Stochastic Gene Expression Modeling with Hill Function for Switch-Like Gene Responses

Haseong Kim and Erol Gelenbe

Abstract—Gene expression models play a key role to understand the mechanisms of gene regulation whose aspects are grade and switch-like responses. Though many stochastic approaches attempt to explain the gene expression mechanisms, the Gillespie algorithm which is commonly used to simulate the stochastic models requires additional gene cascade to explain the switch-like behaviors of gene responses. In this study, we propose a stochastic gene expression model describing the switch-like behaviors of a gene by employing Hill functions to the conventional Gillespie algorithm. We assume eight processes of gene expression and their biologically appropriate reaction rates are estimated based on published literatures. We observed that the state of the system of the toggled switch model is rarely changed since the Hill function prevents the activation of involved proteins when their concentrations stay below a criterion. In ScbA-ScbR system, which can control the antibiotic metabolite production of microorganisms, our modified Gillespie algorithm successfully describes the switch-like behaviors of gene responses and oscillatory expressions which are consistent with the published experimental study.

Index Terms—Stochastic gene expression modeling, gene regulatory networks, switch-like gene responses, Gillespie algorithm.

1 Introduction

The gene expression is the sum of processes that results in a specific level of mRNAs and proteins. So, the understanding of underlying gene expression mechanisms can be a good starting point to solve many biological problems in systems biology. For example, gene regulatory networks (GRNs) have become main tools to uncover the relationships between genes, whose mechanisms are based on the gene expression changes [1], [2]. Though there are many studies of gene expression using mathematical and statistical models, there are still many controversial issues about how such models can provide decisive insight into gene expression processes. In several recent studies, gene expression is described by discrete and probabilistic processes rather than as continuous and deterministic concentration changes [3], [4]. Also, dynamics of GRNs is investigated using a stochastic modeling approach [5] which enables us to detect abnormal gene behaviors in steady states [6].

The stochasticity known as "gene expression noise" can be either disadvantageous for reduction or beneficial in a sense of evolution [7]. But recently the stochastic nature of biological processes can play important functional roles in a cell such as physiological regulation mechanisms and evolutionary adaptation [8]. Ordinary differential equation (ODE)-based deterministic models do not easily explain these noise oriented cellular functions. So, many ODE-based studies added a noise term to their ODE in order to describe various biological processes such as protein folding [9], oscillatory expressions [10], and network dynamics [11]. For example, Navozhay et al. determined how a negative feedback regulation affects to the gene expression of TetR-based synthetic transcriptional networks [12]. They used stochastic simulations for their gene regulatory model, but ordinary differential equations are also employed to describe the dose responses of the negative reaction which shows S-shape expression profiles in their synthetic experiment.

Stochastic simulation models are also widely used to describe the stochastic gene expression mechanisms [3], [4]. The Gillespie algorithm is generally used in most of the stochastic models to find a possible solution of their stochastic reaction equations [13]. In stochastic models, however, switch-like behavior of a gene is only achieved by gene expression cascade consisting of two more genes, which means the model requires additional set of parameters. It could be a disadvantage of simulation-based approach because of the computation time. The situation could be worse when corresponding parameter values are not known.

In this study, we modified Gillespie algorithm by adding Hill function which describes protein activation. Eight processes of the gene expression are assumed while the expression mechanism is conventionally explained by three stage models (promoter activation, transcription, and translation) [14]. For each stage of the eight processes, its corresponding rate parameters are derived from published information. Our approach is applied to the toggle switch model [15] and the ScbA-ScbR antibiotic metabolite production system of microorganisms [16] to simulate their expression data which show switch-like and oscillatory behavior.

• *The authors are with the Department of Electrical and Electronic Engineering, Imperial College London, South Kensington Campus, London SW7 2BT, United Kingdom. E-mail: {hk308, e.gelenbe}@imperial.ac.uk.*

Manuscript received 25 Mar. 2011; revised 3 July 2011; accepted 13 Nov. 2011; published online 29 Nov. 2011.
For information on obtaining reprints of this article, please send e-mail to: tcbb@computer.org, and reference IEEECS Log Number TCBBSI-2011-03-0074.
Digital Object Identifier no. 10.1109/TCBB.2011.153.

2 Assumptions of Gene Expression Process

2.1 Gene Activation

Depending on cell growth conditions, there are several copies of partially replicated chromosomes [4]. Let n_1 and n_1^{max} are the number of active genes and a constant that maximum copies of genes (chromosomes), respectively. Then, each copy of genes spontaneously switches ON and OFF with rate λ_1 and γ_1, respectively. The stationary distribution for the number of active genes is then Binomial, where the probability of being ON is $Pr_{ON} = \lambda_1/(\lambda_1 + \gamma_1)$. However, in this study, we assume the changing rate of the number of active genes is zero. That is, $dn_1/dt = 0$.

2.2 Transcription

The transcription is typically assumed to follow Exponential distribution [3]. Let RP be RNA polymerase and Pro be corresponding promoter in a DNA strand. After the RNA polymerase binds to the promoter, it switches from a closed complex (RPc) to an open complex (RPo), $RP + Pro \leftrightarrow RPc \leftrightarrow RPo$. The initiation of transcription is only activated by the RNA polymerase-promoter open complex. The time interval between successive transcripts is considered to have an exponential distribution with $1/T_{avg}$, where T_{avg} is the average transcript initiation interval. Then, the probability of the transcription initiation reaction in a small time interval Δt is $1/T_{avg}\exp(-t \cdot 1/T_{avg})\Delta t$, where t is the time. Let λ_2 be the activation rate of RPo then $\lambda_2 = 1/T_{avg}$ and the production rate per unit time is $\lambda_2 \cdot n_1$. The transcription initiation rate, λ_2, of mRNA is assumed $0.0067\ \mathrm{sec}^{-1}$ (1 transcript per 2.5 min) [17]. The reaction equation is

$$RPo \overset{\lambda_2}{\rightarrow} RPo + R, \tag{1}$$

where R represents an mRNA. Note that we only concern RNA polymerase-promoter open complex in this transcription modeling. However, we will soon have more discussions about activator proteins such as transcription factors.

2.3 Translation

In prokaryotic cells, ribosomes bind successively to an mRNA as soon as it is accessible behind the transcribing RNA polymerase. Multiple ribosomes space about 80 nucleotides [3]. Corresponding proteins are produced successively from the attached ribosomes. These translation processes are continued until the mRNA is degraded by an RNAse-E. An RNAse-E is known that it directly competing with ribosomes to bind corresponding promoters because these two types of molecules share their binding site of a promoter. So the competition leads to successful translation or degradation of mRNAs. Then, the probability for the number of proteins, n_3, is $Pr(n_3 = n) = p^n(1-p)$, where p and $(1-p)$ are the probabilities that a ribosome and an RNase-E bind to the mRNA, respectively. Let T_D be an average time interval between successive competitions, then translation initiation rate λ_3 is $\lambda_3 = 1/T_D$. Now, p can be obtained using protein burst size, b. Many studies have interpreted variations of proteins in terms of burst which takes place in brief periods of high expression intensity followed by long periods of low intensity [17], [18]. Let's assume the burst size is equal to an average number of proteins per one mRNA. Then, the probability that an ribosome binds to an mRNA is computed by solving $b = p/(1-p)$ [3]. All the appropriate parameter values of the λ_3, γ_3, b, and p will be explained along with degradations of mRNA and proteins. Equation (2) depicts the chemical equation of the translation process

$$R(t) \overset{\lambda_3}{\rightarrow} R(t + \Delta t) + P(t + \Delta t). \tag{2}$$

The time delay, Δt, is employed to describe post-transciptional processes such as RNA splicing [19]. It is also known that the time delays induce stochastic oscillations. Bratsun et al. simulated stochastic models with delayed time in mRNA/protein degradation and negative interactions processes [19] while McAdams and Arkin studied a delayed time for protein activations after their promoters are activated [3]. Δt is assumed 0.5 in this study.

2.4 Protein Dimerization

Proteins act in a form of a dimer or an oligomer rather in isolation. It is known that the protein multimerization is an important feature in regulations of biomolecules such as enzyme and transcription factors [20]. The dimer dissociation constant K_{D1} is varied from 0.1 to 10, where $K_{D1} = [P]^2/[PP] \approx k_{d1}/k_{a1}$; k_{a1} and k_{d1} are the dimer association and dissociation rate, respectively [21]. In this study, $K_{D1} = 0.1, k_{a1} = 1.0 \times 10^6\ \mathrm{M}^{-1}\mathrm{sec}^{-1}$

$$P + P \underset{k_{d1}}{\overset{k_{a1}}{\rightleftharpoons}} PP. \tag{3}$$

The protein concentration growth is affected by additional parameters: initial cell volume and cell growth rate. We fix the cell growth rate and the cell volume is considered 1.6×10^{-15} L. Let c_{a1} be the stochastic rate of the dimerization then c_{a1} can be converted from k_{a1}, $c_{a1} = (2 \times 10^6 \mathrm{M}^{-1}\mathrm{sec}^{-1})/(6.023 \times 10^{23}\mathrm{mol}^{-1} \times 1.6 \times 10^{-15}\mathrm{L}) \approx 0.002\ \mathrm{sec}^{-1}$ (Note that if the reaction is second order, $c_{a1} = (2 \times k_{a1})/(n_A \times V)$ where n_A is Avogadro constant and V is a volume of a cell) [22]. With this dimer association and dimer dissociation constant, we can obtain that the stochastic rate of the monomerization, c_{d1} is $0.0002\ \mathrm{sec}^{-1}$.

2.5 Degradation of mRNA and Protein

Generally, the nth order reaction equation is defined by $(d[A])/dt = k[A]^n$, where k is reaction constant and $[A]$ is the concentration of substance A (mRNA or protein). When the order is one ($n = 1$), the decrease in the concentration of A over time is $\log([A]/[A]_0) = -kt$, where $[A]_0$ is the initial concentration of A at time $t = 0$. So the half-life of A is $T_{half} = \log(2)/k$. Then, the surviving mRNAs, n_2, in the population after transcription is blocked would be $n_2 = n_{2,0}p^{T_{half}/T_D}$. This is equal to $T_{half} = -(\log(2)/\log(p))T_D$ [3]. The average half-life of mRNAs is assumed by 300 sec [21]. Therefore, the translation initiation rate $\lambda_3 = 1/T_D$ can be computed with appropriate p which is obtained by using the protein burst size b. Note that b is varied; 100 for an average *E. coli* gene [4], 40 for lacZ, and 5 for lacA [17]. In this study, b is assumed 20 so the probability that a ribosome binds to the mRNA, p, is 0.952. Then, the average time interval between successive competitions of ribosome and RNase-E, T_D, is $21.1168\ \mathrm{sec}$. Therefore, the translation initiation rate λ_3 is $0.04736\ \mathrm{sec}^{-1}$. The degradation rate of

mRNA, γ_2, is $0.00231\ \mathrm{sec}^{-1}$ $(= log(2)/300\ \mathrm{sec})$. The average of half-life of proteins is known as from 15 to 120 min [17]. It is known that dimers or oligomers are more stable than their monomeric components [21]. In our study, the half-life of monomer proteins and dimer proteins are assumed as 15 and 20 min, respectively. So the degradation rates of monomers (γ_3) and dimers (γ_4) are $0.00077\ \mathrm{sec}^{-1}$ and $0.000578\ \mathrm{sec}^{-1}$.

2.6 Repressor-Operator (RO) Association

A repressor is a DNA binding protein that inhibits the transcription of genes. The activation of an RNA polymerase is blocked when a represser binds to the corresponding gene's operator

$$P + O \underset{k_{d2}}{\overset{k_{a2}}{\rightleftharpoons}} PO, \tag{4}$$

where P, O, and PO represent a repressor protein, operator, and repressor-operator complex, respectively. k_{a2} and k_{d2} are the repressor-operator complex association and dissociation rates. $K_{D2}(= k_{d2}/k_{a2} = [P][O]/[PO])$ is the PO complex equilibrium constants where the rate of dissociation of the PO complex is significantly affected by changes in association rate constants. K_{D2} is varied from 1 to 1,000 nM and we assume $K_{D2} = 10$ [21]. The kinetics of association (k_{a2}) is known to be from $1 \times 10^8\ \mathrm{M}^{-1}\ \mathrm{sec}^{-1}$ to $2 \times 10^9\ \mathrm{M}^{-1}\ \mathrm{sec}^{-1}$ [23]. We use $1 \times 10^8\ \mathrm{M}^{-1}\ \mathrm{sec}^{-1}$ so the converted association stochastic rate, c_{a2}, is $1 \times 10^8\ \mathrm{M}^{-1}\ \mathrm{sec}^{-1}/(6.023 \times 10^{23}\mathrm{mol}^{-1} \times 1.6 \times 10^{-15}\mathrm{L}) = 0.104\ \mathrm{sec}^{-1}$ and the dissociation stochastic rate, c_{d2}, is $1.04\ \ \mathrm{nM/sec}$.

2.7 Activator-Promoter/Enhancer Association

An activator is a DNA bind protein that increases the RNA polymerase activity. This type of transcription factor binds to either enhancer or promoter regions of DNA strand. After the activators bind to the enhancer, the RNA polymerase is stabilized and activates the transcription. So, not only the transcription initiation rate λ_2 in (1), the concentration of activators also affects to the mRNA production. So we have the following reaction equation that produces RNA polymerase-Promoter complex (RPo),

$$PP + Pro \underset{k_{d2}}{\overset{k_{a2}}{\rightleftharpoons}} RPo, \tag{5}$$

where PP can be either activators and RNA polymerases. We use the same kinetic canstants with the repressor-operator association because this reaction is also a kind of DNA-protein associations.

2.8 Phosphorylation and Ubiquitination

Phosphorylation is a process that adds a phosphate (PO_4) group to a protein. This process turns many enzymes' activity on and off, which may cause or prevent the mechanisms of diseases. Enzymes involved in this phosphorylation process are called kinases. The reversible process of phosphorylation is dephosphorylation which removes the phosphate group from a phosphorylated protein. An enzyme called phosphatases performs this dephosphorylation. A protein activated by the phosphorylation has several possible phosphorylation sites on its amino acid residues such as serine, threonine, and tyrosine. Many modeling approaches simplify the expression processes including the multiple phosphorylation by introducing time delays to the model [3], [24]. However, we introduce a reaction equation to represent a phosphorylation process that a kinase phosphorylates an enzyme

$$PP_{kinase} + PP \underset{k_{d3}}{\overset{k_{a3}}{\rightleftharpoons}} PP_{kinase} + PPp, \tag{6}$$

where PPp represents the phosphorylated protein.

Ubiquitination is also a post-translational protein modification process that labels proteins to promote their degradation. Though the ubiquitination process requires E1, E2, and E3 enzymes, we simply present the process with an E2-E3 ligase and a corresponding protein ubiquitinated by the ligase

$$PP_{ligase} + PP \underset{k_{d3}}{\overset{k_{a3}}{\rightleftharpoons}} PP_{kinase} + PPu, \tag{7}$$

$$PPu \xrightarrow{\gamma_5} \phi, \tag{8}$$

where PPu denotes the ubiquitinazed protein. Equation (7) represents the degradation process of the ubiquitinated protein. In this study, all protein-protein association/disassociation rate are assumed to be the same $k_{a3} = k_{a1}$ and $k_{d3} = k_{d1}$. Also the degradation rate γ_5 is set to be the same as the dimer protein degradation rate, $\gamma_5 = \gamma_4$. In these post-translational processes, time delays can be also introduced to describe multiple phosphorylation/ubiquitinations.

3 Hill Function for Protein Activation

In an ordinary differential equation model, the net change of the mRNA concentration is generally described by a simple rate equation, $dn_2/dt = \lambda_2 \cdot g([P]) - \gamma_2 \cdot n_2$, where $g([P])$ is the Hill function of protein P (transcription factor) [26]. This Hill function $g(x)$ is generally defined as follows:

$$g(x) = \frac{x^n}{K^n + x^n}, \tag{9}$$

where K is the threshold for the regulatory influence of x on a target gene, and n is a steepness parameter. This function increases from 0 to 1 as x goes ∞ and the inhibitory function is $1 - g(x)$. In our study, this Hill function is used to compute the propensity of protein activation in the Gillespie algorithm.

The goal of the Gillespie algorithm is to estimate the state vector $\mathbf{X}(t) = (X_1(t), \ldots, X_N(t))$, given the system initial state $\mathbf{X}(t_0) = \mathbf{x}_0$, where $X_i(t)$ denotes the number of molecules of species S_i in the system at time t. Consider M chemical reactions $\{R_1, \ldots, R_M\}$. Each reaction R_j is characterized by two quantities, state-change vector $\mathbf{v}_j = (v_{1j}, \ldots, v_{Nj})$ and propensity function a_j. v_{ij} is the change of the ith molecular population caused by one R_j reaction. If the system is in state $\mathbf{x}$ and R_j occurs than the next state of the system will be $\mathbf{x} + \mathbf{v}_j$. The propensity function a_j can be defined so that $a_j(x)dt = Pr\{R_j \text{ will occur in } [t, t + dt) | \mathbf{X}(t) = \mathbf{x}\}$ and

$$a_j(x) = c_j[x], \tag{10}$$

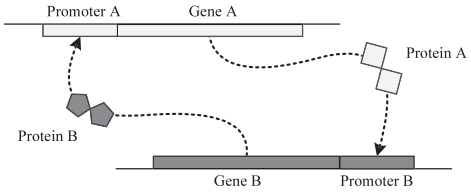

Fig. 1. The toggle switch system.

where $[x]$ the concentration of the involved species x and c_j is the kinetics of the reaction R_j. We add the Hill function (9) to (10) to describe the switch-like behaviors of a gene

$$a_j(x) = c_j[x] \cdot \frac{[x]^n}{K^n + [x]^n}. \quad (11)$$

For example, let's assume a dimer protein PP_A represses the expression of gene B by binding to the operator of gene B

$$Pro_B + PP_A \overset{c_{a2}}{\rightarrow} Pro_B PP_A. \quad (12)$$

Then, the probability that PP_A will undergo this reaction in the next dt is $[Pro_B][PP_A]c_{a2}dt$. So the propensity function is $a_j([PP_A], [Pro_B]) = c_{a2}[Pro_B][PP_A]$. But if the gene B is known to have the switch-like response, its propensity function is redefined as follows:

$$a_j(PP_A, Pro_B) = c_{a2} \cdot [Pro_B] \cdot [PP_A] \cdot \frac{[PP_A]^n}{K^n + [PP_A]^n}. \quad (13)$$

In this study, K and n are assumed 50 and 7, respectively.

4 Modeling Gene Regulatory Networks

4.1 Toggle Switch Gene Expression Model

The toggle switch system, proposed by Gardner et al. [15], consists of two repressor molecules and corresponding two inducers. In their model, a switch is achieved by controlling the levels of inducers. In this study, however, we focus on the effect of Hill functions on the behavior of the two genes without their inducers. Fig. 1 shows our toggle switch system. Each promoter is inhibited by the repressor which is transcribed by the opposing promoter

$$Pro_i(t) + TF_i(t) \overset{\lambda_2}{\rightarrow} Pro_i(t) + TF_i(t) + R_i(t), \quad (14)$$

$$R_i(t) \overset{\lambda_3}{\rightarrow} R_i(t + \Delta t) + P_i(t + \Delta t), \quad (15)$$

$$P_i(t) + P_i(t) \underset{c_{d1}}{\overset{c_{a1}}{\rightleftharpoons}} PP_i(t), \quad (16)$$

$$Pro_i(t) + PP_j(t) \underset{c_{d2}}{\overset{c_{a2}}{\rightleftharpoons}} Pro_i PP_j(t), \quad (17)$$

$$\begin{aligned} R_i(t) &\overset{\gamma_2}{\rightarrow} \phi(t), \\ P_i(t) &\overset{\gamma_3}{\rightarrow} \phi(t), \\ PP_i(t) &\overset{\gamma_4}{\rightarrow} \phi(t), \end{aligned} \quad (18)$$

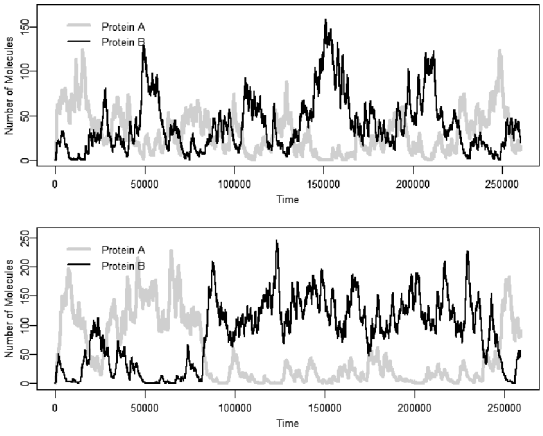

Fig. 2. The protein and mRNA expressions of the toggle switch model with (upper) and without (lower) the Hill funtions.

where $i, j \in \{A, B\}$. Equation (14) shows the transcription process, where TF_i is a transcription factor. Equations (15) and (16) are translation and dimeriztion processes, respectively. Equation (17) represents the repression process of protein A binding to the promoter B. So the expression of protein B can be the reporter of the repressor-inducer binding activity.

It is known that a negative feedback engenders the oscillatory expression of genes. Fig. 2 shows our simulated protein expression profiles of gene A (gray) and B (black) when Hill function is not applied (upper) and is applied (lower). In both cases, the system state switches depending on the concentrations of proteins A and B. However, the system with Hill function does not change its states as much as those of the system without Hill function. It is mainly because Hill function prevents fluctuations of the protein expressions if the signal strength of an activator or an inhibitor is below its threshold.

4.2 ScbA-ScbR System

Many microorganisms produce variety of secondary metabolites which have antibiotic activities to take an advantage in their survival, and some of them are commercially valuable. Antibiotics, however, are often toxic even to their producers so some microorganisms make signalling molecules to control the production of the antibiotics [16]. SCB1 is known that its addition to agar culture develops localized antibiotic production in *S. coelicolor* [27]. In Fig. 3, A gene pair, *scbA* and *scbR* is involved in the regulation of the synthesis of SCB1 [16]. These two genes encode a cytoplasmic receptor protein (ScbR) and an amplifier protein (ScbA), respectively. ScbR can directly bind to the promoter of cryptic type I polyketide synthase gene cluster (cpk) and represses its expression. cpk is a group of genes required for the biosynthesis of a given secondary metabolite and its genes are found to be the principle target of the extracellular regulatory molecule, SCB1 [28].

However, the production of this gene cluster has not been determined. Mehra et al. [16] studied their hypothesis of the formation of a ScbA-ScbR complex using mathematical models and they showed their system exhibits bistable and the states of the system are switched in response to a threshold SCB1. In this system, cells are initially OFF steady-state where ScbR represses the expression of cpk genes. If the concentration of SCB1 is over a certain criterion, the state

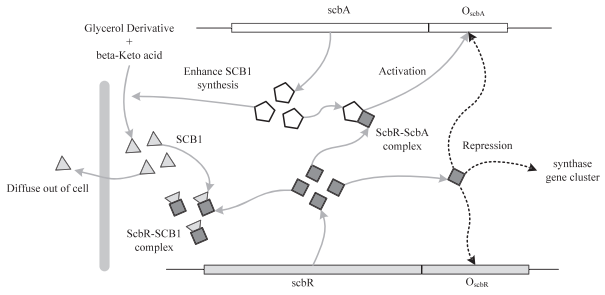

Fig. 3. ScbR/ScbA System.

is turned ON which means SCB1 inhibits the activity of ScbR so cpk genes are expressed. We simulate this system having the Hill type reaction in binding the SCB1 and ScbR molecules which are observed as the reporters of the system.

Following equations represent the reactions of our simulated system (Pol: RNA polymerase, SR: ScbR, SA: ScbA, SB: SCB1). Reactions for ScbR are shown in (19)-(23) and (24)-(28) are for ScbA. ScbR activates SCB1 from the inactive form of SCB1 (P^*_{SB}) in (29). Hill function is applied in (30) for the ScbR-SCB1 binding reaction. Repressions of ScbR are depicted in (33) and (34)

$$Pro_{SR}(t) + TFPol_{SR}(t) \xrightarrow{\lambda_2} Pro_{SR}(t) + TFPol_{SR}(t) + R_{SR}(t), \tag{19}$$

$$TF_{SR}(t) + Pol_{SR}(t) \underset{c_{d2}}{\overset{c_{a2}}{\rightleftharpoons}} TFPol_{SR}(t), \tag{20}$$

$$R_{SR}(t) \xrightarrow{\lambda_3} R_{SR}(t + \Delta t) + P_{SR}(t + \Delta t), \tag{21}$$

$$R_{SR}(t) \xrightarrow{\gamma_2} \phi(t), \tag{22}$$

$$P_{SR}(t) \xrightarrow{\gamma_3} \phi(t), \tag{23}$$

$$Pro_{SA}(t) + TFPol_{SA}(t) \xrightarrow{\lambda_2} Pro_{SA}(t) + TFPol_{SA}(t) + R_{SA}(t), \tag{24}$$

$$P_{SA}P_{SR}(t) + Pol_{SA}(t) \underset{c_{d2}}{\overset{c_{a2}}{\rightleftharpoons}} TFPol_{SA}(t), \tag{25}$$

$$R_{SA}(t) \xrightarrow{\lambda_3} R_{SA}(t + \Delta t) + P_{SA}(t + \Delta t), \tag{26}$$

$$R_{SA}(t) \xrightarrow{\gamma_2} \phi(t), \tag{27}$$

$$P_{SA}(t) \xrightarrow{\gamma_3} \phi(t), \tag{28}$$

$$P_{SA}(t) + P^*_{SB} \xrightarrow{c_{a3}} P_{SA}(t) + P_{SB}(t), \tag{29}$$

$$P_{SB}(t) + P_{SR}(t) \underset{c_{d1}}{\overset{c_{a1}}{\rightleftharpoons}} P_{SB}P_{SR}, \tag{30}$$

$$P_{SB}(t) \xrightarrow{\gamma_3} \phi(t), \tag{31}$$

TABLE 1
Parameters of Stochastic Gene Expression Model

Parameters		Values	References
Transcription initiation	λ_2	$0.0067\ sec^{-1}$	[17]
Translation initiation	λ_3	$0.0474\ sec^{-1}$	[4], [17]
mRNA degradation	γ_2	$0.00023\ sec^{-1}$	[17]
Monomer degradation	γ_3	$0.00077\ sec^{-1}$	[25], [17]
Dimer degradation	γ_4, γ_5	$0.00058\ sec^{-1}$	[25], [17]
Dimer association	c_{a1}	$0.0023\ sec^{-1}$	[25]
Dimer dissociation	c_{d1}	$0.00023\ sec^{-1}$	[25]
DNA-protein association	c_{a2}	$0.1038\ sec^{-1}$	[23]
DNA-protein dissociation	c_{d2}	$1.04\ sec^{-1}$	[23]
protein-protein association	c_{a3}	$0.0023\ sec^{-1}$	
protein-protein dissociation	c_{d3}	$0.00023\ sec^{-1}$	
Burst size	b	20	[4], [17]
Time delay	Δt	0.5	[19]
Michaelis-Menten constant	K	50	
Cell growth rate and its cell volume are fixed.			

$$P_{SB}P_{SR}(t) \xrightarrow{\gamma_4} \phi(t), \tag{32}$$

$$P_{SR}(t) + Rpo_{SR}(t) \underset{c_{d2}}{\overset{c_{a2}}{\rightleftharpoons}} P_{SR}Rpo_{SR}, \tag{33}$$

$$P_{SR}(t) + Rpo_{SA}(t) \underset{c_{d2}}{\overset{c_{a2}}{\rightleftharpoons}} P_{SR}Rpo_{SA}, \tag{34}$$

$$P_{SA}(t) + P_{SR}(t) \underset{c_{d1}}{\overset{c_{a1}}{\rightleftharpoons}} P_{SA}P_{SR}, \tag{35}$$

$$P_{SA}P_{SR}(t) \xrightarrow{\gamma_4} \phi(t). \tag{36}$$

Though, the parameters in Table 1, say "$C1$," are derived from published information to reduce the parameter space, it is necessary to find a set of appropriate parameter values ("$C2$") so as to this Scb-ScR system shows oscillatory expression pattern triggered by the negative feedback of ScbR. The empirically found parameter values, $C2$, include $c_{a1} = 0.0005, c_{d1} = 0.0001, c_{a2} = 0.1, c_{d2} = 0.1, c_{a3} = 0.000001$, and $c_{d3} = 0.00008$. All the other parameters are the same as Table 1.

In Fig. 4, the system clearly shows the oscillatory expression caused by the negative autoregulatory feedback of ScbR [29]. Also Fig. 5 illustrates the switch-like behavior of the system. The levels of ScbA (Right panel) and ScbR (Left panel) decreases and increases, respectively, as SCB1 concentration rises. Hill function applied system (rectangle)

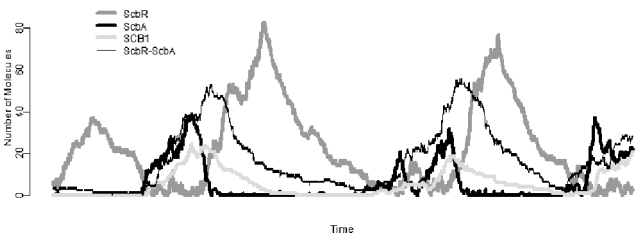

Fig. 4. Oscillatory expression profiles engendered by the negative autoregulatory feedback of ScbR (thick and dark gray). As ScbR level decreases, ScbA (thick and black) expression increases, which causes SCB1 (thick and light gray) level growth. On the other hand, the repression effect of ScbR on itself and ScbA is recovered again as ScbR expression increases.

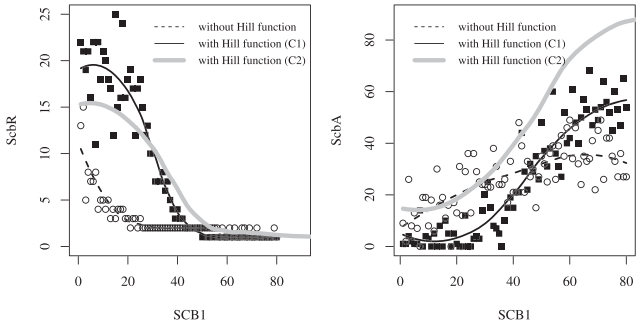

Fig. 5. Switch like behavior of ScbR (Left) and ScbA (Right) where solid line with rectangle and dotted line with circle represent the corresponding molecules with and without Hill function, respectively, in the condition $C1$. Gray line depicts the switch like behavior of each molecule with the condition $C2$. The lines are predicted values using the Local regression method (Loess).

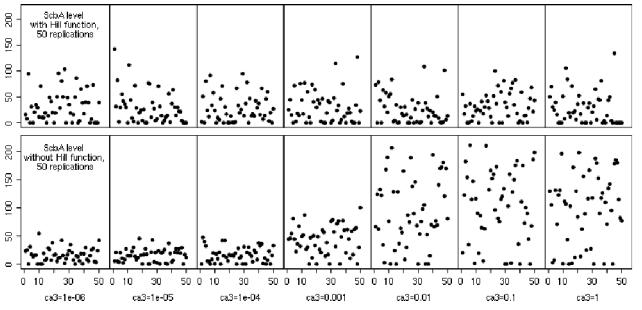

Fig. 6. ScbA expression levels by varying the value of c_{a3} in (29) in condition $C2$. There are 50 replications for each value of c_{a3} and each dot represents the mean of ScbA expression for 50,000 sec. Upper and lower panels show the ScbA level with and without Hill function, respectively.

shows the S-shape curve of each molecules depending on the level of SCB1 in both conditions of $C1$ (black solid lines) and $C2$ (gray solid lines). In these two different conditions, the system shows the similar switch-like expression patterns caused by the Hill function.

Additionally, we looked into the sensitivity of parameter c_{a3} in (29) where Hill function is applied. In Fig. 6, the mean values of ScbA expression in the system with Hill function are distributed between 0 and 100 for all c_{a3} values while the non-Hill function applied system shows sudden expression burst up to 200 when $c_{a3} \geq 0.001$. Note that, in condition $C2$, the ScbA-ScbR system has both oscillatory and switch-like behavior within 100 protein level. So Hill function makes a parameter less sensitive which could be an advantage in modeling with stable gene expression. However, becoming low-sensitive parameter could be also a drawback in finding optimal parameters, which will be discussed in Section 5.

5 Discussion

In ODE models, Hill function is commonly used to describe the chemical reactions of metabolites and the switch-like gene responses. In this study, we propose a delicate stochastic gene expression modeling method with Hill function which can successfully characterize the switch-like gene behavior. Several important biological properties including protein dimerization, time delay of transcription, and DNA-protein, protein-protein interactions are considered and their stochastic rate parameters are estimated based on published information.

Our toggle switch system with Hill function shows robust expression patterns against the internal/external perturbations as its Hill function makes the system stable under the condition below a certain criterion. The resistance of our body to external factors causing a disease might be an example of this robust gene response. Though the original paper of the ScbA-ScbR system shows the possible three steady-states (OFF, ON, and bi-stable) along with their possible parameter spaces, our study shows only switch-like behaviors of the system with fixed parameters as this study focuses on the effect of Hill function.

Though our model shows successful switch-like behavior, Hill function could be a drawback of stochastic models when we search for optimal parameter values because the system with Hill function becomes less sensitive. For example, in Fig. 6, we observed that the system is stable with $c_{a3} \geq 0.001$ while Hill function applied system provide no evidence of this information. So Hill function should be carefully applied to a system that requires detailed understandings of gene expression mechanism. Otherwise, we can apply Hill function to the pathway which is less important but need to be stable expression.

As future work, we will investigate stochastic modeling-based large-scale gene regulatory networks which is not easy to model using conventional simulation method such as the Gillespie algorithm. Probability models that lead to the work that we conduct in the field of gene regulatory networks [30], start with discrete queuing network models and their continuous counterparts [31], [32], related probability models followed by the introduction of G-Networks in the 1990's [33], [34], [35] and their counterparts for biologically plausible spiked neuronal networks [36], [37]. In all of this work, it has been important to express analytical solutions for complex stochastic networks so that the computational effort and potential numerical errors can be reduced as much as possible. Related work with an engineering flavor [38] has aimed at obtaining useful analytical solutions for genetic algorithms, so that closed form expressions can be used to replace lengthy Monte Carlo simulations. Other related work on obtaining analytical solutions to complex mult-iparticle stochastic models relates to the interactions and motions of molecules [37], [39]. Such analytical results also have the advantage of avoiding the need to truncate very large or infinite systems of equations that describe the related Markov processes, thus avoiding the inevitable truncation errors when an infinite or very large system of equations is replaced by a smaller number of equations.

References

[1] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, "Using Bayesian Networks to Analyze Expression Data," *J. Computational Biology*, vol. 7, nos. 3/4, pp. 601-620, 2000.

[2] R. Opgen-Rhein and K. Strimmer, "Learning Causal Networks from Systems Biology Time Course Data: An Effective Model Selection Procedure for the Vector Autoregressive Process," *BMC Bioinformatics*, vol. 8, no. 2, article S3, 2007.

[3] H. McAdams and A. Arkin, "Stochastic Mechanisms in Gene Expression," *Proc. Nat'l Academy of Sciences USA*, vol. 94, no. 3, pp. 814-819, 1997.

[4] J. Paulsson, "Models of Stochastic Gene Expression," *Physics of Life Rev.*, vol. 2, no. 2, pp. 157-175, 2005.

[5] E. Gelenbe, "Steady-State Solution of Probabilistic Gene Regulatory Networks," *J. Physical Rev. E,* vol. 76, p. 031903, 2007.
[6] H. Kim and E. Gelenbe, "Anomaly Detection in Gene Expression via Stochastic Models of Gene Regulatory Networks," *BMC Genomics,* vol. 10, no. Suppl 3, article S26, 2009.
[7] H. Maamar, A. Raj, and D. Dubnau, "Noise in Gene Expression Determines Cell Fate in Bacillus Subtilis," *Science,* vol. 317, no. 5837, pp. 526-529, 2007.
[8] A. Eldar and M. Elowitz, "Functional Roles for Noise in Genetic Circuits," *Nature,* vol. 467, no. 7312, pp. 167-173, 2010.
[9] J. Wang, L. Xu, and E. Wang, "Potential Landscape and Flux Framework of Nonequilibrium Networks: Robustness, Dissipation, and Coherence of Biochemical Oscillations," *Proc. Nat'l Academy of Sciences USA,* vol. 105, no. 34, pp. 12271-12276, 2008.
[10] J. Chahine, R. Oliveira, V. Leite, and J. Wang, "Configuration-Dependent Diffusion Can Shift the Kinetic Transition State and Barrier Height of Protein Folding," *Proc. Nat'l Academy of Sciences USA,* vol. 104, no. 37, pp. 14646-14651, 2007.
[11] P. Ao, "Global View of Bionetwork Dynamics: Adaptive Landscape," *J. Genetics and Genomics,* vol. 36, no. 2, pp. 63-73, 2009.
[12] D. Nevozhay, R. Adams, K. Murphy, K. Josić, and G. Balázsi, "Negative Autoregulation Linearizes the Dose-Response and Suppresses the Heterogeneity of Gene Expression," *Proc. Nat'l Academy of Sciences USA,* vol. 106, no. 13, pp. 5123-5128, 2009.
[13] D. Gillespie, "Stochastic Simulation of Chemical Kinetics," *Ann. Rev. Physical Chemistry,* vol. 58, pp. 35-55, 2007.
[14] V. Shahrezaei and P. Swain, "Analytical Distributions for Stochastic Gene Expression," *Proc. Nat'l Academy of Sciences USA,* vol. 105, no. 45, pp. 17256-17261, 2008.
[15] T. Gardner, C. Cantor, and J. Collins, "Construction of a Genetic Toggle Switch in Escherichia Coli," *Nature,* vol. 403, pp. 339-342, 2000.
[16] S. Mehra, S. Charaniya, E. Takano, and W. Hu, "A Bistable Gene Switch for Antibiotic Biosynthesis: The Butyrolactone Regulon in Streptomyces coelicolor," *PLoS One,* vol. 3, no. 7, p. e2724, 2008.
[17] M. Thattai and A. van Oudenaarden, "Intrinsic Noise in Gene Regulatory Networks," *Proc. Nat'l Academy of Sciences USA,* vol. 98, no. 15, pp. 8614-8619, 2001.
[18] L. Cai, N. Friedman, and X. Xie, "Stochastic Protein Expression in Individual Cells at the Single Molecule Level," *Nature,* vol. 440, no. 7082, pp. 358-362, 2006.
[19] D. Bratsun, D. Volfson, L. Tsimring, and J. Hasty, "Delay-Induced Stochastic Oscillations in Gene Regulation," *Proc. Nat'l Academy of Sciences USA,* vol. 102, no. 41, pp. 14593-14598, 2005.
[20] N. Marianayagam, M. Sunde, and J. Matthews, "The Power of Two: Protein Dimerization in Biology," *Trends in Biochemical Sciences,* vol. 29, no. 11, pp. 618-625, 2004.
[21] N. Buchler, U. Gerland, and T. Hwa, "Nonlinear Protein Degradation and the Function of Genetic Circuits," *Proc. Nat'l Academy of Sciences USA,* vol. 102, no. 27, pp. 9559-9564, 2005.
[22] D. Wilkinson, *Stochastic Modelling for Systems Biology.* Chapman and Hall/CRC, 2006.
[23] D. Goeddel, D. Yansura, and M. Caruthers, "Binding of Synthetic Lactose Operator DNAs to Lactose Repressors," *Proc. Nat'l Academy of Sciences USA,* vol. 74, no. 8, pp. 3292-3296, 1977.
[24] P. Smolen, D. Baxter, and J. Byrne, "Modeling Circadian Oscillations with Interlocking Positive and Negative Feedback Loops," *J. Neuroscience,* vol. 21, no. 17, pp. 6644-6656, 2001.
[25] N. Buchler, U. Gerland, and T. Hwa, "Nonlinear Protein Degradation and the Function of Genetic Circuits," *Proc. Nat'l Academy of Sciences USA,* vol. 102, no. 27, pp. 9559-9564, 2005.
[26] H. De Jong, "Modeling and Simulation of Genetic Regulatory Systems: A Literature Review," *J. Computational Biology,* vol. 9, no. 1, pp. 67-103, 2002.
[27] E. Takano, T. Nihira, Y. Hara, J. Jones, C. Gershater, Y. Yamada, and M. Bibb, "Purification and Structural Determination of SCB1, a γ-Butyrolactone that Elicits Antibiotic Production in Streptomyces Coelicolor A3 (2)," *J. Biological Chemistry,* vol. 275, no. 15, pp. 11010-11016, 2000.
[28] E. Takano, H. Kinoshita, V. Mersinias, G. Bucca, G. Hotchkiss, T. Nihira, C. Smith, M. Bibb, W. Wohlleben, and K. Chater, "A Bacterial Hormone(the SCB 1) Directly Controls the Expression of a Pathway-Specific Regulatory Gene in the Cryptic Type I Polyketide Biosynthetic Gene Cluster of Streptomyces Coelicolor," *Molecular Microbiology,* vol. 56, no. 2, pp. 465-479, 2005.
[29] T. Perkins and P. Swain, "Strategies for Cellular Decision-Making," *Molecular Systems Biology,* vol. 5, no. 1, pp. 326-341, 2009.
[30] H. Kim and E. Gelenbe, "G-Networks Based Two Layer Stochastic Modeling of Gene Regulatory Networks with Post-Translational Processes," *Interdisciplinary Bio Central,* vol. 3, pp. 1-8, 2011.
[31] E. Gelenbe and R. Muntz, "Probabilistic Models of Computer Systems Part I (Exact Results)," *Acta Informatica,* vol. 7, no. 1, pp. 35-60, 1976.
[32] E. Gelenbe, "Diffusion Approximations: Waiting Times and Batch Arrivals," *Acta Informatica,* vol. 12, pp. 285-303, 1979.
[33] E. Gelenbe and R. Schassberger, "Stability of Product form G-Networks," *Probability in the Eng. and Informational Sciences,* vol. 6, no. 3, pp. 271-276, 1992.
[34] E. Gelenbe, "G-Networks with Signals and Batch Removal," *Probability in the Eng. and Informational Sciences,* vol. 7, no. 3, pp. 335-342, 1993.
[35] E. Gelenbe and A. Labed, "G-Networks with Multiple Classes of Signals and Positive Customers," *European J. Operational Research,* vol. 108, no. 2, pp. 293-305, 1998.
[36] E. Gelenbe and J. Fourneau, "Random Neural Networks with Multiple Classes of Signals," *Neural Computation,* vol. 11, no. 4, pp. 953-963, 1999.
[37] E. Gelenbe, "Network of Interacting Synthetic Molecules in Equilibrium," *Proc. Royal Soc. A (Math. and Physical Sciences),* vol. 464, pp. 2219-2228, 2008.
[38] E. Gelenbe, "Genetic Algorithms with Analytical Solution," *Proc. First Ann. Conf. Genetic Programming,* pp. 437-443, 1996.
[39] E. Gelenbe, "Search in Unknown Random Environments," *Physical Rev. E,* vol. 82, no. 6, p. 061112, 2010.

Haseong Kim received the BS degree in horticulture science from the College of Life Science, Korea University, and the MS degree in bioinformatics from Seoul National University. From 2006 to 2008, he was a research assistant in the Laboratory of the Bioinformatics and Biostatistics in the Department of Statistics at Seoul National University. Then, he started the PhD degree at Imperial College London. His research area is mainly focused on the gene regulatory network construction and the dynamics of the networks.

Erol Gelenbe received the MSc and PhD from the Polytechnic Institute of New York University (Brooklyn Poly) and the Docteur-ès-Sciences degree from the University Pierre et Marie Curie (Paris VI). He is the Professor in the Dennis Gabor Chair, Electrical and Electronic Engineering Department, Imperial College London and is an alumnus of the Middle East Technical University in Ankara. He combines theoretical and experimental investigations with the development of software tools and systems. His current research addresses green IT, QoS, and security of computer networks, emergency management and evacuation, auctions in the digital economy, and search in large unknown spaces, winning several recent best paper awards. In recent years, he developed the concept of self-aware networks as part of initiatives in autonomic communications. He also works on analytical solutions for queueing network models, models ond stochastic models of gene regulatory networks and neural networks. His current projects include the UK Technology Strategy Board SATURN Project (2010-2012) with BT, Northrup-Grumman, Oxford, and Warwick Universities; the EU FP7 Fit4Green (2010-2012) with GSI, HP, VTT, ENEL and others. More information can be found at http://www.ee.ic.ac.uk/gelenbe.