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

Analysis of network dynamics including hidden variables by symbolic-numeric approach

Daisuke Tominaga[1] Yasuhito Tokumoto[2]
Masahiko Nakatsui[1] Fuyan Sun[1] Jun Miyake[2,3]
Katsuhisa Horimoto[1]

[1]Computational Biology Research Center, AIST, 2-42 Aomi, Koto, Tokyo 135-0064, Japan
[2]The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan
[3]Research Institute for Cell Engineering, AIST, 1-8-31 Midorigaoka, Ikeda, Osaka 563-8577, Japan

Abstract We propose a symbolic-numeric method for estimating the kinetic constants in a biological network including hidden variables which mean that the behaviors of corresponding molecules cannot be directly measured. In the present method, an algebraic manipulation of the differential equations over the Laplace domain, formulated based on the assumption of linear relationships between the variables, is combined with the numerical fitting of the sampling data. The performance of the method is illustared for a part of MAPK network with the data measured by the transfection cell array in combination of the gene interference by siRNAs.

Keywords Symbolic Computation; Numerical Optimization; Network Dynamics; Hidden Variable; Bi-fan structure

1 Introduction

The clarify of the dymanics of a complex network is one of the important issues in systems biology. By the recent advances of the experimental technology in molecular biology, the behaviors of a large numbers of genes such as gene expression levels can be measured simultaneously in different conditions. However, it is still difficult to measure the time series of gene expression data. Indeed, the transfection cell array[1] is one of most advanced technology for measuring the time series of gene expressions in a living cell, but even by using these experiments, the gene expressions are measured for only a small number of repoter genes, in which the fluorescence protein is artificially encoded. In usual, it encounters frequently the difficulty for measuring the molecule behaviors in biological experiments, and for analyzing the network including hidden variables in the biological networks. Thus, it is challenging to clarify the dynamics of whole network only from the measurement of a small fraction of constituent molecules.

In this paper, we propose a symbolic-numeric approach for estimating kinetic constant in the case when the time series of expressions of reporter genes are measured by the transfection cell array in combination of the interference of the remaining genes by siRNAs. In this case, the number of the reporter genes are limited, and thus time-dependent behaviors

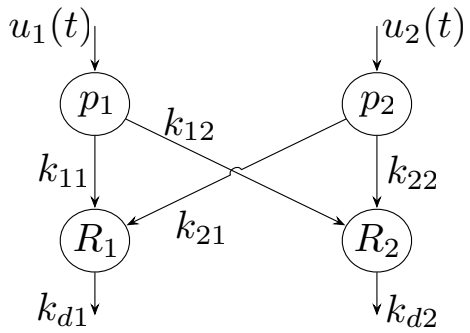

Figure 1: The network model analyzed in the present study.

are not measured in most constituent genes. Here, by using our approach, we present a solution for estimating the network dynamics in a partial model including hidden variables of MAPK pathway.

2 Materials and Methods

2.1 Model

We consider a network in Fig.1. In the network, we assume that the expression levels of two molecules, R_1 and R_2, can be measured as the reporter genes. These two molecules degrade with respective known constant rates, k_{d1} and k_{d2}. We also assume that any expression levels can not be measured in two molecules, p_1 and p_2, which change by unknown external forces, $u_1(t)$ and $u_2(t)$. The kinetic constants between them are k_{11}, k_{12}, k_{21}, and k_{22}.

2.2 Formulation over Laplace domain

The dynamics of the molecules in Fig.1 is expressed by the following ordinary differential equations:

$$\begin{aligned}
R_1^{0\prime}(t) &= k_{11}p_1(t) + k_{21}p_2(t) - k_{d1}R_1^0(t) \\
R_2^{0\prime}(t) &= k_{12}p_1(t) + k_{22}p_2(t) - k_{d2}R_2^0(t) \\
R_1^{-p1\prime}(t) &= k_{21}p_2(t) - k_{d1}R_1^{-p1}(t) \\
R_2^{-p1\prime}(t) &= k_{22}p_2(t) - k_{d2}R_2^{-p1}(t) \\
R_1^{-p2\prime}(t) &= k_{11}p_1(t) - k_{d1}R_1^{-p2}(t) \\
R_2^{-p2\prime}(t) &= k_{12}p_1(t) - k_{d2}R_2^{-p2}(t)
\end{aligned} \tag{1}$$

where R_i^0 R_i^{-X} indicate the expression levels when no genes are surpresse and that when gene X is surpressed by the corresponding siRNA, respectively.

Then, Eqns. (1) are also expressed as a system of the corresponding algebraic equa-

tions, by Laplace transformation, i.e.,

$$\begin{aligned}
sL[R_1^0(t)] - R_1^0(0) &= k_{11}L[p_1(t)] + k_{21}L[p_2(t)] - k_{d1}L[R_1^0(t)] \\
sL[R_2^0(t)] - R_2^0(0) &= k_{12}L[p_1(t)] + k_{22}L[p_2(t)] - k_{d2}L[R_2^0(t)] \\
sL[R_1^{-p1}(t)] - R_1^{-p1}(0) &= k_{21}L[p_2(t)] - k_{d1}L[R_1^{-p1}(t)] \\
sL[R_2^{-p1}(t)] - R_2^{-p1}(0) &= k_{22}L[p_2(t)] - k_{d2}L[R_2^{-p1}(t)] \\
sL[R_1^{-p2}(t)] - R_1^{-p2}(0) &= k_{11}L[p_1(t)] - k_{d1}L[R_1^{-p2}(t)] \\
sL[R_2^{-p2}(t)] - R_2^{-p2}(0) &= k_{12}L[p_1(t)] - k_{d2}L[R_2^{-p2}(t)]
\end{aligned} \tag{2}$$

where $L[R(t)]$ is a function in s obtained by Laplace transformation of $R(t)$.

Apart from the network model, we fit the measured data of expression levels by exponential polynomials, i.e.,

$$R(t) = \sum_{i=1}^{n} a_i \exp(-m_i t). \tag{3}$$

Then, Eqn. (3) are expressed as a system of the corresponding algebraic equations by Laplace transformation, i.e.,

$$L[R(t)] = \sum_{i=1}^{n} \frac{a_i}{s+m_i} \tag{4}$$

2.3 Estimation of kinetic constants over the Laplace domain

We eliminate $L[p_1(t)]$ and $L[p_2(t)]$ from the Eqns.(2), and we obtain the following equations:

$$\begin{aligned}
k_{d1} &= \frac{R_1^0(0) - R_1^{-p1}(0) - R_1^{-p2}(0)}{L[R_1^0(t)] - L[R_1^{-p1}(t)] - L[R_1^{-p2}(t)]} - s \\
k_{d2} &= \frac{R_2^0(0) - R_2^{-p1}(0) - R_2^{-p2}(0)}{L[R_2^0(t)] - L[R_2^{-p1}(t)] - L[R_2^{-p2}(t)]} - s \\
\frac{k_{12}}{k_{11}} &= \frac{(s+k_{d2})L[R_2^{-p2}(t)] - R_2^{-p2}(0)}{(s+k_{d1})L[R_1^{-p2}(t)] - R_1^{-p2}(0)} \\
&= \frac{\frac{R_2^0(0) - R_2^{-p1}(0) - R_2^{-p2}(0)}{L[R_2^0(t)] - L[R_2^{-p1}(t)] - L[R_2^{-p2}(t)]} L[R_2^{-p2}(t)] - R_2^{-p2}(0)}{\frac{R_1^0(0) - R_1^{-p1}(0) - R_1^{-p2}(0)}{L[R_1^0(t)] - L[R_1^{-p1}(t)] - L[R_1^{-p2}(t)]} L[R_1^{-p2}(t)] - R_1^{-p2}(0)} \\
\frac{k_{22}}{k_{21}} &= \frac{(s+k_{d2})L[R_2^{-p1}(t)] - R_2^{-p1}(0)}{(s+k_{d1})L[R_1^{-p1}(t)] - R_1^{-p1}(0)} \\
&= \frac{\frac{R_2^0(0) - R_2^{-p1}(0) - R_2^{-p2}(0)}{L[R_2^0(t)] - L[R_2^{-p1}(t)] - L[R_2^{-p2}(t)]} L[R_2^{-p1}(t)] - R_2^{-p1}(0)}{\frac{R_1^0(0) - R_1^{-p1}(0) - R_1^{-p2}(0)}{L[R_1^0(t)] - L[R_1^{-p1}(t)] - L[R_1^{-p2}(t)]} L[R_1^{-p1}(t)] - R_1^{-p1}(0)}
\end{aligned} \tag{5}$$

Note that the right sides of Eqns. (5) are composed of the terms related with the repoter genes. Thus, we substitute Eqn. (44) obtained by fitting of Eqn. (3) into Eqns. (5), and we obtaine the equations in the form as $c = F(s)/G(s)$, where $F(s)$ and $G(s)$ are polynomials in s, and c is a constant value.

In the actual case, however, the equation, $c = F(s)/G(s)$, does not always hold, due to the noise of data. Thus, we estimate c so as to minimize the following formula:

$$M(c) = \int_0^{u_{max}} (cG(s) - F(s))^2 ds \tag{6}$$

By solving $\frac{\partial M(c)}{\partial c} = 0$, we obtain the following equation:

$$c = \frac{\int_0^{u_{max}} F(s)G(s)ds}{\int_0^{u_{max}} G(s)^2 ds} \tag{7}$$

The values of k_{d1} and k_{d2} are known as the constant values, and the value of u_{max} is estimated so as to minimize the following equation:

$$N(u_{max}) = \left(\frac{\int_0^{u_{max}} F_{d1}(s)G_{d1}(s)ds}{\int_0^{u_{max}} G_{d1}(s)^2 ds} - k_{d1} \right)^2 + \left(\frac{\int_0^{u_{max}} F_{d2}(s)G_{d2}(s)ds}{\int_0^{u_{max}} G_{d2}(s)^2 ds} - k_{d2} \right)^2 \tag{8}$$

By using the value of u_{max}, all constants (k_{d1}, k_{d2}, k_{12}/k_{11}, and k_{21}/k_{22}) are estimated from Eqns. (7). Note that we should check the consistency between estimated and known values of k_{d1} and k_{d2} .

Actually, we can obtain the values of k_{12}/k_{11} and k_{21}/k_{22} by the following way. First, we substitute the fitted equations in Eqns. (3) and (4) into the formulra in Eqn. (5), and then obtain the equation in the form as $c = F(s)/G(s)$, as follows:

$$\begin{aligned} \frac{k_{12}}{k_{11}} &= \frac{(s+k_{d2})\sum_{j=1}^n \frac{a_{22j}}{s+m_{22j}} - \sum_{j=1}^n a_{22j}}{(s+k_{d1})\sum_{j=1}^n \frac{a_{21j}}{s+m_{12j}} - \sum_{j=1}^n a_{21j}} \equiv \frac{F_{1211}(s)}{G_{1211}(s)} \\ \frac{k_{21}}{k_{22}} &= \frac{(s+k_{d1})\sum_{j=1}^n \frac{a_{11j}}{s+m_{11j}} - \sum_{j=1}^n a_{11j}}{(s+k_{d2})\sum_{j=1}^n \frac{a_{12j}}{s+m_{12j}} - \sum_{j=1}^n a_{12j}} \equiv \frac{F_{2122}(s)}{G_{2122}(s)} \end{aligned} \tag{9}$$

where $F(s)$ and $G(s)$ are denoted by $F_{1211}(s)$ and $G_{1211}(s)$ for k_{12}/k_{11}, and by $F_{2122}(s)$ and $G_{2122}(s)$ for k_{21}/k_{22}.

Thus, we obtain the two equations k_{12}/k_{11} and k_{21}/k_{22}, respectively, corresponding to Eqn. (6), i.e.,

$$\begin{aligned} M_{1211}(k_{12},k_{11}) &= \int_0^{u_{max}} (F_{1211}(s) - (k_{12}/k_{11})G_{1211}(s))^2 ds \\ M_{2122}(k_{21},k_{22}) &= \int_0^{u_{max}} (F_{2122}(s) - (k_{21}/k_{22})G_{2122}(s))^2 ds \end{aligned} \tag{10}$$

Table 1: Estimated values of kinetic constants.

u_max	k_{d1}	d_{d2}	k_{12}/k_{11}	k_{21}/k_{22}
-10%	0.00182164	0.00250274	0.642616	2.45827
0%	0.00211409	0.00208827	0.476623	3.16587
+10%	0.00237544	0.00182634	0.380480	3.82652

Finally, we also obtain the equations for the two ratios, corresponding to Eqn. (7), as follows:

$$\begin{aligned} \frac{k_{12}}{k_{11}} &= \frac{\int_0^{u_{max}} F_{1211}(s)G_{1211}(s)ds}{\int_0^{u_{max}} G_{1211}(s)^2 ds} \\ \frac{k_{21}}{k_{22}} &= \frac{\int_0^{u_{max}} F_{2122}(s)G_{2122}(s)ds}{\int_0^{u_{max}} G_{2122}(s)^2 ds} \end{aligned} \tag{11}$$

3 Results

We analyzed actual data measured by transfection cell arrays for a part of a network related with apoptosis in mouse[2]. In the actual network, the reporter genes are p53 and jun (R_1 and R_2 in Fig. 1), and are known to be associated with MAPK8 and MAPK14 (p_1 and p_2) by the same way as those in Fig. 1.

In this study, we set $n = 4$ in Eqn. (3). The given observed data and fitted curves to the data by the differential evolution algorithm which implemented as the NMinimize function in Mathematica 6 are shown in Fig. 2.

Table 3 shows the estimated values of k_{d1}, k_{d2}, k_{12}/k_{11}, and k_{21}/k_{22} when the estimated value of umax and $\pm 10\%$ values are used. Note that both values of kd1 and kd2 are given as 0.00192541. This value shows quite similar to the estimated k_{d1} and k_{d2}. This indicates that kinetic constants are successfully estimated in the present method.

4 Discussion

Our appoach is summarized as follows: i) The relationship between the molecules in the analyzed network is modeled by a system of ordinary differential equations. ii) The time series data of the measurable molecules in the network are numerically fitted by a system of exponential polynomials. iii) The kenectic constant values and ratios of kinetic constants are expressed by fractions of fitted polynomials in s by symbolic (algebraic) computation. iv) Finally, kinetic constants are estimated by the least square method for the fitted polynomials.

In the present study, only the ratio of the kinetic constants is obtained. In very near future, explict values of kinetic constants will be reduced by the symbolic-numeric approach. Indeed, we confirm that the formular for the exclict values from the parameter values estimated by data fitting are obtained, when the three layer model is assumed. At any rate, our approach will be one of the useful approach to reveal the network dynamics including the hidden variables.

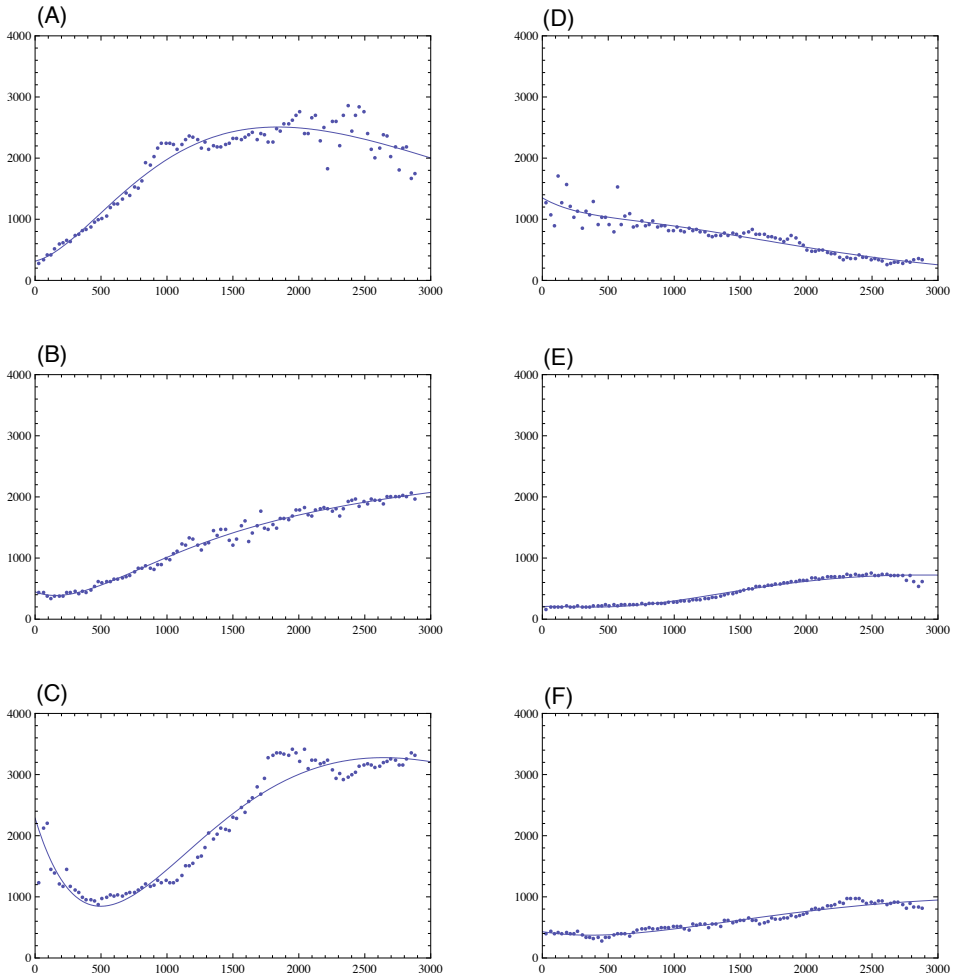

Figure 2: Plots of observed data (dots) and fitted curves (lines) used for calculation of ratios of kinetic parameters. Plots in the left column ((A), (B) and (C)) are p53 (R_1 in Fig. 1), and (D), (E) and (F) are jun (R_2). Two plots at the top ((A) and (D)) are in cases without any interference. Middle (B) and (E) are obtained by interferance in MAPK8 (p_1), and Bottom (C) and (F) are by interferance in MAPK14 (p_2).

References

[1] J. Ziauddin, D.M. Sabatini. Microarray of cel ls expressing defined cDNAs, Nature, 411 (2001) 107-110.

[2] P. Aza-Blanc, C.L. Cooper, K.Wagner, S. Batalov, Q.L. Deveraux, M.P. Cooke. Identification of Modulators of TRAIL-Induced Apoptosis via RNAi-Based Phenotypic Screening, Mol. Cel l, 12 (2003) 627-637.