Rapid Report

Interdisciplinary Bio Central

Open Access, Open Review Journal

G-Networks Based Two Layer Stochastic Modeling of Gene Regulatory Networks with Post-Translational Processes

Haseong Kim[1] and Erol Gelenbe[1,*]

[1]Imperial College, Intelligent Systems and Networks Group, Department of Electrical and Electronic Engineering, Exhibition Road, London SW7 2BT, UK

Subject areas: Bioinformatics/Computational biology/Molecular modeling

Author contribution: H.K. analysed the model and applied it to biological data; E.G. developed the mathematical model for gene regulatory networks and proved its properties.

***Correspondence** and requests for materials should be addressed to E.G (e.gelenbe@imperial.ac.uk).

Editor: Hong Gil Nam, POSTECH, Republic of Korea

Received April 28, 2011;
Accepted May 06, 2011;
Published May 06, 2011

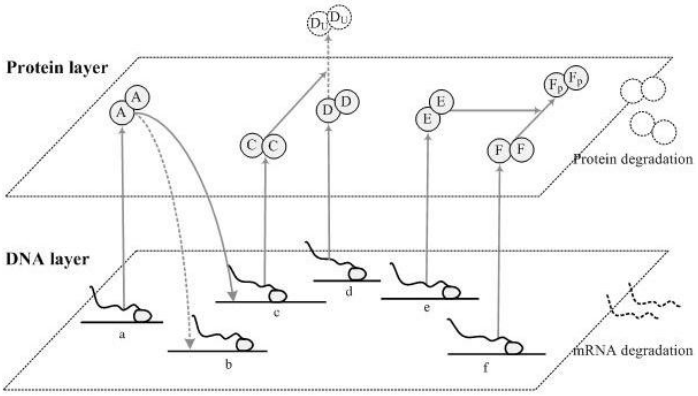
Citation: Kim, H., et al. G-Networks Based Two Layer Stochastic Modeling of Gene Regulatory Networks with Post-Translational Processes. IBC 2011, 3:8, 1-6.
doi: 10.4051/ibc.2011.3.2.0008

Competing interest: All authors declare no financial or personal conflict that could inappropriately bias their experiments or writing.

This article is part of the special issue : the 9th Asia Pacific Bioinformatics Conference (APBC2011).

SYNOPSIS

Background: Thanks to the development of the mathematical/statistical reverse engineering and the high-throughput measuring biotechnology, lots of biologically meaningful gene-gene interaction networks have been revealed. Steady-state analysis of these systems provides an important clue to understand and to predict the systematic behaviours of the biological system. However, modeling such a complex and large-scale system is one of the challenging difficulties in systems biology.
Results: We introduce a new stochastic modeling approach that can describe gene regulatory mechanisms by dividing two (DNA and protein) layers. Simple queuing system is employed to explain the DNA layer and the protein layer is modeled using G-networks which enable us to account for the post-translational protein interactions. Our method is applied to a transcription repression system and an active protein degradation system. The steady-state results suggest that the active protein degradation system is more sensitive but the transcription repression system might be more reliable than the transcription repression system.
Conclusions: Our two layer stochastic model successfully describes the long-run behaviour of gene regulatory networks which consist of various mRNA/protein processes. The analytic solution of the G-networks enables us to extend our model to a large-scale system. A more reliable modeling approach could be achieved by cooperating with a real experimental study in synthetic biology.



Keywords: systems biology, gene regulatory networks, G-networks, queueing system, large-scale modeling

Background

One of the fundamental problems of systems biology is to understand a complex biological system such as a gene regulatory network (GRN). Various mathematical and statistical models have been introduced for the inference of the GRNs[1-4]. Along with these reverse engineering approaches, the development of bioengineering techniques accelerates to uncover the regulatory mechanisms between genes for the GRN construction. Once a GRN is built, it is essential to monitor its cellular behaviour in respond to external stresses or environmental changes. Especially, studying the steady-state effect of the external perturbation to the GRNs allows us to predict the long-term behaviours of a biological system[5].

Recently, probabilistic Boolean Networks (PBNs) are proposed to investigate the steady-state of GRNs[5], but their model is not detailed enough to explain the delicate cellular processes and, in spite of the simplified (binarized) signal, they still suffer from their heavy computing time[6]. A new approach to the steady-state analysis of GRNs based on G-Network theory[7,8] is proposed in[9]. The G-Network is a probabilistic queuing network having special customers which include positive and negative customers, signals and triggers[7,8]. In contrast to a normal queueing theory, G-networks introduced negative customers which describe the inhibitory effect in a system.

In our previous work[10] which is also based on G-networks, GRNs are analyzed to monitor their steady-state behaviours and to detect anomaly expressed genes responding to external perturbations in the yeast cell cycle. However, the model cannot account for the effects of post-translational processes such as phosphorylations and ubiquitinations which play an important role in gene regulation. Mitogen-activated protein (MAP) kinase pathway in mammalian systems is the one example of non-transcriptional and cascade signaling pathways[11]. F. Markowetz *et. al*.[12] proposed a nested effect model (NEM) that can infer the non-transcriptional structure of innate immune response (IIR) in Drosophila melanogaster[12-14].

Our proposed model consists of two layers one of which is a DNA (mRNA) layer and the other is a protein layer. The mRNA expression on the DNA layer is described by a single queueing system and the protein layer is modeled by G-networks. The translation process and transcription factor activities are responsible for the transmission of gene regulatory information between the two layers. We will show a simple numerical example to help the understanding of the proposed model. Also two different regulatory networks responding to external signals are modeled and compared each other in terms of their long-run behaviours.

Results and Discussion

Active degradation for a fast response to a stress

In a GRN, the efficient way of achieving a fast response to a external stress is known to use the active degradation of proteins rather than repressing their transcription[15]. For example, 50% of LexA in E. coli is degraded within 1 minute after its related protein RecA is activated, which looks to be hard to reduce the concentration of LexA by repressing the transcription of LexA in that short time[16]. Here, we show two different types of systems, transcription inhibition model (M_{ti}) and active degradation model (M_{ad}), responding to an extracellular signal. In model M_{ti} (Figure 1 (a)), protein *B* senses the level of signal *C* and their protein complex *BC* represses the expression of gene *A* by binding its operator site. On the other hand, model M_{ad} (Figure 1 (b)) has an active degradation process in responding to the signal *C*. Protein complex *BC* acts as an ubiquitin ligase which is responsible to the active degradation of protein *A*.

These two systems are modeled using our proposed stochastic method and compared in terms of the response efficiency by measuring the steady-state probability (SSP) and the average number of molecules (ANM). Figure 2 shows the SSPs and ANMs with respect to the input rates of external signal *C* which is varied from 0 to 0.1.

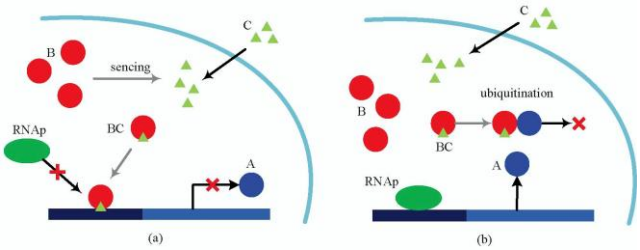


Figure 1. Signal response models. (a) Transcription inhibition model, M_{ti}. *B* senses the intracellular level of a small molecule, *C*, and the complex *BC* represses the expression of *A* by binding to *A*'s operator. (b) Active degradation model (M_{ad}). The protein *A* is degraded by the protein complex *BC* which acts as an ubiquitin ligase.

In Figure 2 (c) and (f), the SSP and ANM of the complex *BC* are almost the same, which means there is no difference between these two systems in detecting the internal signal *C*. The reason that complex *BC* has the steep increment of ANM in (f) is because it's the expected numbers of molecules in the steady-states of the systems. Figure 2 (b) and (c) show that M_{ad} has a faster response system (it reacts in the low level of *C*) than M_{ti} which needs an enough level of *C* (≈0.05) to turn ON its response mechanism. These results are consistent with the previous study[17] which shows the active degradation systems are responsible for the fast responses.

However, if the level of *C* is kept above 0.05 in M_{ti}, protein *A* is completely removed whereas M_{ad} has a possibility that some of protein *A* still stay in a cell even when the concentration of *C* is high enough. It might be because the mRNA of the protein *A* in M_{ad} keep producing its protein molecules even though the molecules are degraded quickly (Figure 2 (a) dashed line). Therefore, we can conclude that M_{ad} is more sensitive and faster response system than M_{ti} but M_{ti} could be a more reliable system in removing the target molecule as a response of the external signal.

Conclusions

In this study, we propose a new approach to analyze the steady-state of GRNs which include post translation processes such as phosphorylation and ubiquitination. The analytic solution of G-

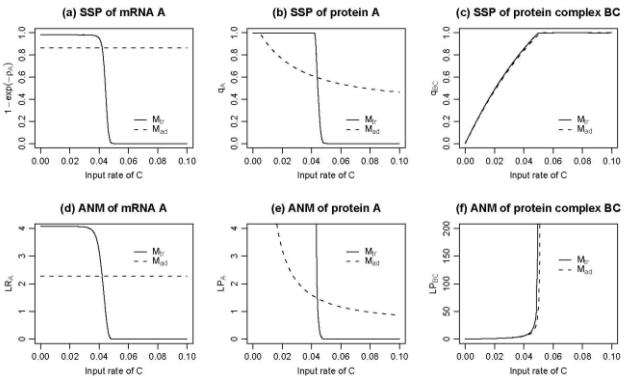


Figure 2. Steady-state probability and average number of molecules in the two response models. (a)-(c) Steady-state probabilities (SSP) that each queue is not empty and (d)-(f) average number of molecules (ANM) in both M_{ti} (solid line) and M_{ti} (dashed line).

networks and the queuing theory enable us to explore the long run behaviour of GRNs without the expensive computation of steady-state analysis. Also the negative signals of G-networks provide a successful way to describe the active degradation of protein molecules by ubiquitination, which is not possible in conventional queueing system. In the stress response example, our method shows that the active degradation system response more quickly to external signal than the transcription repression system, however, the transcription repression could be more appropriate system to wipe out a target protein completely.

Along with the DNA and protein layers, our model can be extended by adding a metabolic layer which has more conveniently observable behaviours of organisms. However, it is necessary to consider the modeling approaches that enable to describe the reverse effect of a chemical reaction before the metabolic layer modeling. Also cooperating with the studies of engineered real organisms in synthetic biology could be helpful to build a more reliable and practical GRN multiple models.

Methods

Biological assumptions for gene regulation

The gene regulation is largely involved with the activities of various protein molecules such as transcription factors, repressors, and activators. Proteins act alone or together with other proteins to promote or block the expression of genes. Based on these protein activities, our study focuses on following seven gene regulation related mechanisms (Figure 3).

(B-I) Transcription (Figure 3 (a)) - Activator proteins such as transcription factors enhance the interaction between an RNA polymerase and the promoter site of a gene by increasing their binding affinity. Some proteins bind to an enhancer which is a short region in a DNA strand to increase transcription levels. Apart from this transcription process, there is another DNA-protein interaction called repression that DNA-binding proteins (repressors) are attached to the operator of a gene in order to prevent an RNA polymerase binds to its target promoter and decrease the transcription rate of the gene.

(B-II) Translation (Figure 3 (b)) - In prokaryotic cells, ribosomes bind successively to an mRNA as soon as it is accessible to the mRNA and produce corresponding proteins successively. These translation processes are continued until the mRNA is degraded by an RNAse-E. It is known that the expression pattern of proteins is different with that of mRNAs. Many studies have interpreted the variation in terms of a burst which takes place in a brief period of high expression intensity followed by a long period of low intensity[17,18].

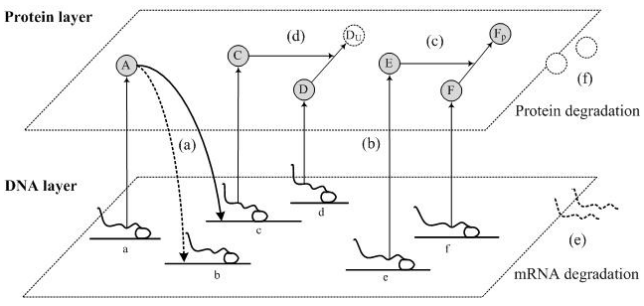


Figure 3. Gene regulation assumption. Gene *a* activates the transcription of gene *c* and represses gene *b*. Protein C add ubiquitin to protein D which will be degraded after ubiquitination. Gene *e* produces kinase E which phosphorylates protein F and then F is activated by this phosphorylation.

(B-III) Phosphorylation (Figure 3 (c)) - Protein phosphorylation is usually involved in the positive regulation of gene expressions. It activates or inactivates many proteins by adding a phosphate (PO_4) group to the substrate proteins or other organic molecules. The phosphorylation is accomplished by a kinase which transfers phosphate groups from high-energy donor molecules. In our study, phosphorylation processes are assumed to be involved with only positive gene regulations.

(B-IV) Ubiquitin dependent protein degradation (Figure 3 (d)) - Ubiquitination is a post-translational modification process that labels proteins for degradation by attaching ubiquitin monomers. This process is performed through a cascade of three enzymes, E1, E2, and E3. E1 activates an ubiquitin which is then transferred to ubiquitin-conjugating enzyme E2. A substrate for the ubiquitination is recognized by ubiquitin ligase E3. The ubiquination is required the hundreds of ubiquitin ligases which is the complex protein of E2-E3 where E2 attaches the ubiquitin to the target protein. Then this tagged substrate is degraded into peptides by 26S proteasome.

(B-V) mRNA degradation (Figure 3 (e)) - The lifetimes of prokaryotic mRNAs range from seconds to more than an hour. An mRNA is cleaved by RNase-E and then is degraded by a combination of ribonucleases including 3'-5' exonucleases. RNase-E shares its binding site with Ribosomes which are responsable for the protein production. In eukaryotic cells, mRNAs can survive from several minutes to days. The degradation of actively translated mRNAs is blocked by ribosomes and poly-A binding protein.

(B-VI) Protease dependent protein degradation (Figure 3 (f)) - Proteases can be any proteins that are responsible for protein breakdown. There are six groups of proteases which are found in lysosomes of animal cells and in vacuoles of yeast and plant cells. These proteases are involved in a multitude of physiological reactions from simple digestion of food proteins to highly-regulated systems such as apoptosis. In our model, the degradation effect of proteases is fixed as a constant while that of ubiquitination depends on the level of E3 enzyme.

(B-VII) Protein-protein interaction - Though various protein-protein interactions play key roles in a living organism, our model concerns a binary reaction channel having following equation $A+B \rightarrow C$ where *A*, *B*, and *C* are all protein molecules. The case where more than two proteins are involved can be modeled using a series of these binary reaction channels. Inducers which bind to repressors and prevent the repressors' activity can be also modeled with this assumption.

DNA layer modeling

Let's imagine a cell where stores mRNA molecules of *i*th gene. Then, this cell producing the mRNA can be considered as the M/M/∞ system in queueing theory where its input and output processes are Poisson (M) and ∞ indicates infinite number of servers. The number of mRNAs in the cell is affected by an input (transcription) rate (λ_i) which depends on the level of its transcription factors. The effect of the transcription factors on the mRNA production will be discussed in Section 0.4. The initial transcription rate of a gene with its transcription factors is assumed to be a constant (e.g. $\lambda_i = 0.0067\text{sec}^{-1}$; one mRNA per 2.5min[18,19]). The degradation rate of an mRNA is also fixed as μ_i which is obtained on the basis of the half-life of an mRNA molecule. However, the total amount of degrade mRNAs depends on the current number of mRNA, *n*. So the average number of mRNA molecules can be computed by modifying birth-death process whose input and output rates are dependent on the current state of the system.

Let p_{i_n} be the probability that a cell has *n* molecules of *i*th mRNA and $\mu_{i_n} = n\mu_i$ where $1 \le n \le \infty$. Then the following balance equations are obtained.

$$p_{i_n} = \frac{1}{n!}\left(\frac{\lambda_i}{\mu_i}\right)^n p_{i_0} \quad (1)$$

Let $\lambda_i/\mu_i = \rho_i$ and $0 \le \rho_i \le 1$. Since the sum of all probabilities of p_{i_n} is one, p_{i_0} is

$$\sum_{n=0}^{\infty} \frac{1}{n!}\rho_i^{\,n}\rho_{i_0} = 1 \quad (2)$$

$$\rho_{i_0} = e^{-\rho_i}$$

Therefore, the steady-state probability that there are n mRNAs in a cell is

$$\rho_{i_n} = \frac{\rho_i^{\,n}}{n!}e^{-\rho_i} \quad (3)$$

Then the average number of mRNAs (LR_i) in a cell is

$$\begin{aligned} LR_i &= \sum_{n=0}^{\infty} n p_{i_n} \\ &= \sum_{n=0}^{\infty} n \frac{\rho_i^{\,n}}{n!}e^{-\rho_i} \\ &= \rho_i \end{aligned} \quad (4)$$

Protein layer modeling

Though the numbers of mRNAs and proteins can be directly/indirectly affected by each other, it is more realistic to model mRNA and protein systems separately. That is, an mRNA only provides itself as a template of corresponding protein production and many modeling approaches assume that the degradation of the mRNA is only depends on the population size of the mRNA. Therefore, apart from the mRNA system, protein expression is modeled using G-network theory[7,8]. A G-network is a stochastic dynamical model having positive and negative agents and signals. Note that the agents and signals represent proteins in this protein layer model. Our model is specially focusing on the number of "activated" proteins rather than protein monomers which is the first product of translation. Monomer proteins could be activated by multimorization or phosphorylation.

(P-I) Activated protein input by translation process

For one mRNA molecule, the probability that the number of proteins (m) produced per burst is $p^m(1\text{-}p)$ where p and $(1\text{-}p)$ are the probabilities that a ribosome and an RNase-E bind to the mRNA, respectively. Let's assume the burst size is equal to the average number of proteins per one mRNA then, $b=p/(1\text{-}p)$. Note that b is varied; 100 for an average *E. coli* gene[20], 40 for lacZ, and 5 for lacA[18]. Let T_D be the average time interval between successive competitions of ribosome and RNase-E. Then the surviving mRNAs, n, in the population after its transcription is blocked would be $n=n_0\, p^{T\text{half}/TD}$ which is equal to $T_{\text{half}} = -(\log(2)/\log(p))T_D$[21].

$$T_D = \frac{\log(b) - \log(1+b)}{\log(2)} T_{half} \quad (5)$$

Therefore, if we know the average half-life of mRNAs and a burst size, the translation initiation rate for a protein monomer, $\lambda_{mono,i} = LR_i / T_D$, can be determined. The geometric distribution of the above protein expression is the discrete version of gamma distribution (sum of exponential distribution) in[18,22] which include the effect of protein delusion by cell cycle. So our assumed protein expression can be considered as a gamma distribution with its average frequency of expression burst per cell cycle is set to one. We assume the burst size b=7.8 and the half-life of mRNA is 150 sec.

Some proteins are activated in a form of a dimer or an oligomer rather in monomer. It is known that the protein multimerization is an important feature in regulations of bio-molecules such as enzymes or transcription factors[23]. The dimer dissociation constant K_D is varied from 0.1 to 10 where $K_D=[P]^2/[PP]$; $[P]$ and $[PP]$ are the monomer and dimer protein concentrations, respectively[24]. In this study, we assume the input rate of an activated protein is

$$\lambda_{p,i} = 10\lambda_{mono,i}^2 \quad (6)$$

(P-II) Activated protein input by phosphorylation

Consider protein A which is activated by a phosphorylation process where kinase B is responsible for the phosphorylation. Then, the reaction equation will be $A_p + B \xrightarrow{\lambda_{p,i}} A + B_p$ where A_p and B_p are phosphorylated A and B, respectively, and $\lambda_{p,i}$ is the phosphorylation rate. However, our protein expression model only concerns about the activated proteins so the reaction can be re-written as $A_p \xrightarrow{\lambda_{p,i}} B_p$. That is when this reaction occurs, A_p and B_p decreases and increases by 1, respectively. We assume $\lambda_{p,i}$ =0.0023 which is a protein-protein association rate in Buchler, N. E. et al.[24].

(P-III) Activated protein output by ubiquitination

As we mentioned in (B-IV), an ubiquitinaton target protein is recognized by E3 enzymes which is bound with E2 enzymes together. This process can be expressed $A_u + B \xrightarrow{\mu_{p,i}} A + B_u$ which is similar with the case of phosphorylation. Then the reduced form of this equation will be $A_u \xrightarrow{\mu_{p,i}} B_u$. The negative customers of G-networks are used to model this active degradation process so $\mu_{p,i}$ is a negative customer input rate. We also use $\mu_{p,i}$ =0.0023 the same as the phosphorylation rate.

(P-IV) Activated protein output by protease dependent degradation

The 1st order reaction equation is defined by $d[A]/dt = k[A]$ where k is reaction constant and $[A]$ is the concentration of protein A. Then the decrement of 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. Therefore, if we know the half-life of a protein, the reaction (degradation) rate can be obtained by $k = \log(2)/T_{\text{half}}$. The average of half-life of proteins is known as from 15min to 120min[18]. We assume the half-life of activated proteins is 30min so the activated protein output rate by protease dependent degradation is k=0.000385.

(P-V) Protein-protein interaction

In our model, two types of protein reaction channels are considered[25]. (i) Protein P_i turns into protein P_j, whose reaction equation is $P_i \xrightarrow{\mu_{pj}} P_j$ which is the same form as that of phosphorylation and ubiquitination processes. (ii) Protein complex P_l is synthesized by binding two proteins P_i and P_j together. In this case, the reaction equation is $P_i + P_j \rightarrow P_l$. Note that the case involving more than two proteins can be modeled by considering the reactants, P_i and P_j, are protein complexes which are previously formed by the same reaction.

Based on the these mathematical assumptions, we can build the balance equation of the protein expression which is the same form of G-networks[9]. Let's consider a random process $\boldsymbol{X}(t)=[X_1(t),...,X_n(t)]$ where $X_i(t)$ be an integer-valued random variable representing the number of ith protein molecules at time t ($t \ge 0$) and an n-vector of non-negative integers $\mathbf{x}=[x_1,...,x_n]$. If $\Pr(\mathbf{x},t)$ is the probability that $\mathbf{X}(t)$ takes $\mathbf{x}$ at time t, then the probability that $\mathbf{X}(t)$ have $\mathbf{x}$ at time $t+\Delta t$ is defined as

$$\begin{aligned} \Pr(\mathbf{x}, t+\Delta t) = \sum_{i=1}^{n} \Big[& (\lambda_{p,i}\Delta t + o(\Delta t))\Pr(\mathbf{x}_i^-, t) I(\mathbf{x}_i > 0) \\ & + (\mu_{p,i}\Delta t + o(\Delta t))\Pr(\mathbf{x}_i^+, t) \\ & + \sum_{j=1}^{n} \Big\{ (p_{ij}^+ r_i \Delta t + o(\Delta t))\Pr(\mathbf{x}_{ij}^{+-}, t) I(\mathbf{x}_j > 0) \\ & + (p_{ij}^- r_i \Delta t + o(\Delta t))\Pr(\mathbf{x}_{ij}^{++}, t) \\ & + (p_{ij}^- r_i \Delta t + o(\Delta t))\Pr(\mathbf{x}_i^+, t) I(\mathbf{x}_j \text{ is } 0) \\ & + \sum_{l=1}^{n} \Big((p_{ijl} r_i \Delta t + o(\Delta t))\Pr(\mathbf{x}_{ijl}^{++-}, t) + (p_{jil} r_j \Delta t + o(\Delta t))\Pr(\mathbf{x}_{ijl}^{++-}, t) \Big) I(\mathbf{x}_l > 0) \Big\} \\ & + (d_i r_i \Delta t + o(\Delta t))\Pr(\mathbf{x}_i^+, t) \\ & + (1 - (\lambda_i + \mu_i + r_i)\Delta t + o(\Delta t))\Pr(\mathbf{x}, t) \Big] \end{aligned} \quad (7)$$

Where $\mathbf{x}_i^+(\mathbf{x}_i^-)$ is a vector that the value of ith element is $\mathbf{x}_i$+1 ($\mathbf{x}_i$-

1); $I(C)$ is indicator function which is 1 if the condition, C, is true or 0 other wise; $o(\Delta t)\to 0$ as $\Delta t\to 0$; $\lambda_{p,i}$ and $\mu_{p,i}$ are the positive and negative input rates of the ith protein, respectively. $d_i \times r_i$ is the protein degradation rate (0.000385) where d_i is the probability that a customer (protein) leaves the system and r_i is a server activation rate. The first term describes the increment of the activated proteins by its translation and the second term explains the active degradation by ubiquitination. $p_{ij}^{+}(p_{ij}^{-})$ is the probability that protein i phosphorylates (ubiquitinates) protein j. So the third term is the probability that protein i phosphorylates protein j in time t. The fourth and fifth terms describe the ubiquitination process where protein i ubiquitinases protein j with probability (p_{ij}^{-}) so both of the i and j are depleted by 1 if this event is happened. The sixth term represents the probabilities that the interaction of protein i and j is creating protein complex l. The seventh term is the probability that the protein i is degraded by the protease dependent (non-ubiquinization) protein degradation (nothing is happened except that protein i lose one molecule). The last term represents nothing happened.

E.Gelenbe proposed a solution of the Equation (7) in Gelenbe, E. et al.[9]. Let q_i be the probability that activated protein i performs its function then q_i is defined by

$$q_i = \left[1, \frac{\Lambda_i^{+}}{r_i + \Lambda_i^{-}}\right] \tag{8}$$

$$\begin{aligned} \Lambda_i^{+} &= \lambda_{p,i} + \sum_{j=1}^{n} q_j r_j p_{ji}^{+} + \sum_{j,l=1,l\neq j}^{n} q_j q_l r_j p_{jli} \\ \Lambda_i^{-} &= \mu_{p,i} + \sum_{j=1}^{n} q_j r_j p_{ji}^{-} + \sum_{j,l=1,l\neq j}^{n} q_l r_l p_{lij} \end{aligned} \tag{9}$$

Then the steady-state probability that there are x_i molecules of activated protein i in a cell is

$$\begin{aligned} \lim_{t\to\infty} p_i(x_i,t) &= p_i(x_i) \\ &= q_i^{x_i}(1-q_i) \end{aligned} \tag{10}$$

Therefore, not only the protein expression from a single mRNA but also the expression considering our biological assumptions including post-translation processes can be explained by a geometric distribution. The average number of activated protein i is $LP_i=q_i/(1-q_i)$ ($0\le LP_i\le\infty$, $0\le q_i<1$). Note that the probabilities, p_{ij}^{+}, (p_{ij}^{-}), and p_{ijl} are all set to $1/N_{out}$ where N_{out} is the number of out degree of a protein in a protein network structure.

Bridging mRNA and protein layers

Since the number of mRNAs depends on the concentration of its transcription factors (activated proteins), the input rate of an mRNA can be a function of its transcription factors. Consider ith mRNA which is transcribed by jth transcription factor. Then the input rate of ith mRNA can be described by a function of the number of the transcription factor, $\lambda_i = f(LP_j)$. f can take various forms depending on the types of ith gene response. In this study, we define f which is similar to the Michaelis-Menten description of enzyme kinetics,

$$f^{+} = \frac{\alpha(LP_j/K)^h}{1+(LP_j/K)^h}, \quad f^{-} = \frac{\alpha}{1+(LP_j/K)^h} \tag{11}$$

where f^{+} and f^{-} indicate activation and repression function, respectively; K is the Michaelis constant; h is a Hill coefficient; is a scaling parameter. We assume h=4 (tetramerized protein) and $K=b$(=7.8) (Average number of protein molecules per one transcription). The scaling parameter α =0.0133 is obtained under the condition of $K=LP_j$ so that the transcription rate with the effect of Equation (11) is to be the same as $1/T_{init}$ which is the initial transcription rate in a single gene expression model (T_{init}=150sec in this study). It is because the initial transcription rate is assumed with the level of its corresponding transcription factors is sufficient enough.

The average number of protein j, LP_j, is depending on its average number of mRNA, LR_j, as described in assumption (P-I). Let $\rho_i^{(0)}$ be the initial value of ρ_i. Then we can compute $LP_i^{(0)}$ using corresponding input and output rates of ith protein. If protein j is the transcription factor of ith gene, $\rho_i^{(1)}$ can be obtained based on the jth protein level $LP_j^{(0)}$ which is computed by the same way as $LP_i^{(0)}$. We can find the optimal average numbers (including ρ_i and q_i) of mRNAs and proteins by iterating these steps until the following join steady-state probability of protein expression is converged.

$$p(\mathbf{x}) = \prod_{i=1}^{N} q_i^{x_i}(1-q_i) \tag{12}$$

where $\mathbf{x}=x_1, x_2, \ldots, x_N$. That is, if we define $p(\mathbf{x})^{(k)}$ is the joint probability which is computed using corresponding parameters in kth iteration, then the iteration is continued until $p(\mathbf{x})^{(k)} - p(\mathbf{x})^{(k-1)} < c$ where c is stopping criterion. Note that, the proof of this product form of joint probability can be found in the G-network theory[9].

Toy example of a single gene expression and the inference of its parameter values

Let's assume there is a gene A which produces mRNA and corresponding protein. The transcription rate of mRNA is λ_A = 0.0067 sec^{-1} by assuming one mRNA is produced every 150 sec. The mRNA half-life is set to 300 sec so the degradation rate of the mRNA is μ_A =0.00231 (=log(2)/300sec). Therefore, the average number of mRNA in a cell is $LR_A=\rho_A$ = 2.9 (4). The steady-state probability that at least one mRNA in a cell is0.945 (=1-$e^{-\rho A}$) (Equation 2).

Now, we need to compute the average number of protein A, LP_A, in a cell. However, a previous experimental study[17] implies the LP_A of our model should be the same as the burst size b(=7.8) when we assume there is no interaction with other proteins and the burst frequency per cell cycle is fixed one. So, the average number of protein A is $q_A/(1-q_A)=b$ =7.8. In order to describe the actual processes in protein layer, we need to know the unknown parameters in the Equation (8) such as r_i, p_{ji}^{+} and p_{ji}^{-}. The translation initiation rate $\lambda_{p,A}$ =0.034 is obtained by using Equation (6) and the ubiquitination rate is $\mu_{p,A}$ =0.0023. Along with the condition, $\sum_{j\in S(i)}(p_{ij}^{+} + p_{ij}^{-}) + d_i = 1$ where $S(i)$ is the set of proteins that are affected by ith protein, we assume these transition probabilities are distributed uniformly. That is $p_{ij}^{+} = p_{ij}^{-} = \cdots = d_i = 1/NS(A)$ where $NS(A)$ is the length of set $S(i)$ which is the out-degree of ith protein. The degradation rate of a protein molecule is d_A r_A = 0.000385 with the 1800 sec half-life of a protein but we can re-write it as $\frac{1}{NS(A)} \times \frac{\beta}{NS(A)} = 0.000385$ by assuming the activation rate r_A is an inverse proportion of the number of out-degrees. Since, in this example, $NS(A)$=1, we use β=0.000385 for the other system to find q_i of Equation (8).

References

1. Friedman, N., Linial, M., Nachman, I., and Pe'er, D. (2000). Using Bayesian networks to analyze expression data. *J Comput Biol 7*, 601-620.
2. Hirose, O., Yoshida, R., Imoto, S., Yamaguchi, R., Higuchi, T., Charnock-Jones, D.S., Print, C., and Miyano, S. (2008). Statistical inference of transcriptional module-based gene networks from time course gene expression profiles by using state space models. *Bioinformatics* 24, 932-942.
3. Martin, S., Zhang, Z., Martino, A., and Faulon, J.L. (2007). Boolean dynamics of genetic regulatory networks inferred from microarray time series data. *Bioinformatics* 23, 866-874.
4. Opgen-Rhein, R., and Strimmer, K. (2007). Learning causal

networks from systems biology time course data: an effective model selection procedure for the vector autoregressive process. *BMC Bioinformatics* 8 Suppl 2, S3.

5. Shmulevich, I., Gluhovsky, I., Hashimoto, R.F., Dougherty, E.R., and Zhang, W. (2003). Steady-state analysis of genetic regulatory networks modelled by probabilistic boolean networks. *Comp Funct Genomics* 4, 601-608.
6. Ching, W.K., Zhang, S., Ng, M.K., and Akutsu, T. (2007). An approximation method for solving the steady-state probability distribution of probabilistic Boolean networks. *Bioinformatics* 23, 1511-1518.
7. Gelenbe, E. (1991). Product-form queueing networks with negative and positive customers. *Journal of Applied Probability*, 656-663.
8. Gelenbe, E. (1993). G-networks with triggered customer movement. *Journal of Applied Probability* 742-748.
9. Gelenbe, E. (2007). Steady-state solution of probabilistic gene regulatory networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 76, 031903.
10. Kim, H., and Gelenbe, E. (2009). Anomaly detection in gene expression via stochastic models of gene regulatory networks. *BMC Genomics* 10 Suppl 3, S26.
11. Marshall, C.J. (1995). Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* 80, 179-185.
12. Markowetz, F., Kostka, D., Troyanskaya, O.G., and Spang, R. (2007). Nested effects models for high-dimensional phenotyping screens. *Bioinformatics* 23, i305-312.
13. Boutros, M., Agaisse, H., and Perrimon, N. (2002). Sequential activation of signaling pathways during innate immune responses in *Drosophila*. *Dev Cell* 3, 711-722.
14. Tresch, A., and Markowetz, F. (2008). Structure learning in Nested Effects Models. *Stat Appl Genet Mol Biol* 7, Article9.
15. Sneppen, K., Krishna, S., and Semsey, S. (2010). Simplied Models of Biological Networks. *Annual review of biophysics*, 39:43
16. Little, J.W. (1983). The SOS regulatory system: control of its state by the level of RecA protease. *J Mol Biol* 167, 791-808.
17. Cai, L., Friedman, N., and Xie, X.S. (2006). Stochastic protein expression in individual cells at the single molecule level. *Nature* 440, 358-362.
18. Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci U S A* 98, 8614-8619.
19. Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell* 123, 1025-1036.
20. Paulsson, J. (2005). Models of stochastic gene expression. *Physics of life reviews* 157-175.
21. McAdams, H.H., and Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A* 94, 814-819.
22. Paulsson, J., Berg, O.G., and Ehrenberg, M. (2000). Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. *Proc Natl Acad Sci U S A* 97, 7148-7153.
23. Marianayagam, N.J., Sunde, M., and Matthews, J.M. (2004). The power of two: protein dimerization in biology. *Trends Biochem Sci* 29, 618-625.
24. Buchler, N.E., Gerland, U., and Hwa, T. (2005). Nonlinear protein degradation and the function of genetic circuits. *Proc Natl Acad Sci U S A* 102, 9559-9564.
25. Gelenbe, E. (2008). Network of interacting synthetic molecules in steady state. Proceedings of the Royal Society A: Mathematical, *Physical and Engineering Science* 464(2096).