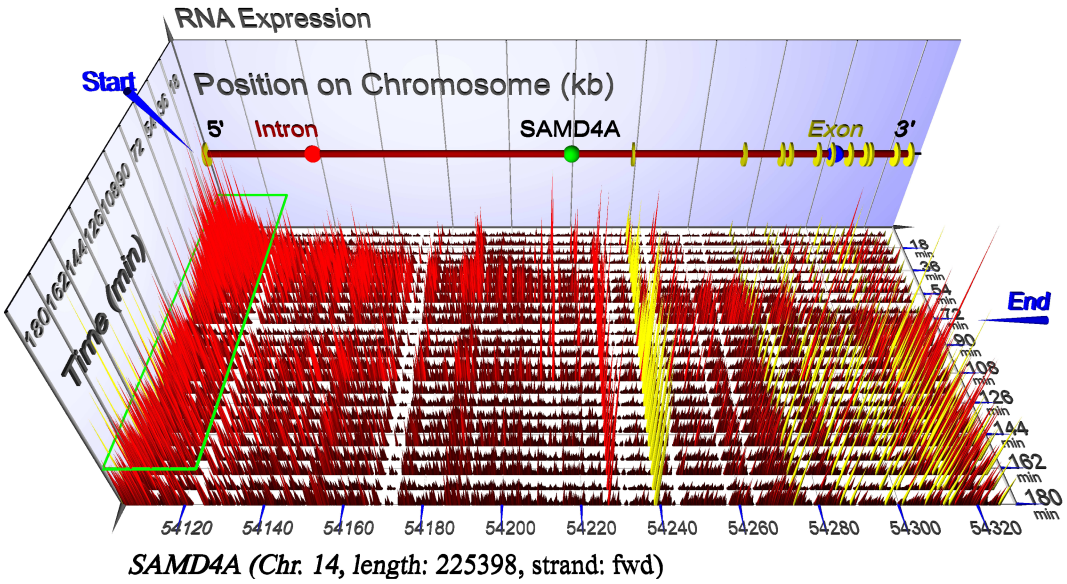

Figure 1. 3D Co-transcriptional Splicing of *SAMD4A*

Transcription waves visualized using microarrays. HUVECs were stimulated with TNFα, samples collected every 7.5 min for 3 h, and total nuclear RNA purified and hybridized to a tiling microarray bearing 25-mers complementary to *SAMD4A*. The vertical axis gives intensity of signal detected by intronic and exonic probes. Gene length and genomic location are shown at the front, probe positions within the gene from left to right; and time after *stimulation* from top to bottom. Arrowheads indicate the "Start" and "End" of the first wave of transcription that sweeps down the gene.

We also showed that in addition to the histone lysine methylation-enriched sites, such as the H3K36me3 sites, RNAPII molecules stall near the CCCTC-binding factor (CTCF) and cohesin [3] binding sites [5]. Since CTCF and cohesin proteins act as nucleosome positioning anchors for DNA, they are closely related with DNA loop formation and are also known as insulators of RNAPII motion [24, 25]. The analysis of

[3]Transcriptional repressor also known as 11-zinc finger protein or CCCTC-binding factor that in humans is encoded by the CTCF gene.

pre-mRNA synthesis in long genes indicates that the velocity of RNAPII in exons is quite a bit slower than in introns [5]; this conjecture is also supported by other groups [9, 16].

By using the features obtained above, we propose a model for describing the transcription dynamics. This model is based on the cellular automaton (CA), in which the boxes are aligned in one-dimension spanning the entire gene from the $5'$ end (left) to the $3'$ end (right), and the balls (i.e., the RNAPII molecules) move according to the rule 184 as proposed by Wolfram [9]. The rule 184 is generally used as a simple model of traffic flow in a single lane of motion: RNAPII molecules move in a single direction. An RNAPII molecule moves at a distance of one box when there is no RNAPII in front of the (right) next box.

§ 2. Formulation

§ 2.1. CA Modeling of RNAPII Elongation

We provide a detailed description of the mathematical formulation by using the features obtained from our experimental analysis [5]. Suppose the system has millions of cells, and each cell has more than twenty thousands of genes. There are M RNAPII molecules which are engaged in transcription of a gene. Hereafter, we focus on the motion of the kth RNAPII ($k \in \{1, 2, \cdots, M\}$).

We simulate the RNAPII dynamics using a CA model. First, we describe a simple cellular automaton model called a box-ball model, where boxes are aligned in one dimension spanning over whole gene from $5'$ end (left) to $3'$ end (right), and balls (= RNAPII molecules) move according to the rule 184 proposed by Wolfram. In addition, we propose a possible mechanism of RNAPII motion by extending the totally asymmetric simple exclusion process (TASEP). First, we define some mathematical symbols, initial conditions, and boundary conditions etc.

A gene under consideration of length L is split into N discretized boxes, called cell units, based on a suitable unit of length δL. For example, the size of one RNAPII molecule η or one nucleosome length is taken as one unit. Then these gene units are allocated a number such as $\{1, 2, \cdots, N\}$ from the edge of the gene called the TSS as follows:

$$\{G_1, G_2, \cdots, G_N\}. \tag{2.1}$$

The initial state is defined at $t = 0$, and the motion of the RNAPII molecules is determined at $t = i\delta t$ ($i = 0, 1, 2, \cdots$) according to a suitable time evolution unit δt. Here we take into account the RNAPII effective velocity difference γ as

$$\frac{\delta L}{\delta t} = v_i = \gamma v_e, \tag{2.2}$$

where v_i is the velocity in the introns, v_e is that in the exons, and we set the base velocity as v_i. As explained above, $\gamma \geq 1$. If a cell unit in front is vacant, the above described properties are expressed on the position x_j^t of G_j as

$$x_j^{t+(i\delta t)} = \begin{cases} x_j^t & (0 \leq i < \gamma_j), \\ x_j^t + \delta L & (i = \gamma_j), \end{cases} \tag{2.3}$$

which i is a nonnegative integer ($0 \leq i \leq \gamma_j$) and t is an arrival time at the position x_j^t. Specifically, we denote by γ_j the γ value of G_j.

The gene is assumed to be constituted by an alternating 1-dimensional alignment of $s+1$ exons and s introns. For cyclic case, $s+1$ exons should read as s exons. RNAPIIs are assumed to move in a single direction in the line from the $5'$ end to the $3'$ end with effective forward velocity. The first exon (1st exon) consists of cell units with numbers $\{1, 2, \cdots, n_1\}$, the first intron (the 1st intron) consists of cell units with numbers

$$\{n_1 + 1, n_1 + 2, \cdots, m_2\}, \tag{2.4}$$

The second exon (2nd exon) consists of cell units with numbers,

$$\{m_2 + 1, m_2 + 2, \cdots, n_2\}, \tag{2.5}$$

and so on. The remaining are similarly defined; the kth intron ($k = 1, 2, \cdots, s$) consists of cell units with numbers

$$\{n_k + 1, n_k + 2, \cdots, m_{k+1}\}, \tag{2.6}$$

and kth exon ($k = 1, 2, \cdots, s+1$) consists of cell units with numbers

$$\{m_k + 1, m_k + 2, \cdots, n_k\}, \tag{2.7}$$

where $m_1 = 0$ and $n_{s+1} = N$.

For example, the total length of the *SAMD4A* gene is 225398 bp, consisting of 12 exons and 11 introns. The average length of each exon is 183 bp with small deviations, but the intron length varies substantially.

§ 2.2. Global Intensity Profile of RNAPII

Previous reports adopted a static picture and attributed these fine structures of the RNAPII density profile to the stalling of RNAPII [8, 9, 15, 16]. To understand the physical process underlying the observed RNAPII density profile, we performed a numerical simulation of RNAPII motion with the additional assumptions that the number of cells in which RNAPII moving molecules has a distribution f_{cell} over time t from *stimulation*.

To define the intensity profile of RNAPII density at G_j, we introduce the following time domain variables instead of using the parameters of displacement. $T_j(k)$ represents the elapsed time of the kth RNAPII to arrive at the jth cell unit G_j from the start cell unit G_1, and this can be obtained by CA. By using this parameter, the residence time of kth RNAPII at G_j, denoted as $\delta T_j(k)$, is given by

$$\delta T_j(k) = T_{j+1}(k) - T_j(k). \tag{2.8}$$

Thus, the number of the kth RNAPII molecules, that reside in the jth cell unit G_j, coincides with the ones that start at time $T_{(s)}$ in the following time interval:

$$(t - T_j(k) - \delta T_j(k) + \epsilon) \leq T_{(s)} \leq (t - T_j(k) + \epsilon)\,, \tag{2.9}$$

where ϵ is denoted as the sum of the time range that results in the stochastic actions of RNAPII including occasional stops and backtracks. Here we introduce $f^{(k)}_{cell}(T_{(s)})$ as the number of activated cells in which kth RNAPII molecule started at time $T_{(s)}$. Consequently, the number of kth RNAPII molecules $I^t_j(k)$, that reside in the jth cell unit G_j at time t, is explicitly given as follows.

$$I^t_j(k) = \sum_{i_t=-\delta T_j(k)'+\epsilon'}^{\epsilon'} f^{(k)}_{cell}(t - T_j(k) + i_t\delta t), \tag{2.10}$$

where i_t takes integers for integrating residence times, and $\delta T_j(k)'$ and ϵ' are approximated integer values that take $\delta T_j(k)/\delta t$ and $\epsilon/\delta t$, respectively.

To evaluate Eq.(2.10) we use time delays of RNAPII molecules. Here we introduce $\Delta T_j(k)$ as a time difference of the elapsed time of kth RNAPII from that of the preceding $(k-1)$th RNAPII to reach G_j, which yields

$$\Delta T_j(k) = T_j(k) - T_j(k-1), \tag{2.11}$$

where $\Delta T_j(1) \equiv 0$. Thus, $T_j(k)$ can take the form by using the integer index l as

$$T_j(k) = T_j(1) + \sum_{l=2}^{k} \Delta T_j(l). \tag{2.12}$$

Specifically, we denote by $\delta t_{in}(k)$ the time difference $\Delta T_1(k)$ which indicates the injection time difference for the start cell unit G_1, since this parameter can be given as initial conditions. By using this parameter, $f_{cell}^{(k)}(T_{(s)})$ is written recursively as

$$\begin{cases} f_{cell}^{(2)}(T_{(s)}) = f_{cell}^{(1)}(T_{(s)} - \delta t_{in}(2)), \\ f_{cell}^{(k)}(T_{(s)}) = f_{cell}^{(1)}(T_{(s)} - \sum_{l=2}^{k} \delta t_{in}(l)). \end{cases} \tag{2.13}$$

From Eqs.(2.10) and (2.13), one can obtain the number of kth RNAPII molecules $I_j^t(k)$ as;

$$\begin{aligned} I_j^t(k) = & \sum_{i_t=-\delta T_j(k)'+\epsilon'}^{\epsilon'} \\ & f_{cell}^{(k)}\left(t - \left(T_j(1) + \sum_{l=2}^{k} \Delta T_j(l)\right) + i_t \delta t\right) \\ = & \sum_{i_t=-\delta T_j(k)'+\epsilon'}^{\epsilon'} \\ & f_{cell}^{(1)}\left(t - T_j(1) - \sum_{l=2}^{k} (\delta t_{in}(l) + \Delta T_j(l)) + i_t \delta t\right), \end{aligned} \tag{2.14}$$

where $f_{cell}^{(1)}(T_{(s)}) \equiv f_{cell}(T_{(s)})$.

In the gene, the time interval of the kth and $(k-1)$th RNAPIIs injected by the time interval of $\delta t_{in}(k)$ will be modulated by CA. Thus, we introduce the arrival time delay from the preceding $(k-1)$th RNAPII to the kth RNAPII at G_j and denote as Δ_j^k;

$$\Delta_j^k = \delta t_{in}(k) + \Delta T_j(k). \tag{2.15}$$

Consequently, the condition below is always satisfied for any j and k in the asymmetric simple exclusion process (ASEP).

$$\Delta_j^k > 0. \tag{2.16}$$

By taking the limit of time unit $\delta t \to +0$, $I_j^t(k)$ using Eq.(2.15) can be given as follows

$$I_j^t(k) = \int_{-\delta T_j(k)+\epsilon}^{\epsilon} f_{cell}\left(u + t - T_j(1) - \sum_{l=2}^{k} \Delta_j^l\right) du, \tag{2.17}$$

where u is a parameter of time on the function f_{cell} as

$$f_{cell}\,(u) := \lim_{\substack{\delta t \to +0 \\ u = i\delta t}} \frac{f_{cell}^{(k)}(i\delta t)}{\delta t}. \tag{2.18}$$

Therefore, it should be noted that I_j^t, the total number of all RNAPII molecules that reside in the jth cell unit G_j at time t, can be given as follows,

$$I_j^t = \sum_{k=1}^{M} I_j^t(k). \tag{2.19}$$

Here one can introduce the RNAPII stop time θ min and the backtrack length $\kappa \simeq 10$ bp [26, 27]. By using these parameters, the time range ϵ of this stochastic action is expressed as

$$\epsilon = \alpha\theta + \beta\frac{\kappa}{v_i}, \tag{2.20}$$

where the constant values α and β are small integer values to represent the frequencies of stops and backtracks in the one cell unit, respectively. Therefore, the left-hand side can be eliminated by applying the phenomenon to Eq.(2.3). Thus, Eq.(2.20) is reduced to

$$\epsilon = \frac{\beta\kappa}{\delta L}\delta t. \tag{2.21}$$

From the molecular biology experiments, a relation $\epsilon \simeq 10^{-2} \cdot \delta t \ll \delta T_j(k)$ should hold. This condition helps us to evaluate the time range of Eq.(2.17) that is represented for $u \in [-\delta T_j(k) + \epsilon, \epsilon]$. Namely, in this envelope analysis, we can ignore the time range ϵ for the reduction to $u \in [-\delta T_j(k), 0]$, since ϵ is sufficiently small in this approximation. Therefore, such stochastic movements of the RNAPII molecule can be ignored in describing the collective motion of the RNAPIIs in a large number of cells.

§ 2.3. Formulation of Flow in the Different γ Units

In order to study the stability of RNAPII flow, we will follow the kth RNAPII motion in a gene. We assume that the kth RNAPII and preceding $(k-1)$th RNAPII molecules reside at G_j and G_{j+1}, respectively. Now we inductively define the subsequent time delay Δ_{j+1}^k by using the arrival time delay Δ_j^k from the preceding $(k-1)$th RNAPII to kth RNAPII for reaching G_j. First, $\emptyset_j^k$ is introduced as the time interval that G_j is vacant before the kth RNAPII arrives, that is given by

$$\emptyset_j^k = \Delta_j^k - \delta T_j(k-1). \tag{2.22}$$

Using this definition, the subsequent time delay Δ_{j+1}^k can be defined as

$$\Delta_{j+1}^k = \max\left(\emptyset_j^k + \gamma_j \delta t, \delta T_{j+1}(k-1)\right). \tag{2.23}$$

In this Eq.(2.23), one can see that the arrival time delay Δ_{j+1}^k is obtained by the two conditions whether the kth RNAPII will proceed smoothly without collision (left hand side), or will be in the congestion that occurs after collision with the preceding $(k-1)$th RNAPII (right hand side) at G_j; in other words, the kth RNAPII at G_j must wait for the $(k-1)$th RNAPII in the cell unit G_{j+1} in front. From Eq.(2.22), we may rewrite the above equation as

$$\begin{aligned}
\Delta_{j+1}^k &= \max\left(\Delta_j^k - \delta T_j(k-1) + \gamma_j \delta t, \delta T_{j+1}(k-1)\right) \\
&= \max\left(\Delta_j^k - \delta T_j(k-1) + \gamma_j \delta t, \delta T_{j+1}(k-1)\right) \\
&\quad + (\delta T_j(k-1) - \gamma_j \delta t) - (\delta T_j(k-1) - \gamma_j \delta t) \\
&= \max\left(\Delta_j^k, \delta T_{j+1}(k-1) + \delta T_j(k-1) - \gamma_j \delta t\right) \\
&\quad - \delta T_j(k-1) + \gamma_j \delta t,
\end{aligned} \tag{2.24}$$

where we used the identity for arbitrary A, B, and X as

$$\max(A, B, \cdots) + X = \max(A + X, B + X, \cdots). \tag{2.25}$$

Eq.(2.24) indicates that the subsequent time delay Δ_{j+1}^k is recursively dependent on the preceding time delay Δ_j^k. For example, let us consider four combinations of two types of successive cell units $\{G_{j-n_1+1}, \cdots, G_j\}$ and $\{G_{j+1}, \cdots, G_{j+n_2}\}$ with length of n_1 and n_2, by dividing into four cases as follows.

1) $\{G_{j-n_1+1}, \cdots, G_j\}$ are introns and $\{G_{j+1}, \cdots, G_{j+n_2}\}$ are exons

2) $\{G_{j-n_1+1}, \cdots, G_j\}$ are exons and $\{G_{j+1}, \cdots, G_{j+n_2}\}$ are introns
3) $\{G_{j-n_1+1}, \cdots, G_j\}$ and $\{G_{j+1}, \cdots, G_{j+n_2}\}$ are both introns
4) $\{G_{j-n_1+1}, \cdots, G_j\}$ and $\{G_{j+1}, \cdots, G_{j+n_2}\}$ are both exons

Here we concentrate on the diversity analysis of the arrival time delay Δ_j^k at G_j and the subsequent time delay Δ_{j+1}^k at G_{j+1} in these four cases.

2.3.1. Fast (*Intron*) to Slow Velocity Region (*Exon*)

At the intron-exon boundary, $\gamma_j = 1$ and $\gamma_{j+1} = \gamma_e$. Here γ_e especially indicates the γ value in the exon region. From the definition (2.3), the residence times of the preceding $(k-1)$th RNAPII at G_j and G_{j+1} become

$$\begin{cases} \delta T_j(k-1) = \max\left(\gamma_j \delta t, \gamma_{j+1}\delta t - \emptyset_j^{k-1}\right) \\ \qquad\qquad = \max\left(\delta t, \gamma_e \delta t - \emptyset_j^{k-1}\right), \\ \delta T_{j+1}(k-1) = \gamma_{j+1}\delta t = \gamma_e \delta t, \end{cases} \tag{2.26}$$

where $\delta T_j(k-1)$, which shows the residence time of $(k-1)$th RNAPII at G_j, can be obtained by the two conditions whether the $(k-1)$th RNAPII proceeds smoothly without collision (left hand side), or in congestion after collision with the preceding $(k-2)$th RNAPII (right hand side) at G_j. As for the conditions of Eq.(2.26), the intron to exon flow will be divided into two more specific cases. In the case of

$$\delta t \geq \gamma_e \delta t - \emptyset_j^{k-1}, \tag{2.27}$$

the $(k-1)$th RNAPII is in the free flow state. On the other hand, if

$$\delta t < \gamma_e \delta t - \emptyset_j^{k-1}, \tag{2.28}$$

the $(k-1)$th RNAPII is in the congested flow.

When the preceding $(k-1)$th RNAPII is in free flow state as in Eq.(2.27), its residence time at G_j becomes $\delta T_j(k-1) = \delta t$ from the Eq.(2.26). Thus, Eq.(2.24) can be reduced to

$$\begin{aligned} \Delta_{j+1}^k &= \max\left(\Delta_j^k, (\gamma_e + 1 - 1)\,\delta t\right) \\ &\quad -1 \cdot \delta t + 1 \cdot \delta t \\ &= \max\left(\Delta_j^k, \gamma_e \delta t\right). \end{aligned} \tag{2.29}$$

In analyzing the stability of the flow states, Eq.(2.29) shows that if the arrival time delay Δ_j^k is in the condition of

$$\Delta_j^k > \gamma_e \delta t, \tag{2.30}$$

the subsequent time delay remains stable as

$$\Delta_{j+1}^k = \Delta_j^k, \tag{2.31}$$

which generates the same homogeneous flow as before, even after the kth RNAPII arriving at the exon region. Otherwise, the arrival time delay satisfies the condition

$$\Delta_j^k \leq \gamma_e \delta t, \tag{2.32}$$

and the subsequent time delay Δ_{j+1}^k will always be set at the constant value as

$$\Delta_{j+1}^k = \gamma_e \delta t, \tag{2.33}$$

which generates a new repetitive pattern for a certain range after RNAPII reaching the exon region under the condition that $(k-1)$th RNAPII is in the free flow state. The above expressions in the time domain indicate intuitively that the RNAPII flow state depends on whether the time interval between nearest RNAPII molecules is larger than the time required for an RNAPII passing the exclusion length between RNAPIIs in the reduction region, $\gamma_e \delta t$, or not.

If the preceding $(k-1)$th RNAPII is in congested flow state, its residence time at G_j becomes $\delta T_j(k-1) = \gamma_e \delta t - \emptyset_j^{k-1}$ from Eq.(2.26). In this case, it should be noted that this $\delta T_j(k-1)$ is recursively applied by Eq.(2.22) as long as the several preceding RNAPII molecules have become congested successively. Consequently, this can be reduced as follows.

$$\begin{aligned}
\delta T_j(k-1) &= \gamma_e \delta t - \left(\Delta_j^{k-1} - \delta T_j(k-2)\right) \\
&= \gamma_e \delta t - \Big(\Delta_j^{k-1} \\
&\quad - \left(\gamma_e \delta t - \left(\Delta_j^{k-2} - \delta T_j(k-3)\right)\right)\Big) \\
&= \cdots \\
&= (k-1-m)\,\gamma_e \delta t - \left(\sum_{i=m+1}^{k-1} \Delta_j^i - \delta T_j(m)\right) \\
&= (k-m-1)\,\gamma_e \delta t - \Delta_j^{(m+1,k-1)} + \delta t,
\end{aligned} \tag{2.34}$$

where m $(1 \leq m \leq k-2)$ is introduced as an integer which takes the condition that the mth RNAPII will proceed smoothly at G_j and all of the preceding RNAPII molecules from the $(m+1)$th to the $(k-2)$th RNAPII will become congested after collisions occur. Namely, the following condition holds for any integer i $(m < i \leq k-1)$ except m as

$$\delta t < \gamma_e \delta t - \varnothing_j^i. \tag{2.35}$$

In Eq.(2.34), $\sum_{i=m+1}^{k-1} \Delta_j^i$ is denoted by $\Delta_j^{(m+1,k-1)}$ and $\delta T_j(m) = \delta t$ from the definition of m. Thus, by using Eq.(2.34), Eq.(2.24) can be reduced to

$$\begin{aligned}
\Delta_{j+1}^k = &\max(\Delta_j^k, \gamma_e \delta t + ((k-1-m)\gamma_e \delta t \\
&-\Delta_j^{(m+1,k-1)} + \delta t) - 1 \cdot \delta t) \\
&-((k-1-m)\gamma_e \delta t - \Delta_j^{(m+1,k-1)} + \delta t) + 1 \cdot \delta t \\
= &\max\left(\Delta_j^k, (k-m)\,\gamma_e \delta t - \Delta_j^{(m+1,k-1)}\right) \\
&- \left((k-1-m)\,\gamma_e \delta t - \Delta_j^{(m+1,k-1)}\right).
\end{aligned} \tag{2.36}$$

Let us analyze the stability of the flow states. Eq.(2.36) indicates that if the arrival time delay Δ_j^k satisfies the condition

$$\Delta_j^k \leq (k-m)\,\gamma_e \delta t - \Delta_j^{(m+1,k-1)}, \tag{2.37}$$

the subsequent time delay Δ_{j+1}^k in Eq.(2.36) becomes

$$\begin{aligned}
\Delta_{j+1}^k &= (k-m)\,\gamma_e \delta t - \Delta_j^{(m+1,k-1)} \\
&\quad - \left((k-m-1)\,\gamma_e \delta t - \Delta_j^{(m+1,k-1)}\right) \\
&= \gamma_e \delta t,
\end{aligned} \tag{2.38}$$

which is always set to the constant value $\gamma_e \delta t$ after the RNAPII reaching the exon region, and then generates a new repetitive pattern for a certain range. Otherwise, if the arrival time delay Δ_j^k is in the condition of

$$\Delta_j^k > (k-m)\,\gamma_e \delta t - \Delta_j^{(m+1,k-1)}, \tag{2.39}$$

the subsequent time delay Δ_{j+1}^k in Eq.(2.36) becomes

$$\begin{aligned}
\Delta_{j+1}^k &= \Delta_j^k - \left((k-m-1)\,\gamma_e \delta t - \Delta_j^{(m+1,k-1)}\right) \\
&= \Delta_j^{(m+1,k)} - (k-m-1)\,\gamma_e \delta t.
\end{aligned} \tag{2.40}$$

Consequently, if the $(k-1)$th RNAPII becomes congested after collision with the preceding RNAPII and the kth RNAPII satisfies the stability condition in Eq.(2.39), the subsequent time delay Δ_{j+1}^{k} becomes a constant value defined by Eq.(2.40). This indicates that the arrival time delay Δ_{j+1}^{k} of Eq.(2.40) is a specific condition for the RNAPII being in transition from the congestion state to the free flow state. The subsequent RNAPII will go back to the condition that the preceding RNAPII is in free flow state as in Eq.(2.27). Refer to [28] for more information about the other conditions, "slow (*exon*) to fast velocity region (*intron*)", "fast (*intron*) to fast velocity region (*intron*)", and "slow (*exon*) to slow velocity region (*exon*)".

2.3.2. Critical Density Near the γ Region

In order to find a solution to density conversion near the γ region, we define the densities at the $(l-1)$th intron, lth exon and lth intron regions, as $\rho_I^{(l-1)}$, $\rho_E^{(l)}$ and $\rho_I^{(l)}$, respectively. For the sake of simplicity, we suppose a model in which the G_j, G_{j+1} and G_{j+2} cell units are the $(l-1)$th intron, lth exon and lth intron, respectively. It should be noted that this simplified model can easily be extended to more common forms of exons or introns of long length that consist of multiple cell units, by using the results of the previous subsection. First, the density $\rho_I^{(l-1)}$ in the $(l-1)$th intron region can be defined as

$$\rho_I^{(l-1)} = \frac{\delta L}{\Delta_j^k v_i}, \tag{2.41}$$

where the density $\rho_I^{(l-1)}$ is defined with Δ_j^k, since the G_j is assumed to be the $(l-1)$th intron. Next, the density $\rho_E^{(l)}$ in the lth exon region similarly takes the form

$$\rho_E^{(l)} = \frac{\delta L}{\Delta_{j+1}^k v_e}, \tag{2.42}$$

which is subject to the conditions in the case of the intron to exon flow as noted previously.

Keeping in mind that the value of density, $\rho_I^{(l-1)}$, $\rho_E^{(l)}$ and $\rho_I^{(l)}$ changes according to the arrival time delay Δ_j^k and the state of the $(k-1)$th RNAPII, we evaluate them as follows. If the $(k-1)$th RNAPII is in a free flow state as in the following upper inequation, while an arrival time delay Δ_j^k is rather short (bounded by the parameters) and satisfies the following lower inequation,

$$\begin{cases} \delta t \geq \gamma_e \delta t - \emptyset_j^{k-1}, \\ \Delta_j^k \leq \gamma_e \delta t, \end{cases} \tag{2.43}$$

or, the $(k-1)$th RNAPII is in a congested flow state as in the following upper inequation, while an arrival time delay Δ_j^k is still rather short (but bounded by the different parameters from inequation Eq.(2.43)) and satisfies the following lower inequation,

$$\begin{cases} \delta t < \gamma_e \delta t - \emptyset_j^{k-1}, \\ \Delta_j^k \leq (k-m)\,\gamma_e \delta t - \Delta_j^{(m+1,k-1)}, \end{cases} \tag{2.44}$$

then, the arrival time delay holds $\Delta_{j+1}^k = \gamma_e \delta t$ from Eqs.(2.33) and (2.38), since G_{j+1} is exon. Therefore, the above density $\rho_E^{(l)}$ is reduced to a constant value as

$$\begin{aligned} \rho_E^{(l)} &= \frac{\delta L}{\gamma_e \delta t} \cdot \frac{\gamma_e}{v_i} \\ &= \frac{\delta L}{\delta t} \cdot \frac{1}{v_i} \\ &= \frac{\delta L}{\delta t} \cdot \frac{\delta t}{\delta L} \\ &= 1. \end{aligned} \tag{2.45}$$

Thus, the highly congested flow without collisions is generated in the lth exon region. Refer to [28] for more information about the other cases of critical density.

2.3.3. List of Definition of Symbols

We show a list of definition of symbols, $T_j(k)$, $\delta T_j(k)$, $\Delta T_j(k)$, Δ_j^k, $\emptyset_j^k$, and $\Delta_j^{(m+1,k-1)}$, using a symbol $t_j(k)$ which denotes the arrival time of the kth RNAPII at the jth cell. Then we have a relation

$$t_1(k) = t_1(k-1) + \delta t_{in}(k), \tag{2.46}$$

and all the other symbols are expressed as follows. The elapsed time of the kth RNAPII, $T_j(k)$, to arrive at the jth cell unit G_j from the start cell unit G_1 is

$$\begin{aligned} T_j(k) &= t_j(k) - t_1(k) \\ &= t_j(k) - \sum_{i=1}^{k} \delta t_{in}(i), \end{aligned} \tag{2.47}$$

where i takes integer for integrating the injection time differences for the start cell unit.

$$\delta T_j(k) = t_{j+1}(k) - t_j(k), \tag{2.48}$$

is the residence time of kth RNAPII at G_j. $\Delta T_j(k)$ is the time difference of the elapsed time of kth RNAPII from that of the preceding $(k-1)$th RNAPII to reach G_j, which yields

$$\begin{aligned}\Delta T_j(k) &= T_j(k) - T_j(k-1)\\ &= \left(t_j(k) - \sum_{i=1}^{k} \delta t_{in}(i)\right) - \left(t_j(k-1) - \sum_{i=1}^{k-1} \delta t_{in}(i)\right)\\ &= t_j(k) - t_j(k-1) - \delta t_{in}(k),\end{aligned} \tag{2.49}$$

from Eqs.(2.11) and (2.47). The arrival time delay from the preceding $(k-1)$th RNAPII to the kth RNAPII at G_j is

$$\Delta_j^k = t_j(k) - t_j(k-1), \tag{2.50}$$

from Eqs.(2.15) and (2.49). $\emptyset_j^k$ is introduced as the time interval that G_j is vacant before the kth RNAPII arrives;

$$\emptyset_j^k = t_j(k) - t_{j+1}(k-1), \tag{2.51}$$

from Eqs.(2.22), (2.48) and (2.50). By using Eq.(2.50), $\sum_{i=m+1}^{k-1} \Delta_j^i$ is denoted by $\Delta_j^{(m+1,k-1)}$ as

$$\begin{aligned}\Delta_j^{(m+1,k-1)} &= \sum_{i=m+1}^{k-1} \Delta_j^i\\ &= \sum_{i=m+1}^{k-1} (t_j(i) - t_j(i-1))\\ &= t_j(k-1) - t_j(m),\end{aligned} \tag{2.52}$$

where m $(1 \le m \le k-2)$ is introduced as an integer which takes the condition that the mth RNAPII will proceed smoothly at G_j and all of the preceding RNAPII molecules from the $(m+1)$th to the $(k-2)$th RNAPII will become congested after collisions occur.

§ 3. Results and Discussion

§ 3.1. Application to the *SAMD4A* Gene

The *SAMD4A* gene was selected because it is immediately responsive to the *stimulation*, as well as being quite long in length. The total length of the *SAMD4A* gene is

225398 bp, consisting of 12 exons and 11 introns. The average length of each exon is 183 bp with only small deviations; however, the intron length varies substantially.

Since the length of RNAPII η is about 35 bp, we discretized the *SAMD4A* gene into separate 35 bp box lengths of one RANPII each, and RNAPII moves one box length at a time. The total number of cell units N becomes 6440 in this application. Here the effective velocity of RNAPII in intron is 3.5 (kb/min) [5]; therefore, the time interval δt becomes 0.01 min. Within this time interval, all the RNAPIIs in the gene move at once. To represent the reduction in the velocity of RNAPII in the exons, we maintained the stoppage of RNAPII for γ_e times per time evolution unit. Since it is known that the $3'$ end and $5'$ end in long genes come close together [29, 30], we applied a periodic boundary condition. As one RNAPII goes out from the $3'$ end, RNAPII is immediately re-introduced from the $5'$ end of the gene.

As shown in Fig.2, the RNAPII density profile envelope structure [5] arises from such long-range RNAPII cooperativity, because the suitable distances maintained by a suitable number of RNAPIIs, not a single RNAPII, reproduce the experimental RNAPII density profile over the entire length of the gene. Therefore, since the envelope profile is not sensitive in the correlation of either the stalling profile of the local RNAPII or the correlation of the RNAPIIs with the nucleosomes [16], the evidence suggests the existence of long-range cooperativity between RNAPIIs.

At $t = 30$ min, we compared the results of our simulation to the ChIP-chip results [5]. The simulation result on the envelop curve (Fig.2(b)) generally provides a good fit with the experimental one (Fig.2(a)), except at the two points indicated by the arrows, (ii) and (iii) in Fig.2(a). We observed an increasing enrichment at these positions at the time of the leading RNAPII arrival at position (i) as shown in Fig.2(a).

At $t = 60$ min, the results are also compared with the same optimized parameters at $t = 30$ min. The leading RNAPII moves to position (iv) of Fig.2(a). From the previously reported experimental data [5], the positions indicted by the arrows were enriched with both CTCF and RAD21, which substantially peaked at $t = 30$ min, and at even $t = 0$ minute. Since the CTCF/cohesin (RAD21) complex is tightly related to the looping of DNA, we conjecture that the points (ii) and (iii) indicated the arrows in Fig.2(a) are very close to the position (i) in the real space of the real cell. In fact, other RNAPII experiments show stalling at these sites all the time, starting from $t = 0$ min. (Refer to [31] for more information about these points, (ii) and (iii).) From our simulation for 30 min and 60 min (Fig.2(b)), the sharp peaks are obtained at the exons and CTCF binding sites, that is, at the same sites displaying peaks in experimental data Fig.2(a). Although the existence of blockade by exons and CTCF/cohesin binding in the gene in the normal cell has been shown, the experimentally fitted parameters provide the free homogeneous flow as shown in Fig.3, where the density profile in a spatiotemporal view

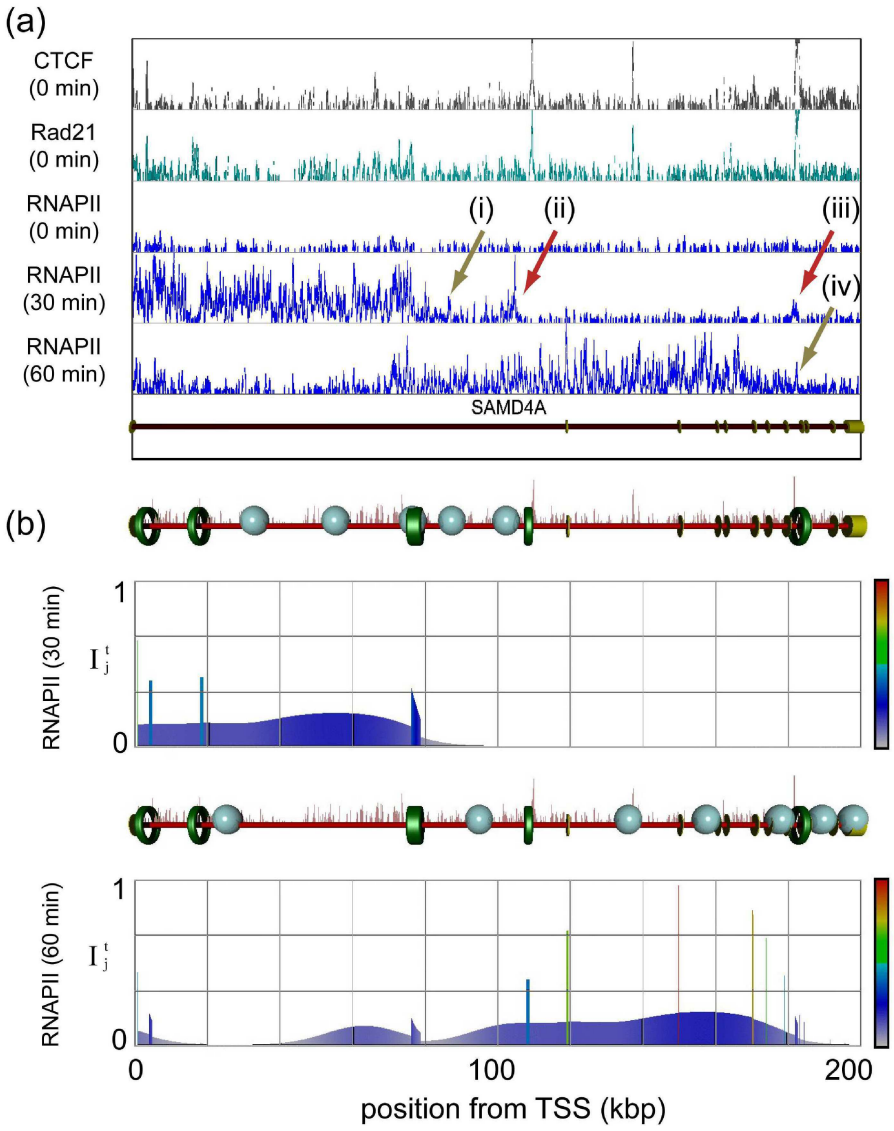

Figure 2. Density Profile of RNAPII and Corresponding CTCF and RAD21 Enrichment (a) The experimental density profile of RNAPII for genes at t=30 and 60 min after the *stimulation*, as observed in the ChIP-chip experiments, and also CTCF/cohesin (RAD21) enrichment at t=0 min [5]. (b) The simulation results at t=30 and 60 min with the optimized parameters: μ=7 and γ=5. The number of RNAPII molecules on the gene body, M=5, and the RNAPII injection timing $\{\delta t_{in}(k)\} = \{0,4,4,6,6\}$ in the gene at t=30 min will be changed to M=7, and $\{\delta t_{in}(k)\} = \{0,4,4,6,6,12,20\}$ at t=60 min.

is provided.

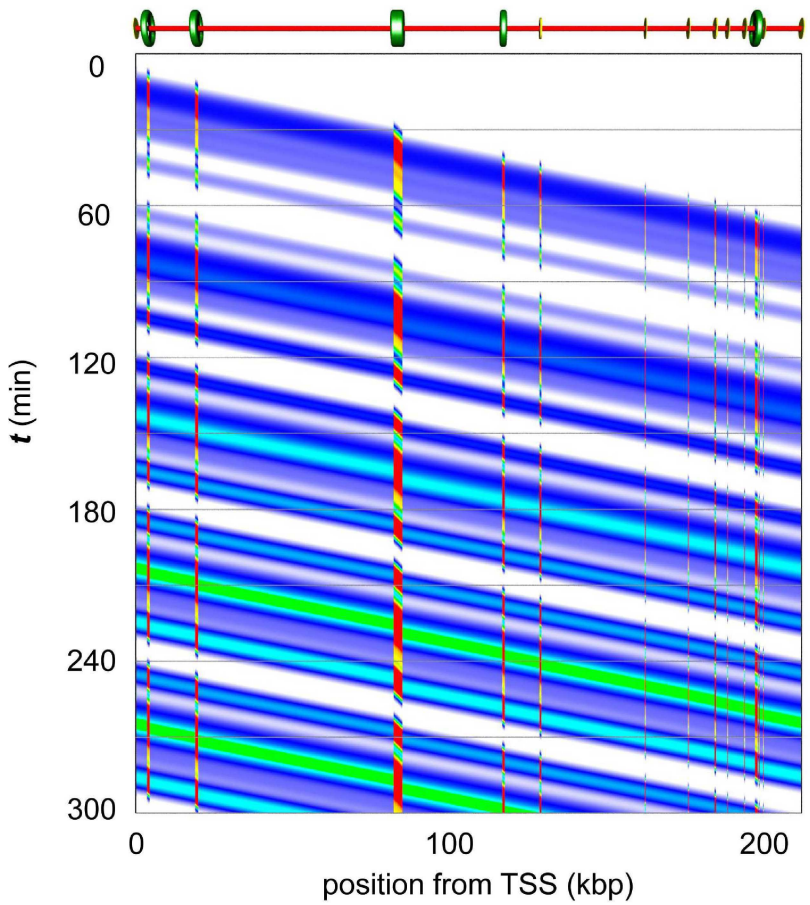

Figure 3. Obtained Time-evolution Pattern in *SAMD4A*

Parameters; γ=10, $0 \leq t \leq 300$ min, and $\{\delta t_{in}(k)\} = \{0,4,4,6,6,12,20,20,20,...\}$ min. The intensity of I_j^t at G_j, in which an RNAPII molecule exists, is shown in Fig.2. From this figure, one can see that the RNAPII molecules gradually assemble in exons and at CTCF/cohesin (RAD21) binding sites. This generates the same homogeneous flow as before, after passing through the γ region, since $\{\delta t_{in}(k)\}$ satisfies the condition in Eq.(2.30).

§ 3.2. Knock Down Cells with a Low Blockade

A lower blockade than that found in normal cells results in both freer and more homogeneous RNAPII motions. Theoretically, knock-down of RAD21 may lead to a partial loosening of the looping of DNA and an inability to retain RNAPII at the cohesin binding positions, and thus the reduction of blockade for RNAPII is achieved. The ChIP-chip experimental result for RAD21 knock-down cell at $t = 60$ min [5] shows that the RNAPII density profile over the entire gene region becomes uniform and more highly enriched than in the control cells (Fig.4(a)). At the well-known hot spots of

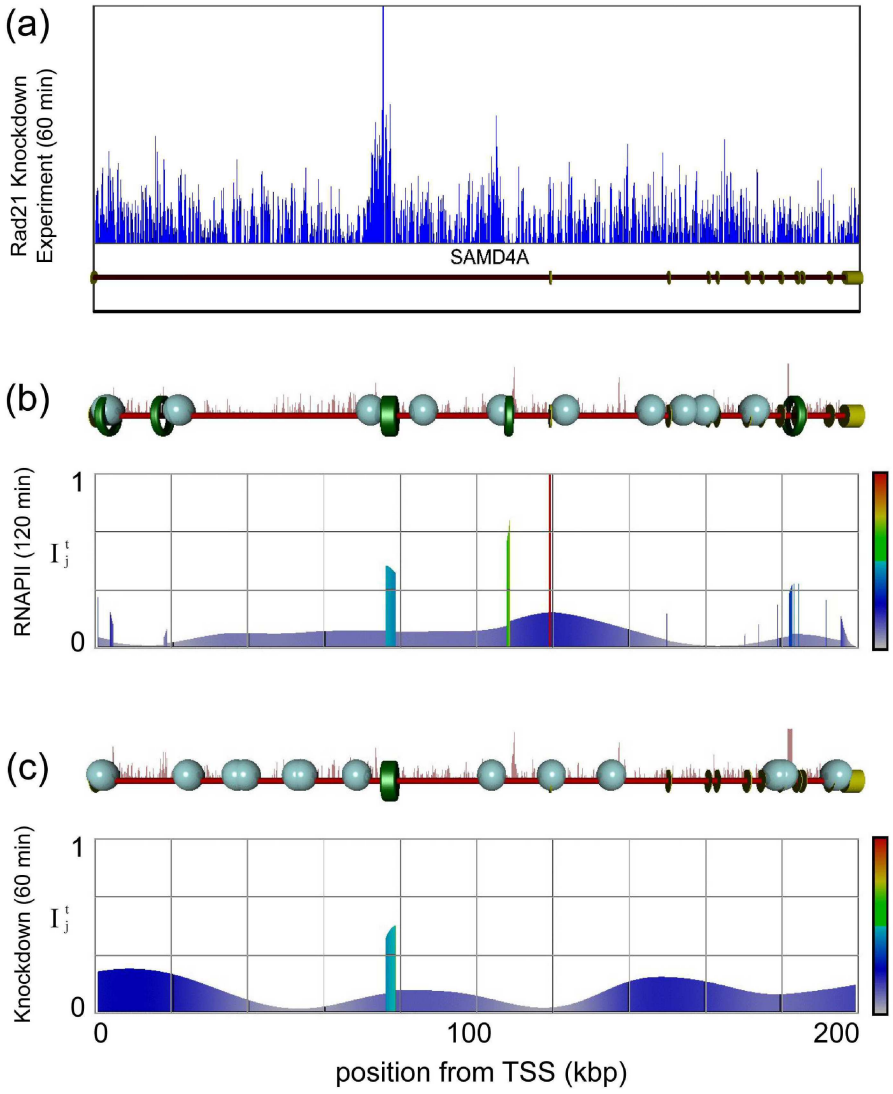

Figure 4. RNAPII Density Profile in Knock-down Cells

(a) Experimental results for the density profile in RAD21 knock-down cells at t=60 min. (b) Simulation results for the RNAPII density profile at t=120 min, where the transcription status of the *SAMD4A* gene is stable and RNAPII is deliberately re-injected, in addition to the transient injections. The parameters are the same as in Fig.2. The number of RNAPII molecules on the gene body M=10, and the RNAPII injection timing is $\{\delta t_{in}(k)\}$ = {0,4,4,6,6,12,20,20,20,20} min. (c) The simulation result of the RNAPII density profile at t=60 min for the knock down cells. Here the simulation parameters are μ=7, γ=1, v_i=v_e=7.0 (kb/min), M=13, and $\{\delta t_{in}(k)\}$ = {0,4,4,6,6,12,20,20,20,20,20,20,20} min.

protein binding sites in the first intron, the RNAPII in the RAD21 knock-down cells is still just as highly enriched as in the normal cells. The experimentally obtained RNAPII density profile for the RAD21 knock-down cells at $t = 60$ min (Fig.4(a)) is similar to the simulation density profile at $t = 120$ min (Fig.4(b)) in normal cells, where more RNAPIIs are assumed to be additionally introduced into the genes than at $t = 60$ min (Fig.4(b)). Figs.4(a) and (b) also indicate that the RNAPII moves rapidly (about two-times faster than that in introns in normal cells) in the knock-down cells.

To simulate the accelerated process in knock-down cells, where a steady state with more RNAPII molecules is already achieved, even at $t = 60$ min, we (tentatively) used the following parameters: $\mu = 7$, $\gamma = 1$, and $v_i = v_e = 7.0$ (kb/min). RNAPIIs with the optimized injection parameters were imposed by means of a periodic boundary condition. RNAPII molecules are allowed to move without stall near the cohesin (RAD21) binding sites, except at hot spot sites where the other protein binding sites are shown by the green ring. The RNAPII density profile thus obtained by the simulation with $\gamma = 1$ (Fig.4(c)) is in good agreement with the experimental result (Fig.4(a)). This indicates that RAD21 knock-down leads to not only a reduction of the RNAPII stall blockade due to the DNA loop, but also a reduction of the RNAPII stall in exons that reported to be regulated by H3K36me3 enrichment. This suggests that the DNA loop formation might occur near exons.

§ 3.3. Global Density Profile of RNAPII with High Congestion

If each arrival time interval satisfies $\Delta_j^k > \Delta_{j(f)}^k$ in the free state (or $\Delta_j^k > \Delta_{j(c)}^k$ in the congested state), the RNAPII flow in the *SAMD4A* gene is free and is governed by the Δ_j^k, even though RNAPII molecules reduce their velocity in the γ regions of exons. Here we used γ=1 for the intron regions, γ=5 for the CTCF-enriched regions and the exons. On the other hand, if one of the time intervals Δ_j^k happens to be satisfied the relation $\Delta_j^k \leq \Delta_{j(f)}^k$ in free state (or $\Delta_j^k \leq \Delta_{j(c)}^k$ in the congested state), the RNAPII flow becomes congested. A spatiotemporal view of the simulation results in this case is given in Fig.5. We also used *SAMD4A* gene structure with the following parameters of the γ regions; γ=1 for the intron regions, γ=5 for the CTCF-enriched regions, and γ=100 for all of the exons.

As shown in Fig.5(a), and from (d) to (f), a regulated repetitive wave with the periodicity $\gamma_{j+1}\eta/v_i$ observed spreading from first encountered exon to the $3'$ end, and then the wave re-enters from the $5'$ end as the result of the periodic boundary condition (Fig.5(g)). Interestingly, the maximum γ occurs at the first encountered exon and the periodicity of the RNAPII density wave passing through the other exons retains the same periodicity. This repetitive wave moves forward and is due to the exon blockade effect. Again, it should be noted that the repetitive pattern shown in Fig.5 is not due to the velocity reduction alone, but also the interaction between the RNAPII molecules, as

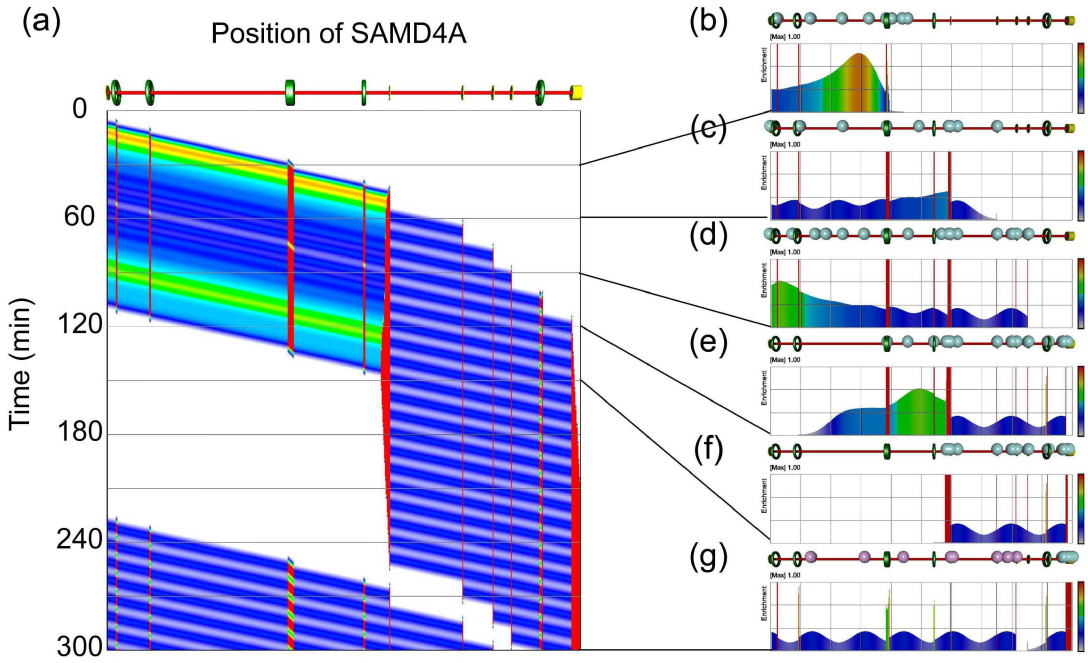

Figure 5. Time-evolution Pattern of Congestion

The obtained time-evolution pattern in *SAMD4A* for the parameters: γ=100, $0 \leq t \leq 300$ min, M=18, and $\{\delta t_{in}(k)\} = \{0,1,2,3,4,5,6,7,8,9,8,7,6,5,4,3,2,1\}$ min. (a) The intensity of I_j^t at G_j, in which an RNAPII molecule exists, is shown. If the injection delay $\delta t_{in}(k)$ is shorter than the critical conversion point in Eq.(2.32), the RNAPII molecules will be congested around the exon region, and after that, they travel with the constant delay $\gamma_{j+1}\eta/v_i$, which resulted in a repetitive pattern of flow over a certain range. (b-g) The RNAPII density profile at (b) t=30, (c) t=60, (d) t=90, (e) t=120, (f) t=150, and (g) t=300 min. In the simulations, a sphere on the gene indicates the preceding RNAPIIs with a distribution for the cell population.

is observed in the prokaryote or the automobile traffic jams, in which the global pattern moves backward with some velocity [21, 22, 32, 33]. In our framework, the RNAPII distribution is mainly due to the change in the difference in the cellular RNAPII density wave, where many solitary waves become uniform due to the broad distribution. The cellular distribution also changes the periodicity of the repetitive pattern, as shown in Fig.5. This clearly indicates that by changing the number of cells, the RNAPII repetitive wave is made experimentally observable.

§ 4. Conclusion

In this report, a cellular automaton model has been successfully applied to RNAPII dynamics in genes. The effect of clearly distinct genomic regions of exons and introns on RNAPII dynamics was investigated, where the RNAPII moves at two different average velocities, to determine the RNAPII spatial distribution. We have described RNAPII dynamics in terms of the properties in time domain such as elapsed time, residence time, and time intervals, instead of those in the space domain; position, length, and distances. As the result of comparing the ChIP-chip experimental results for the *SAMD4A* gene, we showed that long-range RNAPII spatiotemporal cooperativity takes place in RNAPII dynamics. The long-range correlation between RNAPIIs derived here is different from the previously reported for short-range correlations such as local stalling, which are in a range of less than 1 kbp in length and less than 1 min in time. Our model also provides insight into the difference in the RNAPII dynamics of genes in normal cells as well as in cohesin (RAD21) knock-down cells. It is suggested that this formulation of RNAPII dynamics will prove useful in helping to elucidate the RNAPII traffic in other genes.

References

[1] Kornberg RD (2007) The molecular basis of eukaryotic transcription. *Proc. Natl. Acad. Sci. U.S.A.* **104**:12955-12961.
[2] Maniatis T, Reed R (2002) An extensive network of coupling among gene expression machines. *Nature* **416**:499-506.
[3] Orphanides G, Reinberg D (2002) A unified theory of gene expression. *Cell* **108**:439 - 451.
[4] Phatnani HP, Greenleaf AL (2006) Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* **20**:2922-2936.
[5] Wada Y & Ohta Y, et al. (2009) *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18357.
[6] Cook PR, *Principles of Nuclear Structure and Function* (Wiley-Liss Inc., New York, 2001).
[7] Kulaeva OI, Hsieh FK, Studitsky VM (2010) *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11325.
[8] Papantonis A, Larkin JD, Wada Y, Ohta Y, Ihara S, Kodama T, Cook PR (2010) *PLoS Biol.* **8**, e1000419.
[9] Schwartz S, Meshorer E, Ast G (2009) *Nat. Struct. Mol. Biol.* **16**, 990.
[10] Wolfram S (2002) *A New Kind of Science*, Champaign, Illinois.
[11] Klumpp S, Hwa T (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105**:18159.
[12] Tripathi T and Chowdhury D (2008) *Phys. Rev. E* **77**, 011921.
[13] Tripathi T, Schülz GM, Chowdhury D (2009) *J. Stat. Mech: Theo. and Exper*, P08018.
[14] Femino AM, Fay FS, Fogarty K, Singer RH (1998) Visualization of single RNA transcripts in situ. *Science* **280**:585-590.
[15] Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**:77- 88.
[16] Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J (2009) *Nat. Genet.* **41**, 376.
[17] MacDonald C, Gibbs J, Pipkin A (1968) *Biopolymers* **6** (1).

[18] Derrida B (1998) *Phys. Rep.* **301** 65.
[19] Chowdhury D, Guttal V, Nishinari K Schadschneider A (2002) *J. Phys. A*: Math. Gen. **35** pp.L573-L577.
[20] Brooks R, Griffin C, Payne A (2004) *A Cellular, Complexity* **9** , 32.
[21] Nishinari K, Okada Y, Schadschneider A, Chowdlhury D (2005) *Phys. Rev. Lett.* **95** (118101).
[22] Kanai M, Nishinari K, Tokihiro T (2005) *Phys. Rev. E* **72** p.035102 (R).
[23] Greulich P, Garai A, Nishinari K, Schadschneider A, Chowdlhury D (2007) *Phys. Rev. E* **75** (041905).
[24] Wendt KS, et al. (2008) *Nature* **451**, 796.
[25] Mishiro T, et al. (2009) *The EMBO Journal* **28**, 1234.
[26] Cheung ACM, Cramer P, *Nature* **471** (2011) 249.
[27] Galburt EA, et al. (2011) *Nature* **471**, 249.
[28] Ohta Y, Kodama T, Ihara S (2011) *Phys. Rev. E* **86**, 021918.
[29] Yao J, Ardehali MB, Fecko CJ, Webb WW, Lis JT (2007) Intranuclear distribution and local dynamics of RNA polymerase II during transcription activation. *Mol Cell* **28**:978-990.
[30] O'Sullivan JM, Tan-Wong SM, Morillon A, Lee B, Coles J, Mellor J, Proudfoot NJ (2004) Gene loops juxtapose promoters and terminators in yeast. *Nat. Genet.* **36**: 1014-1018.
[31] Ohta Y, Nishiyama A, Wada Y, Ruan Y, Kodama T, Tsuboi T, Tokihiro T, Ihara S (2012) *Phys. Rev. E* **84**, 041922.
[32] Kerner BS, Konhaäser P (1993) *Phys. Rev. E* **48**, R2335.
[33] Sugiyama Y, et al. (2008) *New Journal of Physics* **10**, 033001.