Multivariate Time Series Analysis of Metabolic Network Using the E-CELL Simulation System

Yusuke Saito [13] **Kouichi Takahashi** [12] **Masaru Tomita** [13]
t96406ys@sfc.keio.ac.jp shafi@sfc.keio.ac.jp mt@sfc.keio.ac.jp

[1] Laboratory for Bioinformatics
[2] Graduate School of Media and Governance
[3] Department of Environmental Information
Keio University, 5322 Endo, Fujisawa 252-8520, Japan

1 Introduction

We applied multivariate time series analysis to metabolic models constructed using the E-CELL system. The E-CELL system [1] is a generic simulation environment for modeling cellular processes at level of chemical reaction. A state of the cell at a point in time can be represented with variables such as concentration of molecular species and flux rates of reactions. The output of the system throughout simulation is time series of those variables, and we employed the following two mathematical methods in order to analyze the output of the E-CELL system.

2 Impulse response

Impulse is a sudden shock to the system and *impulse response* is the system's response to the impulse.

Network function $H(s)$ of the system can be expressed as

$$H(s) = \mathcal{L}[ImpulseResponse]$$

where $\mathcal{L}$ is Laplace transformation. Thus, system function $H(\omega)$ can be also defined as

$$H(\omega) = \mathcal{F}[ImpulseResponse] = A(\omega)e^{-j\theta(\omega)}$$

where $\mathcal{F}$ is Fourier transformation, $A(\omega)$ is amplitude spectrum, and $\theta(\omega)$ is phase spectrum.

From the impulse response, information about (a) distances in time between reactions in a metabolic pathway, (b) dependencies among reactions and (c) other characteristics of transfer function between reactions, can be obtained.

3 Power contribution ratio

We applied the multivariate auto regressive (AR) model, which assumes that (a) the value of the series at time t depends only on its previous values and random disturbance, and (b) the dependency can be expressed as a linear combination of the values. The equation of the AR model is as follows:

$$x_i(t) = \sum_{j=1}^{k}\sum_{m=1}^{M} a_{ij}(m)x_j(t-m) + e_i(t)$$

where $x_i(t)$ is the value of x_i at the time t, k is the number of variables, M is the order of the model, $a(m)$ is the AR coefficients of the model, and $e_i(t)$ is the original innovation of x_i.

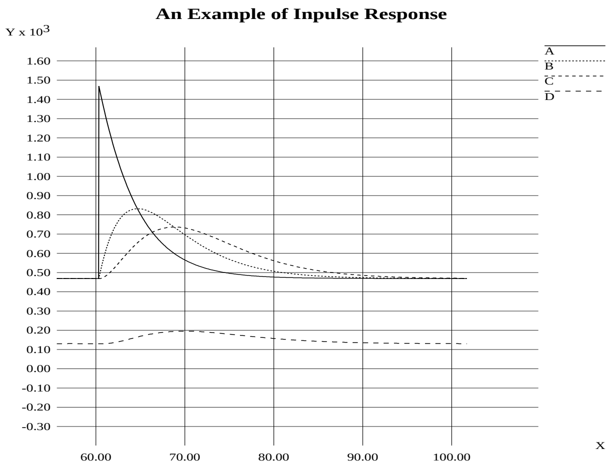

Figure 1: An Example of Impulse Response. For the sake of pilot study on impulse response analysis, we have modeled a small pathway which have only three enzymes and one feedback loop. An impulse is given to substance A. Fig. 1 shows transduction of the impulse to downstream substances (B, C and D). The distance between peaks can be viewed as distance of these two substances in time. The height of peak indicates the degree of correlation to substance A.

For oscillating time series, we analyze them in frequency domain. Specifically, we use power contribution ratio, which is an index of how fluctuation of one variable depends on the other variables at each frequency. The power contribution ratio $r_{ij}(f)$ is defined as follows:

$$r_{ij}(f) = \frac{q_{ij}(f)}{p_{ii}(f)} \qquad p_{ii}(f) = \sum_{j=1}^{k} q_{ij}(f)$$

where $p_{ii}(f)$ is power spectrum of variable x_i and $q_{ij}(f)$ is contribution ratio of variable x_j's innovation.

4 Summary

These multivariate time series analysis methods are supposed to be effective in analyzing metabolic network, particularly when feedback loops are present.

However, it is nearly impossible for laboratory experiments to obtain a large collection of sample values enough to employ these analytical methods. Since computer simulation can produce ample data sufficient for time series analyses, we believe that the E-CELL system will be a promising means for investigation of dynamic behavior of metabolic network.

Acknowledgements

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas 'Genome Science' from The Ministry of Education, Science, Sports and Culture in Japan. We would like to thank other members of the E-CELL project for their technical support in an early stage of this work.

References

[1] Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C., and Hutchison, C., E-CELL: Software environment for whole cell simulation, *Bioinformatics*, (to appear).