The Second International Symposium on Optimization and Systems Biology (OSB'08)
Lijiang, China, October 31– November 3, 2008

Sensitivity Analysis of a Cis-Regulatory Input Function

Ai-Min Chen[1] Tian-Shou Zhou[1,2]

[1]School of Mathematics and Computational Science
Sun Yat-Sen University, Guangzhou 510275, China
[2]State Key Laboratory of Biocontrol and Guangzhou Center for Bioinformatics,
School of Life Science Sun Yat-Sen University, Guangzhou 510275, China

Abstract Stochastic effects of regulatory molecules in low copy numbers per cell in transcription-regulation process have now been recognized as a major physiologically and evolutionarily important factor in the development and function of many living organisms. The transcription rate of a gene is often controlled by several regulators that bind specific sites in the gene's cis-regulatory region. The combined effect of these regulators is described by a cis-regulatory input function(CRIF), which is usually nonlinear. Since the average of a nonlinear function is generally not the same as the function of the average, in addition neither analytical nor numerical solutions of the Chemical Master equation are in general available, we take mass fluctuation kinetics(MFK) in place of mass action kinetics(MAK) to analyze the stochastic effects on sensitivity of cis-regulatory input functions with AND- and OR-logic gate.

Keywords Cis-regulatory module; cis-regulatory input function; stochastic focusing; mass fluctuation kinetics; sensitivity analysis

1 Introduction

Most genes are regulated by multiple transcription factors (TFs) that bind specific sites in DNA regulatory regions. These cis-regulatory regions perform a computation: the rate of transcription is a function of the active concentrations of each of the input transcription factors. This point has been demonstrated, for example, for the endo-16 gene during sea-urchin development, where multiple transcription factors combine to perform an intricate logical computation. Cis-regulatory regions are usually studied by genetic methods, by deleting the various transcription factors or mutating sites in the regulatory region. The picture that emerges from such studies is often stated in terms of logic gates such as AND and OR gates[1, 2].

To illustrate how different regulatory functions can be implemented by using the model description, we consider the response of AND in Fig.1a, which corresponds to the logic function AND, and the implementation of which is referred bo as the AND gate. It can be obtained by choosing weak binding sites for both X_1 and X_2 and placing them adjacent to each other (see Fig.1a) such that each TF alone cannot bind to its site, but when both are present binding occurs with the help of the additional cooperative interaction.

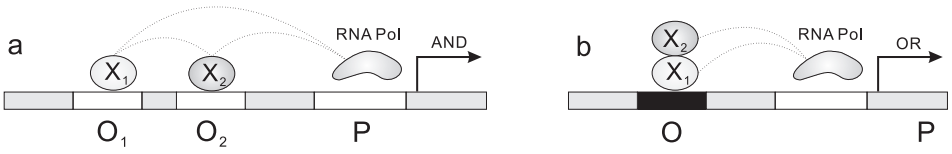

Figure 1: Different Cis-regulatory constructs of the AND (a) and OR (b). Filled and open boxes denote strong and weak binding sites, respectively. Overlapping boxes in (b) indicate repulsive interaction (or competitive binding).

Similarly, one can implement the responses for the target gene in Fig.1b corresponding to the OR gate with strong binding site. Perhaps most strikingly, Transcription factors binding sites often overlap with one another. Here we discuss the two transcription factors competing for binding to the promoter region for OR-gate[2, 3, 4, 5].

Recently, a great wealth of genomic data and experimental findings have confirmed that transcription networks contain significantly recurring wiring patterns termed “network motifs". Among the significant networks in bacterium *Escherichia coli* and yeast *Saccharomyces cerevisiae*, the most common one is the feed-forward loop (FFL)[6, 7, 8, 9]. The FFL consists of a transcription factor X_1, which regulates a second transcription factor X_2. X_1 and X_2 both bind the regulatory region of target gene X_3 and jointly modulate its transcription rate. In this report, we analyze two kinds of input signals: 1)two relate signals in FFL combinatorially regulate the target gene; 2)two simple independent input signals.

Noise inherently exists in cellular process, such as gene expression, signaling transduction and metabolic activities because of the limited number of molecules for typical molecular species[10, 11, 12]. For example, the number of LacI tetrameric repressor proteins in *E.coli* has been estimated to be of the order of 10 to 50 molecules per cell. In previous study, the researchers mainly investigate the structure and function of the deterministic properties. Less attention, however, has been paid into the theoretical investigation on how the fast fluctuation of transcription factors, e.g., stochasticity of input signals, influence the steady state average response and response sensitivity. Our numerical experiments as well as theoretical analysis show that noisy signals exhibit a significant influence on the mode of dynamic response, e.g., stochastic dynamic response shows different results from that of macroscopic deterministic response.

In order to capture the inherent stochasticity of a system of chemical reactions, the present most proper method is using the chemical master equation to track the evolution of the joint probability distribution of the molecule numbers[13]. However, neither analytical nor numerical solutions of chemical master equations are in general available. By using the moment closure approximation Mass fluctuation kinetics (MFK)[14, 15] describe the evolution of the system in terms of the mean and covariance values of concentrations in the course of time.

The main idea that is presented in this report is that the coherent type I network motif with AND- and OR- logic operate in Alon’s eight types of network motifs can carry out specific information-processing functions by using MFK in place of mass action kinetics (MAK). With mathematical modeling, we analyze the sensitivity of coherent type I FFL structural configuration with AND- and OR-gate logic and compare the results of that with two simple independent input signals.

2 Multi-Dimensional Input Functions

Many genes are regulated by more than one transcription factor. The combined effects of these regulators can be described by a multi-dimensional input function. Here we examine and discuss the multi-Dimensional input functions with two TFs.

Cells receive a wide variety of intracellular and extracellular signals, which are often processed combinatorially to generate specific genetic response [2, 3, 4, 6, 5, 8]. Many of these combinational effects are performed at the level of cis-regulatory transcription control, which can function as analogous implementations of logic gates. Every cis-regulatory module contains a series of binding sites of different TFs that control the activation or repression of a gene. These TFs may be activators enhancing the binding or the activity of the RNA polymerase in the cognate promoters, or repressors blocking this binding, or both via the mechanism of "regulated recruitment". Note that in our case, two signaling molecule X_1 and X_2 are taken as input activators. The cis-regulatory constructs are shown in Fig. 1, although more complex cis-regulatory modules can occur in the natural setting.

2.1 Input Function That Integrates Two Activators: AND- and OR-Logic Operation

Let us take a look at an input function that integrates two activators X_1 and X_2 at a promoter. How can two activators work together? A common situations that the activators bind the promoter independently on two different sites. Therefore, there are four binding states of promoter D: D, D_{X_1}, D_{X_2}, and $D_{X_1X_2}$, where $D_{X_1X_2}$ means that both X_1 and X_2 bind to D. Transcription occurs mainly from the state $D_{X_1X_2}$, in which activators X_1 and X_2 both bind with weak binding sites. In the following we use X_1 and X_2 to denote the active forms X_1^* and X_2^*.

The AND represents a regulatory system where the gene expression turns on only when the two activator is present, which can be materialized as a promoter that is activated by two TFs. In such a setup, if one activator is not bound to the promoter, the activator is unable to act. Moreover, this function requires a weak binding sites for both TFs which ensures that each TF alone cannot bind to its site, but when both are present binding occurs with the help of the additional cooperative interaction. For example, the comprehensively studied arabinose-utilization system in the bacterium *Escherichia coli* can be regarded as the realization of AND function with two input signals, namely the concentrations of TFs (cyclic AMP receptor protein and arabinose), and one output signal (the expression level of the operon)[4]. In our case, signal molecules X_1 and X_2 both act as activators.

The OR performs a function that the output is turned "Off" only when the two inputs is low. In our case, it means that the target gene is not expressed ("Off") only when the signal molecules X_1 and X_2 both are absent. Here we assume the two transcription factors competing for binding to the overlapping promoter region.

2.2 Two Kinds of Input Signals

case I: Here, we consider two independent signals shown as follows:

$\varnothing \xrightarrow{k_1} X_1, X_1 \xrightarrow{\lambda_1} \varnothing, \varnothing \xrightarrow{k_2} X_2, X_2 \xrightarrow{\lambda_2} \varnothing$

The stationary solution is a joint distribution of two independent Poisson distribution. The reason is that the molecules X_1 and X_2 are created and annihilated independently of

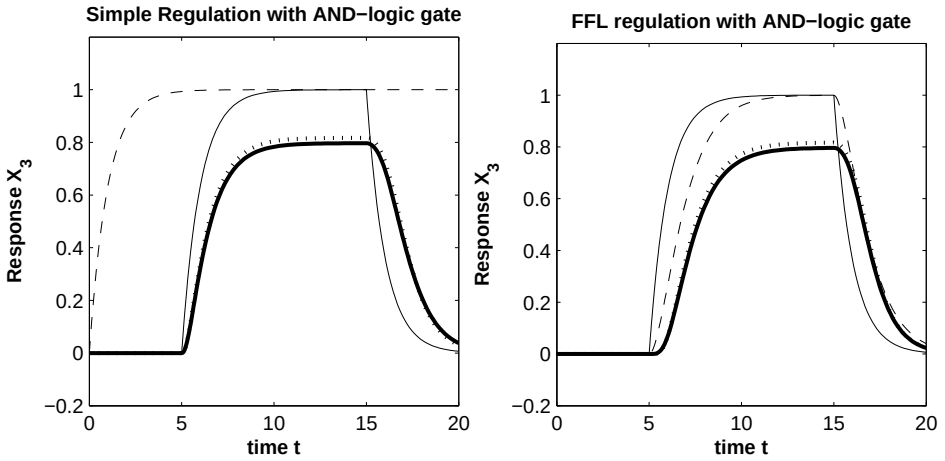

Figure 2: Response of target protein X_3 regulated by two signals with AND-logic operating: (a)two simple independent signals and (b)two relate signals in feed-forward loop. Numerical analysis verified that the C1-FFL with AND input function both shows delay after stimulus X_1 addition and after the stimulus removal. The thin lines show the evolution of signal X_1 with time, the dashed lines designate the signal X_2, the thick black line show the response, X_3 with MFK, while the dotted lines show the response, X_3 with MAK.

each other. In reaction schemes in which several molecules X_1 and X_2 react together they are no longer independent and deviations from the poisson distribution will occur.

case II: Feed-forward loops work as two dependent signals:

$\varnothing \xrightarrow{k_1} X_1, X_1 \xrightarrow{\lambda_1} \varnothing, X_1 \xrightarrow{k_2} X_1 + X_2, X_2 \xrightarrow{\lambda_2} \varnothing$

Though this is a linear Chemical Master equation, its solution is still not available analytically.

2.3 Derivation of Mathematical Models

In this section, we derive our mathematical models in much detail. Throughout this paper we let $X \equiv \{X_1, X_2, X_3\}$ be the chemical species, $X \equiv \{X_1, X_2, X_3\}$ the numbers of molecules, $x \equiv \{x_1, x_2, x_3\}$ the molecules concentration . Molecule numbers are normalized by the system size, given by $\Omega = Av$ where A is Avogadro's number, to yield concentrations in moles per unit volume, so that $x = X/\Omega$. Specifically in this report, X_1 and X_2 designate the two TFs (input signals), X_3 is the target protein (the output). Through MAK, we are easily to written out the deterministic ordinary equations of the chemical reactions[**?**]. Here we assume the rates of binding and dissociation of TFs and DNA binding sites are much more quickly compared to degradation, so that they are taken as to lead to the chemical equilibrium easily (we take concentrations as our dynamical variables):

$$\begin{aligned} K_1 d_{x_1} &= dx_1 & K_2 d_{x_2} &= dx_2 \\ K_3 d_{x_1 x_2} &= d_{x_1} x_2 & K_4 d_{x_2 x_1} &= d_{x_2} x_1 \end{aligned} \tag{1}$$

where $d, d_{x_1}, d_{x_2}, d_{x_1 x_2}, d_{x_2 x_1}$ express the concentration of the corresponding binding states of the promoter respectively. We point out that the fast reaction equilibrium recipe based

on quasi-steady state approximation approach is widely applied to reduce the complexity of multiscale problems.

Using the conservation law of total DNA-binding sites without considering the diffusion of signaling molecules, the transcription rate with AND-logic gate is easily obtained, f/d$_T$.

$$d_T = d + d_{x_1} + d_{x_2} + d_{x_1x_2} + d_{x_2x_1} \tag{2}$$

where d_T describes the total concentration of the promoter. Making use of the fast reaction equilibrium Eqs. (1) in combination with Eqs. (2), we can obtain

$$d = \frac{d_T}{1 + x_1/K_1 + x_2/K_2 + x_1x_2/(K_1K_3 + K_2K_4)} \tag{3}$$

Let $K_3 = K_2/2, K_4 = K_1/2$, then with AND-gate cis-regulatory module, the CRIF of two transcription activators is

$$f = d_T \frac{(x_1/K_1)^{H_1}(x_2/K_2)^{H_2}}{1 + (x_1/K_1)^{H_1} + (x_2/K_2)^{H_2} + (x_1/K_1)^{H_1}(x_2/K_2)^{H_2}} \tag{4}$$

where H_1,H_2 are the Hill coefficients. Similarly, for an OR-gate cis-regulatory module (with the two transcription factors competing for binding to the promoter region), the CRIF is easily derived out:

$$f = d_T \frac{(x_1/K_1)^{H_1} + (x_2/K_2)^{H_2}}{1 + (x_1/K_1)^{H_1} + (x_2/K_2)^{H_2}} \tag{5}$$

Setting fast reactions to be in equilibrium, the transcription and degradation process then simplifies to:

$$\varnothing \xrightarrow{kf} \mathrm{X}_3, \qquad \mathrm{X}_3 \xrightarrow{\lambda} \varnothing \tag{6}$$

Furthermore, based on mass action kinetic mechanism, we can write mass balances for each of reactants of interest

$$\frac{dx_3}{dt} = kf - \lambda x_3 \tag{7}$$

3 Mass Fluctuation Kinetics

However, the intrinsic stochastic effects in chemical reactions, and particularly in biochemical networks, may result in behaviors significantly different from those predicted by deterministic mass action kinetics (MAK). Analyzing stochastic effects even with the chemical Master equation (CME), however, is often computationally taxing and complex. The authors describe here the derivation and applications of what Gómez-Uribe et. al.[15] termed the mass fluctuation kinetics (MFK), a set of deterministic equations to track the means, variances, and covariances of the concentrations of the chemical species in the system.

To capture the inherent stochasticity of a system of chemical reactions, the present method to describe the joint probability distribution of the molecule numbers is the chemical master equation (CME), namely,

$$\frac{\partial P(X,t)}{\partial t} = \sum_{i=1}^{R} (\nu(X - S^i)P(X - S^i) - \nu(X)P(X)) \tag{8}$$

where the vector $X \equiv \{X_1, X_2, X_3\}$ describes the reaction molecules number of the corresponding components. ν and S designate the propensity function of the R reactions and the stoichiometry matrix respectively. Starting from the CME, the evolution of the average, variance and covariance of molecules concentration would be well approximated by a closed system of equations when third-order moments about the mean are negligible. To state the MFK equations, we designate J, Γ and σ the Jacobi matrix, diffusion matrix and variance matrix of concentration respectively. Then the coupled MFK equations are described as follows:

$$\begin{aligned} \frac{d\langle x\rangle}{dt} &= S \cdot \nu(\langle x\rangle) + \sum_{i,j=1,2,3} \frac{\partial^2 \nu(\langle x\rangle)}{\partial x_i x_j} \sigma_{ij} \\ \frac{d\sigma}{dt} &= \Gamma\sigma + \sigma\Gamma^T + \frac{1}{\Omega} S \cdot diag(\nu(\langle x\rangle)) \cdot S^T \end{aligned} \tag{9}$$

where $\langle\cdot\rangle$ shows average. The validity of macroscopic approaches to describe averages cannot be taken for granted obtained from the MAK rate laws because the average of a nonlinear function is generally not the same as the function of the average. In the above, the cis-regulatory function f is usually nonlinear. Let S designate the chemical stoichiometric matrix, (σ_{ij}) designate the variance covariance matrix. Here, we investigate the average of nonlinear functions by mass fluctuation kinetics (MFK).

3.1 Definition of Direction Sensitivity

Since recent studies define the response sensitivity of one input signal as:
Sensitity $\triangleq \frac{\Delta Response/Response}{\Delta signal/signal} \approx \frac{d\ln Response}{d\ln signal}$,
here X_3 is taken as response of two input signals X_1 and X_2 we give directional sensitivity (DS) for two input signals as:
$DS_1 \triangleq \frac{\partial \ln x_3}{\partial \ln x_1} cos\theta + \frac{\partial \ln x_3}{\partial \ln x_2} sin\theta$
in the direction paralleling to any one of the coordinate axes, or
$DS_2 \triangleq \frac{\frac{x_3(x_1+\Delta x_1, x_2+\Delta x_2) - x_3(x_1,x_2)}{x_3(x_1,x_2)}}{\frac{\sqrt{\Delta x_1^2 + \Delta x_2^2}}{\sqrt{x_1^2+x_2^2}}} \approx \frac{\sqrt{x_1^2+x_2^2}}{x_3(x_1,x_2)} (\frac{\partial x_3}{\partial x_1} cos\theta + \frac{\partial x_3}{\partial x_2} sin\theta)$
in the direction otherwise.

4 Outlook and Future work

We have analyzed the stationary state properties of stochastic reaction cis-regulatory modules based on the mass fluctuation kinetics. Through analysis, we mainly obtained:

- **Result I:** Cis-regulatory Input Functions(CRIFs) with AND or OR gate logic operation of two activators show different sign-sensitive delay[4]. AND both shows

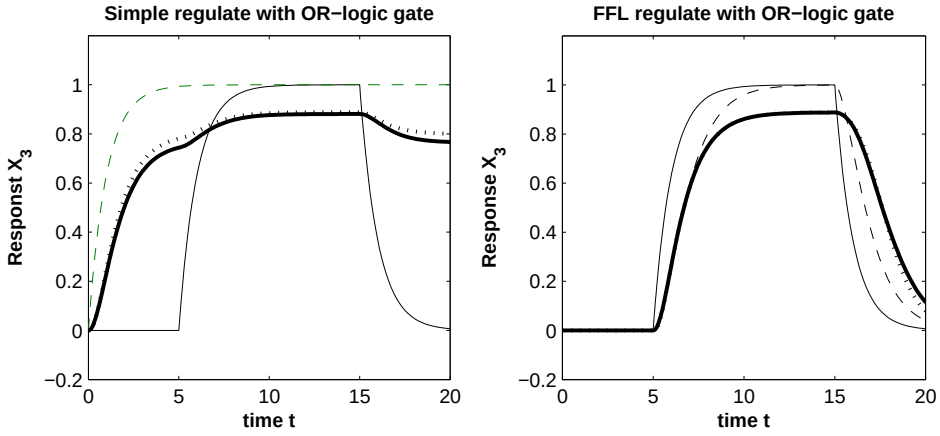

Figure 3: Response of target protein X_3 regulated by two signals with OR-logic operating: (a)two simple independent signals and (b)two relate signals in feed-forward loop. Numerical analysis verified that the C1-FFL with OR input function shows no delay after stimulus X_1 addition and and delay after the stimulus removal. The thin lines show the evolution of signal X_1 with time, the dashed lines designate the signal X_2, the thick black line show the response, X_3 with MFK, while the dotted lines show the response, X_3 with MAK.

delays after stimulation and after the stimulation stops. When the X_3 promoter has OR logic, it shows delay after the stimulation stops while not delay after the addition of stimulation. However, when the two input signals are independent and one is fixed constant, the other's ON step causes a sign-sensitive acceleration in X_3 expression, the other's OFF step causes a sign-sensitive delay in X_3 removal. The numerical results are shown in Fig.2 and Fig.3.

- **Result II:** With mass fluctuation kinetics (MFK), not mass action kinetics (MAK), it has been shown that AND and OR gate CRIFs both exhibit stochastic focusing, the phenomenon of stochastic focusing is not very obvious, the numerical results not shown here for the limited pages. [16].
- **Result III:** Based on definition of one dimensional sensitivity, we give the directional sensitivity of systems with two input signals. Stochasticity results show different results from the deterministic analysis with deterministic MAK. Numerical simulation validate the analytic results.

In this paper, we propose two biologically feasible cis-regulatory modules designs: AND and OR. In our proposed models, signaling molecules can serve only as activators by the introduction of an alternative promoter [17]. In spite of the simplicity of the two cis-regulatory modules, quite complex functionality at the stochastic level can emerge as shown in the article. However, there are many other combinatorial transcription logic operations with two transcription factors, we will analyze the detailed stochastic effects on detailed map of different cis-regulatory input functions in the following work.

Acknowledges

We acknowledge the support from the Natural Science Key Foundation of People's Republic of China (No. 60736028) and support from Research Fund of Youth Scholars for the Doctoral Program of Higher Education of China (No. 20070558053).

References

[1] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman & Hall/CRC, London, 2006).

[2] N. E. Buchler, U. Gerland, and T. Hwa, On Schemes of Combinatorial Transcription Control. Proc. Natl. Acad. Sci. U.S.A. **100**, 5136 (2003).

[3] R. Hermsen, S. Tans, and P. R. ten Wolde, Transcriptional regulation by competing transcription factor modules. PLoS Comput. Biol. **2**, e164 (2006).

[4] S. Mangan and U. Alon, Structure and function of the feed-forward loop network motif. Proc. Natl. Acad. Sci. U.S.A. **100**, 11980 (2003).

[5] A. E. Mayo, Y. Setty, S. Shavit, A. Zaslaver and U. Alon, Plasticity of the cis-Regulatory Input Function of a Gene. PLoS Biol. **4**, e45 (2006).

[6] Y. Setty, A. E. Mayo, M. G. Surette, and U. Alon, Detailed map of a cis-regulatory input function. Proc. Natl. Acad. Sci. U.S.A. **100**, 7702 (2003).

[7] U. Alon, Network motifs: theory and experimental ap-proaches. Nat. Rev. Genet., **8(6)**, 450–461 (2007).

[8] Shen-Orr, S. S., Milo, R., Mangan, S. and Alon, U. Network motifs in the transcriptional regulation network of Escherichia coli. Nat. Genet. **31**, 64–68 (2002).

[9] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. Network Motifs: Simple Building Blocks of Complex Networks. Science **298**, 824–827 (2002).

[10] M. Thattai and A. van Oudenaarden, Attenuation of noise in ultrasensitive signaling cascades Biophys. J. **82**, 2943 (2002).

[11] J. Paulsson, Summing up the noise in gene networks Nature **427**, 415 (2004).

[12] T. Shibata and K. Fujimoto, Noisy signal amplification in ultrasensitive signal transduction. Proc. Natl. Acad. Sci. U.S.A. **102**, 331 (2005).

[13] N. G. van Kampen, Stochastic Processes in Physics and Chemistry (North-Holland, Amsterdam, 1992).

[14] K. H. Kim, H. Qian, H. M. Sauro, Sensitivity Regulation based on Noise Propagation in Stochastic Reaction Networks arXiv:0805.4455 (2008).

[15] C. A. Gómez-Uribe and G. C. Verghese. Mass fluctuation kinetics: capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations. J. Chem. Phys., **126**(2), 024109 (2007).

[16] J. Paulsson, O. G. Berg, and M. Ehrenberg. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. Proc. Natl. Acad. Sci. U.S.A., **97**(13), 7148–7153 (2000).

[17] K. A. Egland and E. P. Greenberg. Quorum Sensing in Vibrio fischeri: Analysis of the LuxR DNA Binding Region by Alanine-Scanning Mutagenesis. J. Bacteriol. **183**, 382–386 (2001).