FEBS Letters 580 (2006) 3511–3518

Inferring biomolecular regulatory networks from phase portraits of time-series expression profiles

Kwang-Hyun Cho[a,b,*], Jeong-Rae Kim[b], Songjoon Baek[b], Hyung-Seok Choi[c], Sang-Mok Choo[d]

[a] College of Medicine, Seoul National University, Jongno-gu, Seoul 110-799, Republic of Korea
[b] Bio-MAX Institute, Seoul National University, Gwanak-gu, Seoul 151-818, Republic of Korea
[c] Interdisciplinary Program in Bioinformatics, Seoul National University, Gwanak-gu, Seoul 151-747, Republic of Korea
[d] School of Electrical Engineering, University of Ulsan, Ulsan 680-749, Republic of Korea

Received 6 February 2006; revised 2 May 2006; accepted 9 May 2006

Available online 22 May 2006

Edited by Robert B. Russell

Abstract **Reverse engineering of biomolecular regulatory networks such as gene regulatory networks, protein interaction networks, and metabolic networks has received an increasing attention as more high-throughput time-series measurements become available. In spite of various approaches developed from this motivation, it still remains as a challenging subject to develop a new reverse engineering scheme that can effectively uncover the functional interaction structure of a biomolecular network from given time-series expression profiles (TSEPs). We propose a new reverse engineering scheme that makes use of phase portraits constructed by projection of every two TSEPs into respective phase planes. We introduce two measures of a slope index (SI) and a winding index (WI) to quantify the interaction properties embedded in the phase portrait. Based on the SI and WI, we can reconstruct the functional interaction network in a very efficient and systematic way with better inference results compared to previous approaches. By using the SI, we can also estimate the time-lag accompanied with the interaction between molecular components of a network.**

Keywords: Regulatory networks; Phase portraits; Time-series expression profiles; Reverse engineering

1. Introduction

The complexity of biological phenomena is primarily caused by interactions of biochemical components in the underlying biomolecular regulatory networks at different layers including gene regulatory networks, protein interaction networks, and metabolic networks [1,2]. Hence, identification (or reverse engineering) of the functional interaction structure of a biomolecular network is of pivotal importance if we want to understand the essential principles prevailing the observed complex phenomena. As more high-throughput time-series measurements become available, various reverse engineering schemes have been developed to reconstruct the functional interaction structure from given time-series expression profiles (TSEPs) [3–5]. For instance, reverse engineering of gene regulatory networks from time-series microarray experiments has been getting increased attention although there still remain many problems to be resolved such as dealing with the dimensionality (i.e., relatively many network nodes but only few measurements available [6]) and computational complexity, and estimating the unknown time-lags accompanied with the interactions. To tackle such problems, scientists have combined approaches by making use of sequence information about binding motifs from databases or designing additional experiments to complement the insufficient information [7–9].

The previous studies for reverse engineering of gene regulatory networks include Boolean networks [3,10,11], Bayesian networks [12–15], dynamic Bayesian networks [16,17], and ordinary differential equations (ODEs) based methods [18–21]. The Boolean networks, Bayesian networks, and dynamic Bayesian networks allow us to infer the relation between network nodes (genes in this case), but we cannot identify the detailed regulatory relation by using these methods without additional information such as genomic-sequences [22] or degradation rates [23]. On the other hand, ODE based methods allow us to investigate underlying regulatory relations in more detail. Using this method, we can represent the expression level change of the ith node (x_i) in a network with n nodes as follows:

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \ldots, x_n) \quad (i = 1, \ldots, n).$$

Based on this framework, the influence of x_j on x_i can be represented by $\frac{\partial f_i}{\partial x_j}$. In other words, if $\frac{\partial f_i}{\partial x_j}$ is positive (negative), x_j activates (inhibits) x_i. The common drawback for all of the previous approaches is however that the computational complexity increases exponentially as the number of network nodes increases. Instead of solving ODEs, there is another approach of finding the sign of $\frac{\partial f_i}{\partial x_j}$ through perturbation of each network node [24–26], but this requires many perturbation experiments which is unfeasible for large networks.

We present in this paper a new reverse engineering scheme that can resolve the previous computational complexity problem and can be applied to various reverse engineering problems in a more efficient way. The main idea is to analyze the interaction properties embedded in the phase portrait which is drawn on a phase plane by projection of two TSEPs. If the given TSEP is dense in its sampling time intervals and the regulation is strong enough to be easily captured, we can directly infer the regulatory relation from the phase portrait. However, as this is not the case in many practical situations, we further intro-

*Corresponding author. Fax: +82 2 887 2692.
E-mail address: ckh-sb@snu.ac.kr (K.-H. Cho).

Abbreviations: TSEP, time-series expression profile; ODE, ordinary differential equation; SI, slope index; WI, winding index

0014-5793/$32.00
doi:10.1016/j.febslet.2006.05.035

duce two measures to quantify the interaction properties and to systematically infer the underlying regulatory relation: a slope index (SI) and a winding index (WI). The SI is a measure to determine the regulatory type (activation or inhibition) and the WI is for the direction of such regulation. The proposed scheme can be applied to more general cases whenever the time-dependent interaction among TSEPs is of importance. In addition, we can use this scheme to estimate the time-lag accompanied with the regulation process between two biomolecular components in a network.

The scheme we present is intuitive and very simple. Most of all, it can be easily implemented with almost negligible computational complexity. Although the two measures of SI and WI seem to be similar with the notion of correlation coefficients, the proposed scheme more clearly explains the underlying directionality of activation or inhibition than those focusing only on correlation of distributions without considering the temporal information of the TSEPs. We illustrate the proposed scheme with examples of a synthetic gene network and the chemotactic signaling network of *Dictyostelium*. In addition, we show how the proposed scheme can be applied to estimation of the time-lag accompanied with the regulation process between two genes through an example of the gene regulatory network involved in the oxidative stress of *Escherichia coli*.

2. Materials and methods

If we look at all the TSEPs at the same time, it is difficult to get any insight about the underlying interaction network as the TSEPs look just messy. Hence, the main idea of the proposed scheme is to choose every two TSEPs at a time and to infer the regulatory relation of the corresponding two nodes by investigating the dynamical characteristic of their phase portrait. The whole interaction network can be then constructed by integrating all these results. To illustrate the idea, let us first consider a regulatory network composed of only two nodes, x_1 and x_2, and assume that there exists a time-invariant regulatory relation. If x_1 activates x_2, a local maximum (minimum) of x_1 is followed by a local maximum (minimum) of x_2 as illustrated in Fig. 1A. On the other hand, if x_1 inhibits x_2, a local maximum (minimum) of x_1 is followed by a local minimum (maximum) of x_2 as shown in Fig. 1B. If we assume continuous-time expression profiles of two nodes x_1 and x_2, we define the phase portrait of these expression profiles as follows:

$$X_{12}(t) = (x_1(t), x_2(t)).$$

The phase portrait X_{12} is a curve on the phase plane spanned by x_1 as x-axis and x_2 as y-axis. Note that X_{12} loses the information about the influence from any other network nodes due to the projection from the multi-dimensional space into the two-dimensional phase plane. Thus, whenever we use this concept of a phase portrait, we implicitly assume that there is always one dominant node among the multiple nodes interacting with a given network node. Fig. 1C and D show the phase portraits of Fig. 1A and B, respectively. Since Fig. 1C still carries most of the dynamical characteristics of Fig. 1A, we can use Fig. 1C to infer the regulatory relation between x_1 and x_2. From this observation, if the phase portrait X_{12} of x_1 and x_2 locates in a positive diagonal direction and a point on X_{12} moves counter clockwise (CCW) along with time elapses like Fig. 1C, we can infer that x_1 activates x_2. Similarly, if the phase portrait of x_1 and x_2 locates in a negative diagonal direction and a point on X_{12} moves clockwise (CW) along with time elapses like Fig. 1D, we can infer that x_1 inhibits x_2. In addition, we can infer that x_2 activates x_1 if X_{12} locates in the positive diagonal direction and a point on X_{12} moves CW; x_2 inhibits x_1 if X_{12} locates in the negative diagonal direction and a point on X_{12} moves CCW.

For a network with only two nodes, we can easily identify the regulatory relation following the above observational rule. However, it might not be so evident in general for a network with multiple nodes. To deal with such general cases in a more systematic way, we introduce two measures, SI and WI in the following. In many practical situations, we can take measurements for TSEPs instead of the continuous-time expression profiles as we assumed above. In this case, we can apply the same idea to the reconstructed continuous-time expression profiles obtained by interpolating the sampled discrete-time points of TSEPs. In other words, given k discrete-time points $(t_1, \ldots, t_k)$ of a TSEP, we can construct a phase portrait by connecting each point by the line segment $\overline{X_{12}(t_i)X_{12}(t_{i+1})}$ $(i = 1, \ldots, k-1)$. In this case, we note that the phase portrait can lose the true information about the expression pattern depending on the number of available discrete-time points.

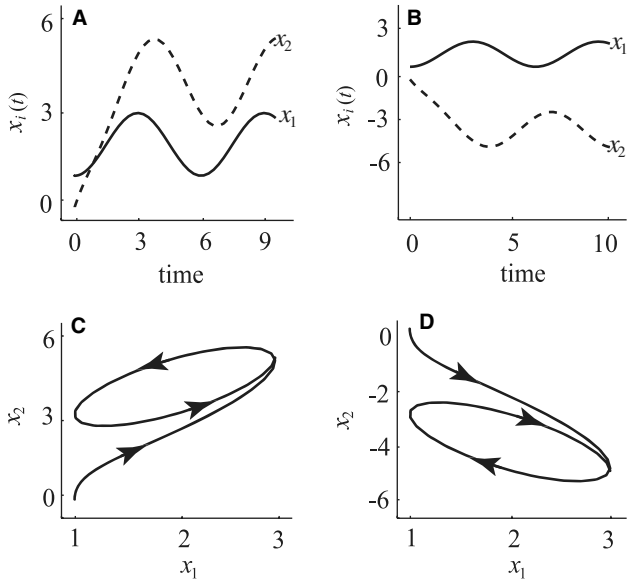

Fig. 1. A and B show two example sets of TSEPs for x_1 (solid line) and x_2 (dotted line). C and D illustrate the phase portraits of A and B, respectively.

Fig. 2A exemplifies a phase portrait of x_1 and x_2, and Fig. 2B illustrates the phase portrait of TSEPs of x_1 and x_2. Note here that the pattern of Fig. 2B can become similar to that of Fig. 2A by increasing the number of sampling points. However, since not so many sampling points are available in most real experiments at present, we need to consider a phase portrait like Fig. 2B and should infer the underlying regulatory relation from this. The dynamic pattern of Fig. 2B seems different from that of Fig. 2A, but we notice that many line segments in Fig. 2B appear in the same positive diagonal direction as the pattern of Fig. 2A.

To represent the diagonal distribution (positive diagonal or negative diagonal) and the moving direction (CW or CCW) in a quantitative way, we define the measures of SI and WI. For two network nodes x_1, x_2, and their TSEPs measured at k even sampling time points, the SI of x_1 and x_2 is defined as follows:

$$\mathrm{SI}(x_1, x_2) = \frac{1}{k-1} \sum_{i=1}^{k-1} \mathrm{sign}\left(\frac{x_2(i+1) - x_2(i)}{x_1(i+1) - x_1(i)} \right),$$

where $x_j(i)$ denotes the value of x_j at the ith sampling time point and $\mathrm{sign}(x) = 1$ for $x > 0$, $\mathrm{sign}(x) = 0$ for $x = 0$, and $\mathrm{sign}(x) = -1$ for $x < 0$ (we exclude those terms of $x_1(i+1) - x_1(i) = 0$). We also define the WI of x_1 and x_2 as follows:

$$\mathrm{WI}(x_1, x_2) = \frac{1}{k-2} \sum_{i=1}^{k-2} \mathrm{sign}(d(i)),$$

where

$$d(i) = \det \begin{bmatrix} x_1(i) & x_1(i+1) & x_1(i+2) \\ x_2(i) & x_2(i+1) & x_2(i+2) \\ 1 & 1 & 1 \end{bmatrix}$$

and $\det A$ denotes the determinant of a square matrix A. From these definitions, if the time intervals of a TSEP are uneven, there is possibility of assigning the same measure value to the ordered pairs of actually different sampling intervals. To compensate for such cases, we can