Reverse Engineering in Systems Biology: Towards Inference of a Gene Regulatory Network

Mineo Morohashi[1,2] **Koji M. Kyoda**[1,3] **Shuichi Onami**[1]
moro@mineo.com kyoda@kyoda.org sonami@symbio.jst.go.jp
Eric Mjolsness[1,4] **Hiroaki Kitano**[1,5]
mjolsness@jpl.nasa.gov kitano@symbio.jst.go.jp

[1] Systems Biology Group, Kitano Symbiotic Systems Project, ERATO, JST, M-31 Suite 6A 6-31-15 Jingumae Shibuya-ku, Tokyo
[2] Medical Research Institute, Tokyo Medical and Dental University, 1-5-45 Yushima Bunkyo-ku, Tokyo 113-0034, Japan
[3] Department of Computer Science, Keio University, 3-14-1 Hiyoshi Kohoku-ku, Yokohama 223-0061, Japan
[4] Jet Propulsion Laboratory, NASA, MS 126-347 Pasadena, CA 91109-8099, USA
[5] Sony Computer Science Laboratories, Inc., 3-14-13 Higashi-Gotanda Shinagawa-ku, Tokyo 141-0022, Japan

Keywords: reverse engineering, gene regulatory network, large-scale expression profiles, systems biology

1 Introduction

With the advance of technology in the field of biology, large–scale sequencing data of genomes, or massive gene expression data of mRNA are available, recently. High-throughput technology could thus measure plenty data. In particular, such as DNA microarray technology enables us to handle massive gene expression data simultaneously during a single experiment. Here arises a question: how should useful and information be mined from the data? For example, can we extract or infer a gene regulatory network from the data obtained by biological experiment? So far, in order to answer the question, some attempts were carried out, yet efficient method have not proposed.

In this paper, we introduce our work, which is an attempt to develop an advanced scheme for inferring a gene regulatory network, and for design of biological experiment, from such various amounts of experimental data. Although we have not yet established quite useful and efficient scheme of inferring a gene network, this is an initial report to show our activities up to this time.

2 Concept

Our concept for developing the scheme is to infer the most plausible and reliable gene network, providing a powerful and efficient reverse engineering method. Besides, from the viewpoint of biology, most efficient way for utilizing the reverse engineering method is to predict or to verify a putative pathway. These results will lead to design biological experiments, in which order they should be carried out. In order to achieve this, we deal with two profiles: one is time series expression profiles, and the other is steady-state expression profiles. By using both different type of data, we can extract sophisticated and reliable network, leading to develop a hybrid inferring system as briefly described in Figure 1. Needless to say, even either the data is lacked, we should be able to extract consistent gene network. We are currently working on developing several methods from different perspectives, as are briefly introduced in next section.

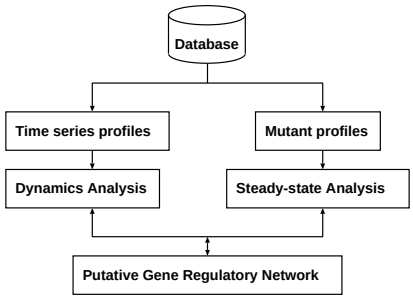

Figure 1: Schematic Representation of the Hybrid Inferring System

3 Past Works

3.1 *In silico* Sampling and Screening

In previous paper [1], we have proposed a method, *in silico* Sampling and Screening, which uses wild–type and mutant expression profiles to resolve plausible network candidates. The method is mainly consists two stage: sampling stage, and screening stage. First, in sampling stage, wild–type profile is used so that candidate network topologies and their parameters are optimized. Next, in screening stage, using mutant expression profiles, final plausible candidate networks are resolved. Using 3-node network, 8 network topologies are extracted as final candidate networks.

3.2 Hypothesis Generator

The hypothesis generator [3] is not only for identifying gene regulatory networks from gene expression profiles such as gene disruption and gene over-expression, but also for describing the experiment design and plans. Assuming the several expression patterns of genes in *Drosophila* embryo, consistent network could be extracted.

3.3 Application for Actual Expression Profiles

While the two methods mentioned above still rely on artificial fake data as the target, we have also applied a method to infer a gene network for the actual experimental data of *Saccharomyces cerevisiae* [2]. We employed clustering technique and neural network based model, and successfully achieved to reduce the dimensionality and to infer networks.

4 Concluding Remarks

We particularly focus on 1) data translation technique, and 2) parameter optimization technique currently, which we think are both the most significant issues in establishing a sophisticated reverse engineering method. The technique of translating expression data from raw experimental data, which makes easy to embed as input data in computer, would accelerate to reveal underlying gene network. For fitting model with actual data, parameter optimization should be necessary to get more reliable network. We believe that these two techniques would allow us to provide an efficient reverse engineering method.

References

[1] Morohashi, M. and Kitano, H., Identifying gene regulatory networks from time series expression data by *in silico* sampling and screening, *Proc. the 5th European Conference on Artificial Life*, 477–486, 1999.

[2] Mjolsness, E., Mann, T., Castañ̃o, R., and Wold, B., From Coexpression to Coregulation: An Approach to Inferring Transcriptional Regulation among Gene Classes from Large–Scale Expression Data, *Technical Report JPL-ICTR-99-4*, Jet Propulsion Laboratory, NASA, 1999.

[3] Kyoda, M.K., Hypothesis Generator for Gene Regulatory Networks,*Master Thesis*, Dept. of Computer Science, Keio University, 2000.